

**Universidade Federal de São Carlos
Departamento de Genética e Evolução
Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular**

Caracterização dos genomas do tatu-peba e tatu-canastra e genômica comparativa na superordem Xenarthra

Alexandre Romero Inforzato

São Carlos – SP

2024

Universidade Federal de São Carlos
Departamento de Genética e Evolução
Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular

Caracterização dos genomas do tatu-peba e tatu-canastra e genômica comparativa na superordem Xenarthra

Aluno: Alexandre Romero Inforzato

Orientadora: Profa. Dra. Patrícia Domingues de Freitas

Dissertação apresentada ao Programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular do Departamento de Genética e Evolução da Universidade Federal de São Carlos, como parte dos requisitos para obtenção do título de Mestre.

São Carlos – SP

2024

*To see a World in a Grain of Sand
And a Heaven in a Wild Flower,
Hold Infinity in the palm of your hand
And Eternity in an hour.*

William Blake

Dedico este manuscrito aos meus pais, que provavelmente não o entenderiam, mas cujas vidas também não foram inteiramente compreendidas pelos seus respectivos pais, numa corrente de afetos feita de elementos fragmentários, onde o completo entendimento do todo é – assim como em certos aspectos deste trabalho – perfeitamente opcional.

Meus profundos agradecimentos:

A todos que, de alguma maneira, contribuíram para a realização deste trabalho.

À Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de mestrado.

Ao Instituto de Conservação de Animais Silvestres (ICAS), pela atuação em projetos de conservação e pelo fornecimento das amostras de tecidos para os sequenciamentos de DNA que possibilitaram a realização deste trabalho.

Ao programa de Pós-Graduação em Genética Evolutiva e Biologia Molecular pela oportunidade da formação em mestrado.

À minha orientadora Profa. Dra. Patrícia Domingues de Freitas, pela inigualável dedicação, sensibilidade, conhecimento e persistência em seu trabalho de orientação.

Ao professor Dr. Pedro Galetti, cuja prática de ensino constituiu a motivação inicial para que eu me aproximasse do LabBMC e todas as excelentes pessoas que ali laboram.

Ao professor Dr. Julian Dupuis, pelo continuado e solícito auxílio nos aspectos mais práticos do meu trabalho, e por ser uma fonte absolutamente invulgar de inspiração como pessoa e como cientista.

Aos meus pais Nicolau e Neusa, pelo afeto verdadeiramente incondicional, que não se encontra em outro lugar.

LISTA DE FIGURAS

Figura 1 – *Pipeline* desde a predição das proteínas expressas, passando pela identificação dos ortólogos de cópia única, conversão dos arquivos FASTA de aminoácidos em FASTA de nucleotídeos, alinhamento múltiplo dos códons, criação da árvore filogenética, análise de seleção, teste de saturação de substituições, até a anotação dos genes positivamente selecionados e agregação dos resultados 15

Figura 2 – Árvores filogenéticas obtidas com a ferramenta IQ-Tree2 v2.1.2, a partir de coleções de genes ortólogos de cópia única (774 genes para o grupo de 11 espécies, e 1347 para o subgrupo de 10 - com a remoção do tatu-canastra, *P. maximus*) para os xenartros abordados neste trabalho 26

Figura 3 – Gráfico da heterozigosidade autossômica observada para as onze espécies de xenartros abordadas, incluindo o *status* atualmente atribuído a cada espécie pela IUCN (LC: Pouco Preocupante; NT: Quase Ameaçada; VU: Vulnerável) 33

LISTA DE TABELAS

Tabela 1 – Espécies de xenartros cujas montagens genômicas estavam disponíveis no NCBI até a realização do presente trabalho (último acesso em 6 de fevereiro de 2024)	9
Tabela 2 – Informações sobre os genomas de xenartros utilizados neste trabalho	10
Tabela 3 – Informações acerca dos dados de sequenciamento a serem utilizados na análise de heterozigidade autossômica envolvendo as 10 espécies de xenartros abordadas neste trabalho	16
Tabela 4 – Características das montagens genômicas para o tatu-peba e tatu-canastra, obtidos com a ferramenta QUAST v3.2, análise de completude com a ferramenta BUSCO v.5.4.6, e número de proteínas preditas com a ferramenta GeMoMa v1.9	21
Tabela 5 – Completude das montagens genômicas de xenartros, avaliada com a ferramenta BUSCO v5.4.6, utilizando o conjunto mammalia_oddb10, criado em 08 de janeiro de 2024, contendo 9226 genes ortólogos de mamíferos	23
Tabela 6 – Número de proteínas preditas pela ferramenta GeMoMa v1.9, e aumento na quantidade de proteínas preditas como resultado da desfragmentação realizada com a ferramenta RagTag v2.1.0	24
Tabela 7 – Número de genes ortólogos de cópia única identificados com a ferramenta OrthoFinder v2.5.5 utilizando os arquivos de proteínas preditas (.FAA) antes e após a desfragmentação dos respectivos genomas	25
Tabela 8 – Informações parciais referentes à anotação dos genes positivamente selecionados, com a ferramenta eggNOG v5.0.2	28
Tabela 9 – Resultados do teste de heterozigidade autossômica	32

SUMÁRIO

1. INTRODUÇÃO	1
2. JUSTIFICATIVAS E OBJETIVOS	6
3. MATERIAIS E MÉTODOS	7
4. RESULTADOS E DISCUSSÃO	20
5. CONCLUSÕES	35
6. REFERÊNCIAS BIBLIOGRÁFICAS	36

Resumo: Este estudo investigou a diversidade genética e genes sob seleção positiva dos xenartros com genomas e dados de sequenciamento atualmente disponíveis na base de dados pública GenBank (NCBI), com foco em *Euphractus sexcinctus* (tatu-peba) e *Priodontes maximus* (tatu-canastra), cujas montagens genômicas foram produzidas neste trabalho. Para as análises de genômica comparativa, as montagens inéditas de *E. sexcinctus* e *P. maximus* foram adicionadas a nove outras montagens para xenartros (*Bradypus variegatus*, *Cabassous unicinctus*, *Chaetophractus vellerosus*, *Choloepus didactylus*, *Choloepus hoffmanni*, *Dasyopus novemcinctus*, *Myrmecophaga tridactyla*, *Tamandua tetradactyla*, e *Tolypeutes matacus*), disponíveis na base de dados GenBank, aumentando a representatividade genômica de um clado com apenas 31 espécies viventes. As montagens genômicas foram feitas usando dados de short-reads com baixa cobertura. A ferramenta RagTag v2.1.0 foi utilizada para reduzir a fragmentação dos genomas, e melhorar a predição de proteínas expressas e genes ortólogos. Para a análise de seleção, as proteínas expressas foram preditas com a ferramenta GeMoMa v1.9, os genes ortólogos de cópia única foram identificados utilizando OrthoFinder v2.5.5, os alinhamentos múltiplos de códons foram realizados com PRANK v.170427, as árvores filogenéticas foram construídas com IQ-Tree2 v2.1.2, e a análise de seleção positiva foi conduzida com HyPhy v2.5.8, usando a função BUSTED. Os genes candidatos identificados como estando sob seleção positiva foram validados com o teste de saturação de substituições de Xia no software DAMBE v7.3.32 e anotados funcionalmente com a ferramenta emapper v2.1.12 do pacote eggNOG v5.0.2. Foram identificados 15 genes sob seleção positiva, incluindo o gene codificador da queratina (KRT10), que podem ter conferido vantagens evolutivas aos xenartros. A análise de heterozigossidade autossômica mostrou variações significativas entre as espécies. O tamanduá-bandeira (*Myrmecophaga tridactyla*) e o tatu-canastra (*P. maximus*), classificados como "Vulneráveis" pela IUCN, apresentaram baixa heterozigossidade, indicando diversidade genética reduzida. Em contraste, o mataco (*T. matacus*), apesar de classificado como "Quase Ameaçado", exibiu alta heterozigossidade, e a preguiça-real (*C.s didactylus*) e preguiça-de-hoffmann (*C. hoffmanni*) – apesar de classificadas como "Pouco Preocupantes" – apresentaram heterozigossidade reduzida. Este estudo contribui com dados genômicos relevantes para a melhor compreensão de aspectos evolutivos e conservacionistas destas espécies.

Abstract: This study investigated the genetic diversity and genes under positive selection in xenarthrans using genomes and sequencing data currently available in the public database GenBank (NCBI), focusing on *Euphractus sexcinctus* (six-banded armadillo) and *Priodontes maximus* (giant armadillo), whose genome assemblies were produced by the author. For comparative genomics analyses, the novel assemblies of *E. sexcinctus* and *P. maximus* were added to the nine existing xenarthran genomes in the GenBank database (*Bradypus variegatus*, *Cabassous unicinctus*, *Chaetophractus vellerosus*, *Choloepus didactylus*, *Choloepus hoffmanni*, *Dasyopus novemcinctus*, *Myrmecophaga tridactyla*, *Tamandua tetradactyla*, and *Tolypeutes matacus*), increasing the genomic representation of a clade with only 31 living species and a rich evolutionary history. The genomic characterization revealed challenges in genome assembly, especially for *P. maximus*, whose low coverage and high fragmentation limited gene content completeness and, consequently, the prediction of expressed proteins. To mitigate this effect, the RagTag v2.1.0 tool was used to reduce fragmentation in several of the assemblies, significantly increasing the prediction of expressed proteins and orthologous genes. For selection analysis, expressed proteins were predicted using the GeMoMa v1.9 tool, single-copy orthologous genes were identified using OrthoFinder v2.5.5, multiple codon alignments were performed with PRANK v.170427, phylogenetic trees were constructed with IQ-Tree2 v2.1.2, and positive selection analysis was conducted with HyPhy v2.5.8, using the BUSTED function. Candidate genes identified as being under positive selection were validated with Xia's substitution saturation test in the DAMBE v7.3.32 software and functionally annotated with the emapper v2.1.12 tool from the eggNOG v5.0.2 package. This led to the identification of 15 genes under positive selection, including keratin (KRT10), which may have conferred evolutionary advantages to xenarthrans. The autosomal heterozygosity analysis showed significant variations among the species. The giant anteater (*Myrmecophaga tridactyla*) and the giant armadillo (*Priodontes maximus*), classified as "Vulnerable" by the IUCN, exhibited low heterozygosity, indicating reduced genetic diversity. In contrast, the southern three-banded armadillo (*Tolypeutes matacus*), although classified as "Near Threatened", displayed high heterozygosity, and the Linnaeus's two-toed sloth (*Choloepus didactylus*) and Hoffmann's two-toed sloth (*Choloepus hoffmanni*) – despite being both classified as "Least Concern" – exhibited reduced heterozygosity. The analyses provided detailed data on the genetic diversity and evolution of xenarthrans, contributing to the understanding of these species and their conservation needs.

1. Introdução

1.1. Xenarthra

O grupo Xenarthra (do grego *xenos*, “estranho”; e *arthros*, “articulação”), assim designado devido à peculiaridade das articulações vertebrais exibidas por todas as espécies compreendidas, é atualmente considerado uma superordem que contém duas ordens monofiléticas: Cingulata, cujos representantes vivos incluem 21 espécies de tatus; e Pilosa, que se subdivide em dois clados, Vermilingua e Folivora, cujos representantes atuais são, respectivamente, quatro espécies de tamanduás e seis espécies de preguiças (SUPERINA & LOUGHRY, 2015). Os Xenartros representam um dos quatro clados basais dos placentários, sendo o único grupo que se originou na América do Sul. Seus registros fósseis compreendem centenas de espécies já extintas, ressaltando a importância das iniciativas de estudo e conservação das espécies atuais, uma vez que a história evolutiva de um vasto conjunto de mamíferos placentários está contida em 31 espécies na atualidade – um número relativamente pequeno de representantes vivos (DELSUC et al., 2001).

Uma das características mais notáveis dos Xenartros diz respeito à sua dentição. Os dentes nesse grupo são geralmente reduzidos em tamanho, complexidade e número, sendo separados por diastemas curtos e compostos por osteodentina, sem esmalte (VIZCAÍNO, 2009). Por esse motivo, os Xenartros eram antigamente classificados como membros da ordem Edentata (significando “sem dentes”), embora essa classificação tenha sido revista na década de 1980, apontando uma série de evidências que demonstraram que as semelhanças entre os grupos são devido à convergência, e não a uma ancestralidade comum (SUPERINA & LOUGHRY, 2015).

Além dos Xenartros serem um grupo relativamente pouco estudado, há poucos dados de pesquisas científicas publicadas abordando temas relacionados à sua conservação. Muitas de suas espécies são classificadas como Vulneráveis ou Quase Ameaçadas (IUCN, 2024), apresentando populações em declínio. A perda de habitat e a caça representam ameaças significativas, tornando a pesquisa de campo essencial para avaliar seu impacto, e publicar resultados nos idiomas falados localmente é vital para aplicar o conhecimento científico em políticas de conservação e aumentar a conscientização pública sobre sua importância. (SUPERINA et al., 2013).

1.2. *Euphractus sexcincus*, o tatu-peba

A espécie *Euphractus sexcinctus*, conhecida como tatu-peba ou tatu-peludo, é endêmica da América do Sul, presente em todos os biomas brasileiros. Habita principalmente áreas de vegetação aberta e bordas de florestas, sendo comum em regiões do Cerrado e do Pantanal. No Cerrado paulista é encontrado em fitofisionomias de Campo Sujo, enquanto no Pantanal da Nhecolândia, uma das oito sub-regiões do Complexo do Pantanal, localizada no Pantanal Sul, prefere áreas mais fechadas. Os adultos podem medir mais de 40 cm e pesar mais de 6 kg, apresentando uma carapaça pardo-amarelada, com 6 a 8 cintas móveis, coberta de pêlos esbranquiçados. Esta espécie é carnívoro-onívora, alimentando-se de pequenos invertebrados e vertebrados, material vegetal e carniça. A reprodução ocorre ao longo de todo o ano, com um período de gestação de 60 a 64 dias, resultando em ninhadas de um a três filhotes que atingem a maturidade aos nove meses. A atividade pode ser diurna ou noturna, variando conforme o habitat em que ocorre (FISCHER, 2016; JABER et al., 2024).

Os tatus-pebas são caracterizados por possuírem temperaturas corporais baixas e taxas metabólicas basais reduzidas, sendo considerados homeotérmicos imperfeitos com capacidade termorregulatória limitada. Eles apresentam hábitos fossoriais adaptados para cavar e viver em tocas, o que os torna tolerantes a longos períodos de apneia e hipóxia (GASPAROTTO et al., 2017). Eles constroem tocas com uma única entrada em forma de U invertido, e as placas dérmicas ossificadas que cobrem a cabeça, flancos, costas e cauda são uma característica distintiva que proporciona proteção. A espécie possui uma ampla distribuição geográfica na América do Sul, a leste dos Andes, abrangendo países como Brasil, Suriname, Bolívia, Paraguai, Uruguai e norte da Argentina (JABER et al., 2024).

Em termos de conservação, embora a União Internacional para a Conservação da Natureza (IUCN) classifique o tatu-peba como “Pouco Preocupante” (*Least Concern*, LC) devido à sua ampla distribuição e grande população, as atividades humanas têm causado declínios populacionais em áreas específicas (JABER et al., 2024). Uma estratégia para a conservação da espécie é a criopreservação de sêmen. Ainda que estudos tenham demonstrado que as mitocôndrias de seus espermatozoides sejam altamente sensíveis ao protocolo de congelamento, a criopreservação pode ser empregada como uma estratégia para fornecer um reservatório genético de longo prazo, valioso em programas de reprodução assistida de espécies selvagens (SOUSA et al., 2016).

1.3. *Priodontes maximus*, o tatu-canastra

O tatu-canastra, *Priodontes maximus*, é o maior tatu existente da atualidade, pesando em média 30 kg, e podendo alcançar até 80 kg em cativeiro; sua carapaça é escura, com uma faixa clara na parte inferior, e cobre parcialmente os lados do corpo, impedindo-o de se enrolar completamente (CARTER et al., 2015). Alcançando até 150 cm de comprimento (incluindo a cauda), o tatu-canastra possui garras grandes em forma de cimitarra, sendo a terceira garra a mais desenvolvida, podendo atingir mais de 15 cm (DESBIEZ & KLUYBER, 2013). Sua dieta consiste principalmente de cupins e formigas, mas o tatu-canastra também consome outros invertebrados e, ocasionalmente, pequenos vertebrados e frutos; suas fezes são secas e firmes, compostas majoritariamente por formigas e cupins (CARTER et al., 2015).

Esta espécie é conhecida por ser rara e de difícil observação, com uma ampla distribuição geográfica que abrange Argentina, Bolívia, Brasil, Colômbia, Equador, Guiana Francesa, Guiana, Paraguai, Peru, Suriname e Venezuela (PITMAN et al., 2004). Um estudo conduzido no Pantanal mostrou que os tatus-canastra preferem habitats de floresta ripária e constroem tocas com uma largura média de $42,0 \pm 5,9$ cm e altura de $35,0 \pm 5,9$ cm em terrenos inclinados, reutilizando essas tocas frequentemente. São predominantemente noturnos, com atividades registradas principalmente entre as 22:00 e 00:00 horas, e sua densidade populacional foi estimada em 5,8 animais por 100 km² (AYA-CUERO et al., 2017).

Os tatus-canastra alteram significativamente seu ambiente através de suas escavações, criando habitats que influenciam recursos para pelo menos 24 outras espécies de vertebrados no Pantanal brasileiro (DESBIEZ & KLUYBER, 2013). Um outro estudo, realizado na Reserva Cisalpina, no leste do estado de Mato Grosso do Sul, registrou que 35 espécies de vertebrados utilizam as tocas do tatu-canastra como abrigo térmico, proteção contra predadores, e locais de alimentação e descanso (FÁVERO MASSOCATO & DESBIEZ, 2017).

Os tatus-canastra estão classificados como “Vulneráveis” pela IUCN e estão ameaçados principalmente pela perda de habitat e caça; no Pantanal, trabalhadores rurais matam esses animais por acreditarem que trazem má sorte (DESBIEZ & KLUYBER, 2013; AYA-CUERO et al., 2017). A diminuição de suas populações ou extinção local da espécie pode prejudicar as comunidades e diversidade do habitat, afetando diretamente as espécies que utilizam suas escavações; sendo assim, é de fundamental importância implementar medidas de conservação para proteger os tatus-canastra e seu papel ecológico, garantindo a manutenção da heterogeneidade do habitat e da biodiversidade nas áreas onde ocorrem (DI BLANCO et al., 2020; FONTES et al., 2020).

1.4. Heterozigossidade Autossômica e Análises de Seleção

A genômica e os estudos filogenômicos são de grande importância para a conservação da biodiversidade, pois fornecem dados detalhados que aprimoram a identificação de espécies, a compreensão da mistura genética, a detecção de alelos adaptativos, e a facilitação do resgate evolutivo. Essas abordagens dão suporte ao desenvolvimento de ações que visam mitigar ameaças às espécies relacionadas com a degradação do habitat e mudanças climáticas, aumentando a precisão na definição de unidades de conservação e possibilitando estratégias de manejo mais eficazes e fundamentadas (SUPPLE & SHAPIRO, 2018).

A variabilidade genética dentro das populações está diretamente ligada ao potencial adaptativo e é essencial para decisões de gestão genética, que visam reduzir a endogamia e aumentar a variação genética. Em espécies ameaçadas, os estudos populacionais podem contribuir com informações valiosas sobre os níveis de diversidade genética que, quando altos, podem evitar a depressão endogâmica e garantir a sobrevivência a longo prazo (SCHMIDT et al., 2020).

A descrição de genomas completos tem possibilitado acessar o nível de heterozigossidade autossômica de uma espécie a partir da análise de dados em larga escala, oferecendo uma medida mais confiável e consistente em comparação com as análises baseadas em conjuntos mais limitados de polimorfismos de nucleotídeos únicos – ou SNPs (TSARTSIANIDOU et al., 2021; SCHMIDT et al., 2020).

A manutenção de altos níveis de heterozigossidade impede que alelos recessivos prejudiciais se manifestem em homozigose. Além disso, os heterozigotos mantêm a diversidade de alelos que pode ser crucial para a adaptação a novas pressões ambientais e desafios evolutivos (VAN DER VALK et al., 2024). Populações com alta heterozigossidade são mais robustas e capazes de sobreviver a pressões seletivas, enquanto populações com baixa heterozigossidade são mais vulneráveis a ameaças como doenças, mudanças climáticas e perda de habitat. A análise da heterozigossidade permite identificar populações e espécies em risco devido à endogamia, que resulta em alelos deletérios e redução na capacidade adaptativa (TSARTSIANIDOU et al., 2021).

O estudo da heterozigossidade fornece uma base sólida para estratégias de conservação genética, ajudando a identificar e preservar a variabilidade genética necessária para a manutenção da saúde e sustentabilidade das populações, sendo útil para auxiliar intervenções, visando a introdução de novos indivíduos ou a proteção de habitats geneticamente mais diversos (VAN DER VALK et al., 2024).

Além das estimativas de diversidade pela heterozigosidade, as análises que buscam sinais de seleção positiva são importantes porque revelam como os organismos se adaptam aos seus ambientes, identificando mudanças genéticas que conferem características vantajosas às espécies. Esse conhecimento é crucial para compreender aspectos relacionados com as pressões evolutivas que moldaram a diversidade e complexidade genética ao longo do tempo, e os mecanismos que impulsionam a diversificação e evolução das espécies (MURRELL et al., 2015).

A análise de seleção positiva envolve a detecção de regiões em genes que evoluíram sob seleção positiva, ou seja, mutações nessas regiões que foram benéficas e favorecidas pela seleção natural. O processo de busca por esses sinais começa a partir do alinhamento de sequências do genoma de diferentes espécies ou populações focais para identificar posições homólogas. Em seguida, é construída uma árvore filogenética para representar as relações evolutivas entre essas sequências. Finalmente, utilizam-se modelos estatísticos de substituição de códons para diferenciar mutações sinônimas, que não alteram a proteína, de mutações não sinônimas, que as alteram. A razão entre substituições não-sinônimas e sinônimas (dN/dS ou ω) é uma métrica chave: uma razão maior que um sugere seleção positiva, menor que um indica seleção purificadora, e igual a um implica evolução neutra (MURRELL et al., 2015).

Métodos como o BUSTED (*Branch-Site Unrestricted Statistical Test for Episodic Diversification*), podem identificar a seleção positiva em um gene sem a necessidade de que sejam previamente especificados quais ramos da árvore filogenética estão sob seleção, e assim modelar a seleção como variável entre sítios e ramos, fornecendo Razões de Evidência (*Evidence Ratios*) para identificar sítios sob seleção positiva episódica. Esse método ajuda a detectar seleção positiva mesmo quando ela é fraca ou “suave”, ou seja, distribuída por vários sítios (BOOKER et al., 2017).

Através da identificação de genes sob seleção positiva, é possível descobrir os mecanismos genéticos que permitem a tolerância a estressores ambientais, como temperaturas extremas, doenças e mudanças de habitat. Além disso, a análise de seleção positiva pode revelar processos biológicos fundamentais e vias metabólicas essenciais para a sobrevivência e reprodução dos organismos. Isso pode levar ao desenvolvimento de estratégias de melhoramento e conservação, focando em características genéticas que conferem maior resistência e adaptabilidade (HUANG et al., 2022).

2. Justificativas e Objetivos

Uma característica marcante dos xenartros é a regressão dentária, já estudada por meio da análise de seleção de genes dentários, em conjunto com a identificação de mutações inativadoras compartilhadas (SIMs) e estimativas de pressão seletiva usando modelos de razão dN/dS ; tais estudos revelaram padrões distintos de pseudogenização (processo pelo qual um gene funcional sofre mutações que resultam em sua inativação, transformando-o em um pseudogene). Nos tamanduás e preguiças, houve evidência de perda completa do esmalte dentário no ancestral comum, sugerida pela inativação dos genes AMELX, AMTN, MMP20 e ENAM. Enquanto os tamanduás continuaram a perder genes como DSPP e ODAPH, resultando na ausência de dentes, essa regressão foi interrompida nas preguiças, possivelmente devido à adaptação a uma dieta herbívora. Nos tatus, genes associados à junção gengival-dente, como AMTN e ODAM, foram inativados antes da perda do esmalte. Esses achados destacam a complexa história evolutiva dos xenartros e sublinham a importância da análise de seleção na compreensão dos processos evolutivos e das adaptações funcionais em diferentes linhagens (EMERLING et al., 2022). Além da relevância dos estudos de seleção, acessar e comparar os níveis de heterozigidade autossômica das diferentes espécies de Xenartros pode gerar *insights* sobre suas histórias demográficas e evolutivas e padrões de diversificação, incluindo questões relativas aos seus status de ameaça.

Considerando o contexto apresentado, o presente trabalho tem como objetivo geral realizar análises genômicas nessas espécies visando compreender aspectos relacionados à evolução e diversificação do grupo, bem como gerar insights relevantes à sua conservação. Para isso, pretende-se alcançar os seguintes objetivos específicos de trabalho:

- 2.1. Caracterizar os genomas de *Euphractus sexcinctus* e *Priodontes maximus*, visando identificar genes e seus respectivos produtos proteicos;
- 2.2. Realizar análises de filogenômica comparativa em espécies da superordem Xenarthra cujos genomas estão publicamente disponíveis, incluindo a identificação dos genes ortólogos, com a finalidade de investigar a existência de sinais de seleção positiva;
- 2.3. Estimar a diversidade genômica para espécies de xenartros com dados de sequenciamento e montagens genômicas de alta qualidade disponíveis, medindo as heterozigosidades autossômicas e levando em consideração as diferenças relacionadas aos seus respectivos status de conservação.

3. Material e Métodos

3.1. Amostragem e Sequenciamento

As coletas das amostras biológicas foram realizadas seguindo todas as recomendações propostas pela legislação brasileira, mediante autorizações do Sistema de Autorização e Informação em Biodiversidade do Instituto Chico Mendes de Conservação da Biodiversidade (SISBIO/ICMBIO 53798-7), do Sistema Nacional de Gestão do Patrimônio Genético e do Conhecimento Tradicional Associado do Ministério do Meio Ambiente (SISGEN/MMA A05D558), e da Comissão de Ética no Uso de Animais da Universidade Federal de São Carlos (CEUA/UFSCAR 3597261118).

Para o sequenciamento do genoma de *E. sexinctus* foi utilizada uma amostra de tecido de orelha de uma fêmea amostrada em 2013, proveniente do Pantanal do Mato Grosso do Sul (19° 15' 39.785" S, 55° 46' 52.588" O). O tecido foi armazenado em etanol absoluto e o DNA foi extraído usando o método Salino modificado de Aljanabi e Martinez (1997); em seguida, foi quantificado por fluorimetria no Qubit® (Invitrogen), com o uso do Qubit® dsDNA HS Assay Kit. Para o sequenciamento do genoma de *P. maximus*, foi usada uma amostra de tecido de orelha de um indivíduo subadulto macho, coletada em 2015 no Pantanal do Mato Grosso do Sul, próximo ao Hotel Fazenda Baía das Pedras (19° 10' 18.5" S, 55° 40' 46.4" O), e fixada em etanol absoluto. O DNA desse espécime foi extraído usando o protocolo de Fenol-Clorofórmio (Sambrook & Russell, 2001) e quantificado por espectrofotometria com Nanodrop da Thermo Fisher Scientific (Waltham, MA, EUA).

Para construção das bibliotecas genômicas (*shotgun*) e sequenciamento next-gen, alíquotas de DNA de ambas as espécies foram utilizadas em protocolos padrão, usando-se o kit Illumina Nextera DNA Library (Illumina, San Diego, EUA), e plataforma Illumina (Illumina, CA, USA) para sequenciamento de ambas as extremidades (*paired-end*).

Para *E. sexinctus*, foi corrida uma lane em plataforma Hi-Seq 2000, tendo sido gerados fragmentos (*short reads*) com 100 pb. Para *P. maximus*, foi corrida meia lane na célula de fluxo em plataforma Hi-Seq 2500, tendo sido gerados fragmentos (*short reads*) com 126 pb. O sequenciamento de ambas as espécies foi feito no Centro de Genômica Funcional da Escola Superior de Agricultura Luiz de Queiroz, da Universidade de São Paulo, ESALQ-USP.

3.2. Montagem, Análise e Melhoramento dos Genomas

Para caracterizar o genoma das espécies *E. sexinctus* e *P. maximus*, os conjuntos totais de *short reads* sequenciados para cada espécie foram inicialmente analisados com o *software* FastQC v0.10.1 (Andrews, 2010). A ferramenta Trimmomatic v0.36 (Bolger, 2014) foi usada para a triagem dos *shorts reads*, descartando-se os de baixa qualidade; o SPAdes v3.10.0 (Bankevich, 2012) foi utilizado para a montagem *de novo* e obtenção dos *scaffolds* finais; as ferramentas QUAST v3.2 (Gurevich, 2013) e BBMap v38.26 (Bushnell, 2014) foram utilizadas para calcular a profundidades de cobertura e as estatísticas de contiguidade, como N50 e L50. Essas operações foram realizadas no servidor do Laboratório Genômica Integrativa (LGI) do Instituto de Biociências (IB) da Unesp, Campus de Botucatu. O acesso ao servidor e o monitoramento de processos foram realizados remotamente, utilizando o PuTTY (TATHAM, 2000). O acesso ao servidor para transferência de arquivos também foi realizado remotamente, utilizando o WinSCP (PRIKRYL, 2000).

Para melhorar a contiguidade das montagens genômicas do tatu-peba e o tatu-canastra, utilizamos a ferramenta RagTag v2.1.0 (ALONGE et al., 2022), com a opção *scaffold*, que alinha os *contigs* a um genoma de referência, ordenando e orientando esses *contigs* conforme sua posição presumida, estimando o tamanho das lacunas entre eles e produzindo uma montagem menos fragmentada. O genoma de referência utilizado foi o do tatu-galinha, *Dasytus novemcinctus* (NCBI, GenBank GCA_030445035.1, RefSeq GCF_030445035.1), que está resolvido ao nível cromossômico, tendo sido extensivamente anotado. As ferramentas QUAST v3.2 e BBMap v38.26 foram utilizadas novamente para avaliar os resultados. As operações para melhorar a contiguidade das montagens foram realizadas no MCC (*Morgan Compute Cluster*) da University of Kentucky.

3.3. Análise de Seleção Positiva em Xenartros

Para as análises de seleção positiva, além dos dados genômicos do tatu-peba e do tatu-canastra, utilizamos dados genômicos disponíveis para nove outras espécies de xenartros, a saber: *Bradypus variegatus*, *Cabassous unicinctus*, *Chaetophractus vellerosus*, *Choloepus didactylus*, *Choloepus hoffmanni*, *Dasytus novemcinctus*, *Myrmecophaga tridactyla*, *Tamandua tetradactyla*, e *Tolypeutes matacus*. As montagens genômicas para estas espécies, em formato FASTA de nucleotídeos (.FNA), descritas na **Tabela 1**, foram baixadas do GenBank. Na **Tabela 2**, estão descritas as características das montagens (fragmentação, cobertura, tamanho, e nível das montagens).

Tabela 1: Espécies de xenartros cujas montagens genômicas estavam disponíveis no NCBI até a realização do presente trabalho (último acesso em 6 de fevereiro de 2024).

Xenartro	Nome comum	GenBank
<i>Bradypus variegatus</i>	Preguiça-comum	GCA_004027775.1
<i>Cabassous unicinctus</i>	Tatu-de-rabo-mole-pequeno	GCA_029593785.1
<i>Chaetophractus vellerosus</i>	Tatu-peludo	GCA_004027955.1
<i>Choloepus didactylus</i> *	Preguiça-real	GCA_015220235.1
<i>Choloepus hoffmanni</i>	Preguiça-de-Hoffmann	GCA_000164785.2
<i>Dasyopus novemcinctus</i> *	Tatu-galinha	GCA_030445035.1
<i>Myrmecophaga tridactyla</i>	Tamanduá-bandeira	GCA_004026745.1
<i>Tamandua tetradactyla</i>	Tamanduá-mirim	GCA_023851605.1
<i>Tolypeutes matacus</i>	Mataco	GCA_026826555.1

* Genomas de referência com dados de proteínas anotadas disponíveis.

Tabela 2: Informações sobre os genomas de xenartros utilizados neste trabalho. Para *E. sexinctus* e *P. maximus*, os valores de N50, L50, e os tamanhos dos genomas foram estimados com a ferramenta QUAST v3.2, enquanto as coberturas foram estimadas com a ferramenta BBMap v38.26. Para as demais espécies, esses valores foram obtidos de suas respectivas páginas no GenBank.

Xenartros (NCBI)	Tecnologia de Sequenciamento	N50 (kb)	L50	Cobertura	Genoma (Gb)	Nível da Montagem
<i>B. variegatus</i>	Illumina HiSeq	1,9	138079	57,9x	1,5	<i>scaffolds</i>
<i>C. unicinctus</i>	Illumina HiSeq	7,4	152746	60,0x	6,1	<i>scaffolds</i>
<i>C. vellerosus</i>	Illumina HiSeq	1,6	572611	22,9x	5,3	<i>scaffolds</i>
<i>C. didactylus</i>	PacBio e outras *	~146200	9	56,8x	3,2	cromossomos
<i>C. hoffmanni</i>	Sanger + Illumina	366,4	2423	65,0x	3,3	<i>scaffolds</i>
<i>D. novemcinctus</i>	PacBio e outras **	~130100	11	17,0x	3,6	cromossomos
<i>M. tridactyla</i>	Illumina HiSeq	41,3	22180	28,9x	3,5	<i>scaffolds</i>
<i>T. tetradactyla</i>	PacBio e outras *	~174600	8	66,5x	3,2	cromossomos
<i>T. matacus</i>	Illumina HiSeq	~12600	68	70,0x	3,8	<i>scaffolds</i>
Xenartros que são o foco deste trabalho:						
<i>E. sexinctus</i>	Illumina HiSeq	4,4	173582	14,7x	2,6	<i>scaffolds</i>
<i>P. maximus</i>	Illumina HiSeq	1,1	417254	9,6x	1,4	<i>scaffolds</i>

* PacBio Sequel I CLR; Illumina NovaSeq; Arima Genomics Hi-C; Bionano Genomics DLS.

** PacBio Sequel II HiFi; Arima Hi-C v2; Bionano Genomics DLS.

Para a determinação dos genes ortólogos de cópia única, dados em formato FASTA de aminoácidos (.FNA) para cada espécie foram convertidos em arquivos FASTA de aminoácidos (.FAA) contendo as proteínas expressas para cada genoma. Para tal foi necessário realizar uma etapa de predição para produzir arquivos .FAA para nove espécies, com exceção da preguiça-real, *C. didactylus* (NCBI, GenBank GCA_015220235.1, RefSeq GCF_015220235.1); e o tatu-galinha, *D. novemcinctus* (NCBI, GenBank GCA_030445035.1, RefSeq GCF_030445035.1), cujos genomas já estavam anotados e resolvidos ao nível cromossômico. Para predição proteica, usamos os genomas (.FNA) de referências dessas duas espécies, com seus respectivos arquivos de anotação (.GFF), e a ferramenta GeMoMa v1.9 (*Genome-based Gene Model Prediction*) (KEILWAGEN et al., 2018), uma vez que esta prediz modelos de genes em genomas de

espécies-alvo usando informações de proteínas de espécies de referência. Deste modo, utilizamos essa ferramenta para prever as proteínas expressas em *E. sexinctus* e *P. maximus*, bem como para sete espécies cujos genomas não estavam anotados (**Tabela 1**).

Como apenas um único genoma-referência deve ser empregado no processo de predição de proteínas com a ferramenta GeMoMa, a escolha do genoma-referência (*C. didactylus* ou *D. novemcinctus*) dependeu da proximidade filogenética com a espécie-alvo: Para espécies de tatus (*E. sexinctus*, *P. maximus*, *C. uncinctus*, *T. matacus*, *C. vellerosus*), foi utilizado o genoma do tatu-galinha, *D. novemcinctus*. Para preguiças (*B. variegatus*, *C. hoffmanni*) e tamanduás (*M. tridactyla*, *T. tetradactyla*), foi utilizado o genoma da preguiça-real, *C. didactylus*. Com isso, obtivemos nove arquivos em formato FASTA de aminoácidos (.FAA). Para completar o conjunto de arquivos com as proteínas expressas nas onze espécies, os proteomas de *C. didactylus* e *D. novemcinctus*, em formato FASTA de aminoácidos (.FAA), foram também baixados do NCBI.

Para avaliar a qualidade (completude) das montagens genômicas, a ferramenta BUSCO v5.4.6 (Benchmarking Universal Single-Copy Orthologs) (MANNI et al., 2021) foi utilizada; esta ferramenta fornece uma medida da completude com base em conjuntos de genes ortólogos de cópia única esperados para um determinado grupo taxonômico. No presente trabalho, utilizamos o conjunto mammalia_odb10 (genes ortólogos de mamíferos, criado em 8 de janeiro de 2024) para avaliar todas as 11 montagens genômicas. Os genes que fazem parte dos conjuntos BUSCO para um determinado grupo, como “mamíferos”, são selecionados de conjuntos ortólogos presentes em pelo menos 90% desse grupo.

Empregamos a ferramenta RagTag v2.1.0 para aprimorar a contiguidade das montagens das espécies que apresentaram baixa detecção de BUSCOs (<70%), utilizando genomas de alta qualidade (resolvidos ao nível cromossômico) de espécies filogeneticamente próximas como referência: para *B. variegatus*, utilizamos *C. didactylus* como referência; para *C. uncinctus* e *C. vellerosus*, utilizamos *D. novemcinctus*; para *M. tridactyla*, utilizamos *T. tetradactyla*. Em seguida, utilizamos novamente a ferramenta GeMoMa v1.9 para gerar arquivos atualizados (.FAA) com as proteínas expressas para cada espécie.

Para a identificação de conjuntos de genes ortólogos de cópia única, utilizamos a ferramenta OrthoFinder v2.5.5 (EMMS & KELLY, 2019), tendo como parâmetro de entrada os arquivos de proteínas preditas (.FAA). A partir desse ponto, a análise se dividiu em dois grupos: O primeiro, com todas as 11 espécies de xenartros já mencionadas; e um segundo, formado do mesmo grupo, porém excluindo-se *P. maximus* e, portanto, com 10 espécies. Essa decisão foi tomada com base no fato de que, mesmo após o aprimoramento do genoma de *P.*

maximus com a ferramenta RagTag, a análise de completude com a ferramenta BUSCO ainda revelou uma porcentagem baixa (26%) de BUSCOs completos. Como isso poderia comprometer nossos resultados subsequentes, escolhemos repetir todas as análises para um subconjunto das 11 espécies do qual somente o tatu-canastra foi excluído. A partir de então, para cada grupo (um com 11 espécies, e outro com 10 – excluindo-se o *P. maximus*), um respectivo conjunto de ortólogos de cópia única foi identificado.

Para realizar a etapa da análise de seleção, os arquivos em formato FASTA de aminoácidos (.FAA), gerados com a ferramenta OrthoFinder v2.5.5, contendo as proteínas expressas por genes ortólogos correspondentes nas espécies-alvo, foram convertidos, cada um, em formato FASTA de nucleotídeos (.FNA), num alinhamento múltiplo de códons. Para isso, foi necessário escrever uma série de *scripts* em Python, utilizando os arquivos de anotação (.GFF) gerados pela ferramenta GeMoMa v1.9 ou baixados do GenBank como guias para resgatar, nos arquivos (.FNA) de cada montagem genômica, os devidos *exons* em cada gene, concatenando-os de maneira apropriada e de forma a reconstruir os arquivos produzidos pelo OrthoFinder v2.5.5 em formato FASTA de nucleotídeos (.FNA).

Em seguida, um *script* foi desenvolvido e utilizado para substituir os cabeçalhos das sequências de nucleotídeos por *tags* identificando a espécie de origem de cada sequência. Posteriormente, outro *script* foi usado para remover todos os *stop codons* (TAA, TAG, TGA) – um requerimento do *software* usado na análise de seleção, mais adiante. Em seguida, utilizamos a ferramenta PRANK v.170427 (LÖYTYNOJA, 2013) com a opção *-codon* para realizar o alinhamento múltiplo das sequências em cada arquivo. Como o processo de alinhamento com a ferramenta PRANK resulta em arquivos de saída cujas sequências FASTA estão desordenadas em relação aos arquivos de entrada, um *script* adicional e que usa os *tags* de espécies, introduzidos na etapa anterior ao alinhamento, foi empregado para reordenar as sequências de nucleotídeos e restaurar o arranjo original, facilitando as análises subsequentes.

Antes de finalmente proceder para a etapa de análise de seleção, utilizamos os alinhamentos múltiplos de códons para gerar, para cada grupo de espécies, uma árvore filogenética – para ser utilizada como dado de entrada na etapa seguinte. Para isso, um *script* em Python foi usado para concatenar os alinhamentos de códons de todos os arquivos referentes aos genes ortólogos em um único arquivo, produzindo, assim, uma longa sequência de nucleotídeos para cada espécie. Esse arquivo foi filtrado com a ferramenta TrimAl v1.2rev59 (CAPELLA-GUTIÉRREZ et al., 2009), que identifica e remove automaticamente regiões alinhadas que não são confiáveis, baseando-se em vários critérios, como quantidade de *gaps*, conservação de colunas e homologia de sequência. O arquivo filtrado foi então utilizado como

parâmetro de entrada para a ferramenta IQ-Tree2 v2.1.2 (MINH et al., 2020), que emprega uma abordagem sofisticada para a seleção de modelos para construir árvores filogenéticas, com o objetivo de representar de forma precisa as relações evolutivas entre as sequências. Inicialmente, o software testa uma ampla gama de modelos de substituição, que descrevem como as sequências de DNA, RNA ou proteínas mudaram ao longo do tempo (KALYAANAMOORTHY et al., 2017). Com base em critérios estatísticos (Critério de Informação de Akaike, Critério de Informação de Akaike Corrigido e Critério de Informação Bayesiano), o IQ-Tree2 seleciona o modelo que melhor se ajusta aos dados, e este modelo é então usado para inferir uma árvore filogenética, aplicando algoritmos que maximizam a verossimilhança, ou seja, a probabilidade de a árvore proposta refletir as mudanças observadas nas sequências ao longo da evolução. As árvores geradas pelo IQ-Tree2 v2.1.2 foram visualizadas com a ferramenta Interactive Tree Of Life (iTOL) v6 (LETUNIC & BORK, 2021).

Para a análise de seleção positiva, utilizamos o HyPhy v2.5.8 (KOSAKOVSKY POND et al., 2019), um *software* para análise estatística de dados de sequências genéticas baseado em hipóteses evolutivas. Usamos sua função BUSTED, que detecta eventos de seleção positiva em sítios específicos em determinados ramos de uma filogenia. Como parâmetros de entrada, utilizamos arquivos de genes ortólogos de cópia única (contendo sequências de códons alinhadas de cada espécie) e a árvore filogenética correspondente gerada pelo IQ-Tree2 v2.1.2, em formato Newick (.NWK). A saída do HyPhy BUSTED inclui testes estatísticos que indicam se há evidência significativa de seleção positiva em algum ramo da árvore, com as seguintes métricas de interesse para cada gene ortólogo de cópia única: **(a)** $\text{Log}(L)_{\text{restrito}}$, que é o logaritmo da verossimilhança do modelo restrito. Um valor de verossimilhança (*log-likelihood*) é uma medida de quão bem o modelo se ajusta aos dados e, no contexto do modelo restrito, este valor indica o ajuste do modelo que não permite seleção positiva; **(b)** $\text{Log}(L)_{\text{não-restrito}}$, que é o logaritmo da verossimilhança do modelo não-restrito. Assim como o $\text{Log}(L)_{\text{restrito}}$, este valor indica o ajuste do modelo aos dados, mas neste caso, permitindo a seleção positiva; **(c)** Proporção de Seleção Positiva, que fornece uma estimativa, em valores percentuais, de quantos sítios estão sendo favorecidos pela seleção positiva ($\omega > 1$) no modelo não-restrito; **(d)** e o P-valor, que indica a significância estatística ao comparar os modelos restrito e não-restrito. Se o P-valor for menor que um nível de significância pré definido (em nossa análise, adotamos $p < 0,05$), pode-se rejeitar a hipótese nula de que não há seleção positiva (modelo restrito) em favor da hipótese alternativa de que há seleção positiva (modelo não-restrito). Um P-valor baixo sugere que há evidências estatísticas significativas de seleção positiva em pelo menos um sítio em um ou mais ramos da árvore. Além disso, podemos utilizar $\text{Log}(L)_{\text{restrito}}$ e $\text{Log}(L)_{\text{não-restrito}}$

para calcular a Razão de Evidência (*Evidence Ratio*), dada por $ER = e^{[\text{Log}(L)_{\text{não-restrito}} - \text{Log}(L)_{\text{restrito}}]}$. A Razão de Evidência fornece uma métrica direta da força da evidência a favor do modelo não-restrito em relação ao modelo restrito, sendo que valores maiores indicam uma maior evidência a favor da seleção positiva (MURRELL et al., 2015).

Escrevemos um novo *script* em Python para coletar essas métricas, calcular a Razão de Evidência e organizar essas informações em uma planilha (.VCF) para processamento posterior. Para filtrar esses resultados e identificar genes candidatos à seleção positiva, escrevemos um outro *script* em Python que lê e filtra esta planilha, descartando os genes cujas métricas não cumpriram os seguintes critérios: Proporção de Seleção Positiva > 5%; Razão de Evidência > 10 (isto é, o modelo não-restrito, que permite seleção positiva, ajusta-se aos dados ao menos 10 vezes melhor do que o modelo restrito, que não permite seleção positiva); e P-valor < 0,05, para que a seleção positiva seja considerada presente com alto nível de confiança.

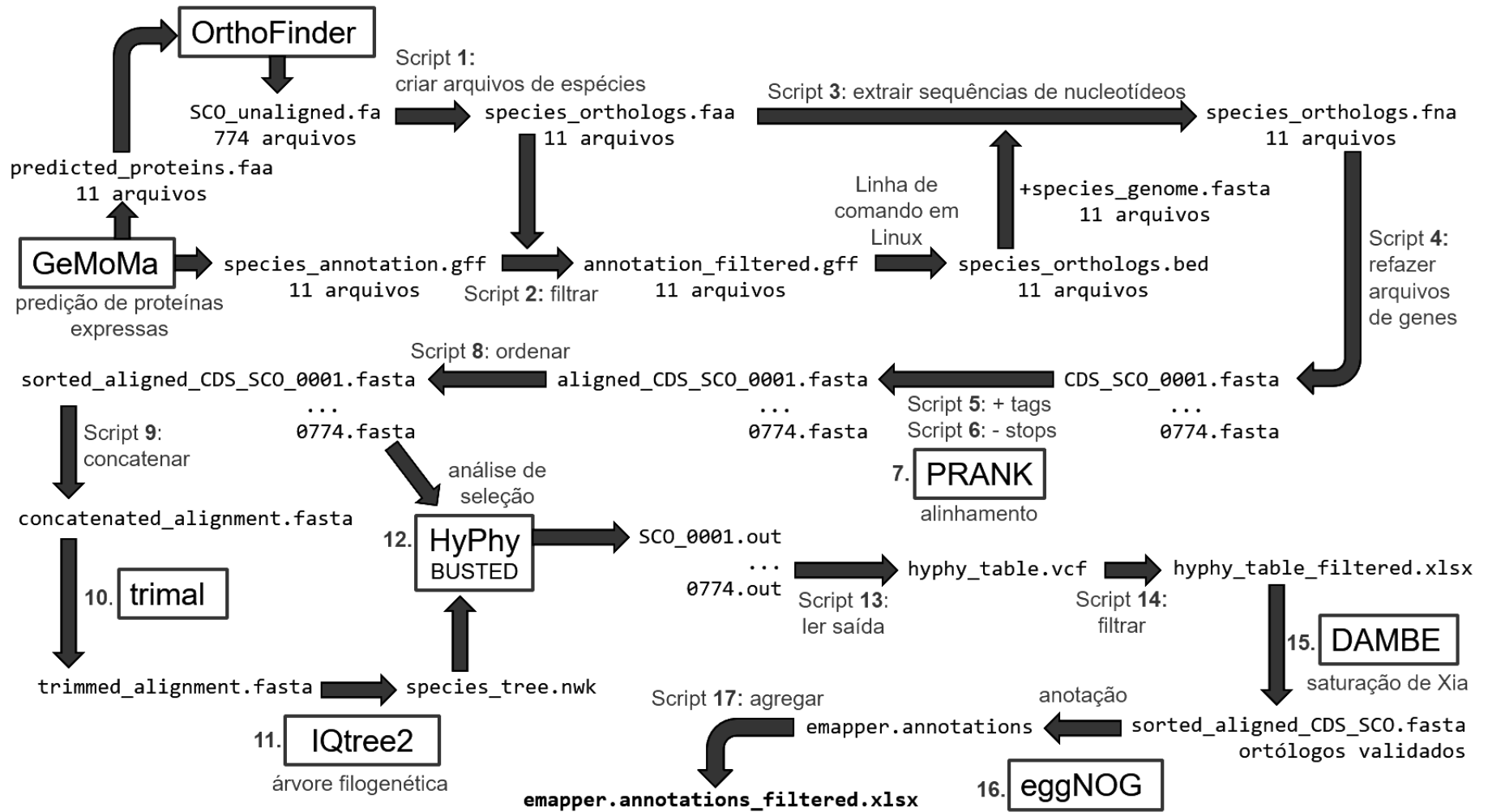
Uma vez identificados os genes candidatos, realizamos, para cada um deles, o teste de saturação de substituições de Xia (2009), com o *software* DAMBE v7.3.32 (XIA, 2018). A saturação de substituições ocorre quando múltiplas substituições acontecem no mesmo sítio ao longo do tempo, tornando difícil distinguir entre substituições recentes e antigas. Ou seja, há um “obscurecimento” do sinal evolutivo, que pode levar a erros nas inferências filogenéticas e análises de seleção. O teste de saturação de substituições de Xia, implementado no *software* DAMBE v7.3.32, avalia a adequação dos alinhamentos de sequências de DNA para análises evolutivas. Ele calcula o índice de saturação de substituições observadas (*Iss*) e o compara com um valor crítico (*Iss.c*) que varia conforme o número de sequências no alinhamento. Se *Iss* for menor que *Iss.c*, os dados não estão saturados e são adequados para análises filogenéticas; se *Iss* for maior que *Iss.c*, os dados estão saturados e podem não ser adequados para análises precisas. Este teste é essencial para garantir que múltiplas substituições não distorçam os resultados em estudos de seleção positiva e filogenética (XIA & LEMEY, 2009).

Tendo identificado os genes candidatos para seleção positiva e os tendo validado através do teste de saturação de substituições de Xia, realizamos uma etapa final de anotação funcional destes genes com a ferramenta emapper v2.1.12, do pacote eggNOG v5.0.2 (Huerta-Cepas et al., 2018). Utilizamos esta ferramenta para anotar funcionalmente as sequências de aminoácidos correspondentes aos genes sob seleção positiva, comparando-as com as sequências registradas no banco de dados eggNOG (eggNOG DB v5.0.2). Esse banco de dados contém grupos ortólogos e anotações funcionais, oferecendo identificação de ortólogos, anotações de função gênica, mapeamento para ontologias como GO (Gene Ontology), análises de conservação gênica, além de informações sobre vias metabólicas e funcionais através de mapeamentos para

KEGG pathways (Kyoto Encyclopedia of Genes and Genomes), proporcionando uma ampla gama de informações biológicas. Escrevemos um *script* em Python para congregiar os resultados da anotação e condensá-los em uma única planilha.

Todas as operações descritas nesta seção foram realizadas no MCC (Morgan Compute Cluster), da University of Kentucky. O processo completo para a identificação dos genes positivamente selecionados, bem como suas anotações, está descrito na **Figura 1**.

Figura 1: Pipeline desde a predição das proteínas expressas, passando pela identificação dos ortólogos de cópia única, conversão dos arquivos FASTA de aminoácidos em FASTA de nucleotídeos, alinhamento múltiplo dos códons, criação da árvore filogenética, análise de seleção, teste de saturação de substituições, até a anotação dos genes positivamente selecionados e agregação dos resultados. As quantidades de arquivos indicadas na figura dizem respeito às análises realizadas com todas as 11 espécies de xenartros.



3.4. Análise de Heterozigosidade Autossômica

Realizamos a análise de heterozigosidade autossômica para 10 espécies de xenartros – incluindo o tatu-peba e o tatu-canastra – utilizando, como ponto de partida, os respectivos dados de sequenciamento (Illumina *short reads*) e um genoma de alta qualidade de *D. novemcinctus* (NCBI, GCA_030445035.1) como referência. Para as espécies com genomas disponíveis previamente, os arquivos contendo *reads* (em formato FASTQ) foram baixados do SRA (*Sequence Read Archive*), um repositório de dados de sequenciamento mantido pelo *National Center for Biotechnology Information* (NCBI). Das 11 espécies abordadas neste trabalho, a espécie *C. unicinctus* não foi incluída nesta análise, pois não pudemos localizar dados de sequenciamento Illumina compatíveis com os das demais espécies disponíveis para *download* no SRA. As informações sobre as espécies abordadas e seus respectivos dados de sequenciamento estão descritas na **Tabela 3**.

Tabela 3: Informações acerca dos dados de sequenciamento a serem utilizados na análise de heterozigosidade autossômica envolvendo as 10 espécies de xenartros abordadas neste trabalho.

Xenartro	Número de Acesso (SRA)	Tecnologia de Sequenciamento	Nº de bases (G bases)
<i>Bradypus variegatus</i>	SRR14084425	Illumina HiSeq 2500	87,1
<i>Chaetophractus vellerosus</i>	SRR7704807	Illumina HiSeq 2500	64,0
<i>Choloepus didactylus</i>	SRR7637805	Illumina HiSeq 2500	50,7
<i>Choloepus hoffmanni</i>	SRR866928	Illumina HiSeq 2000	20,5
<i>Dasypus novemcinctus</i>	SRR12697336	Illumina HiSeq 2500	19,1
<i>Myrmecophaga tridactyla</i>	SRR11431893	Illumina HiSeq 2500	87,3
<i>Tamandua tetradactyla</i>	SRR7704834	Illumina HiSeq 2500	93,8
<i>Tolypeutes matacus</i>	SRR7704836	Illumina HiSeq 2500	69,8
<i>Euphractus sexinctus</i>	N/A	Illumina HiSeq 2000	39,3
<i>Priodontes maximus</i>	N/A	Illumina HiSeq 2500	21,7

Todos os arquivos de *reads* (.FASTQ) foram submetidos a um controle de qualidade. O *software* FastQC v0.10.1 (Andrews, 2010) foi utilizado para verificar a qualidade inicial dos dados sequenciados, e utilizamos a ferramenta fastp v0.23.4 (CHEN et al., 2018) para identificar e remover adaptadores de sequenciamento *paired-end* (*adapter sequences*) e

sequências sobre-representadas, bem como caudas poly-G (artefatos de sequenciamento), estabelecendo também alguns critérios de filtragem para a manutenção dos *reads*: um comprimento mínimo de 35 bases e uma qualidade média (PHRED *Score*) mínima de 20. Bases em ambas as extremidades dos *reads* foram removidas caso seus respectivos PHRED *Scores* fossem inferiores a 3, e uma janela deslizante de tamanho 4 removeu as bases dentro de si ao longo de cada *read* caso o PHRED *Score* médio ficasse abaixo de 15. A janela deslizante (*sliding window*) é um método adaptativo de controle de qualidade, onde um bloco de tamanho fixo (no caso, 4 bases) percorre o *read* de sequenciamento e remove as bases a partir de uma certa posição, caso a qualidade média das bases dentro desse bloco fique abaixo de um determinado limite (no caso, PHRED *Score* = 15). Após esse tratamento, a qualidade dos dados de sequenciamento foi re-avaliada com a ferramenta FastQC, a fim de verificar os resultados desta etapa.

Para utilizar o genoma de *D. novemcinctus* como referência no alinhamento dos *reads* e análises subsequentes, realizamos etapas preparatórias: Primeiro, utilizamos o *software* bwa-mem2 v2.2.1 (VASIMUDDIN et al., 2019), com a ferramenta *index*, para criar arquivos de índice que facilitariam o alinhamento eficiente dos *reads*, permitindo compressão e busca rápida. Em seguida, usamos o *software* samtools v1.12 (DANECEK et al., 2021), com a ferramenta *faidx*, para gerar um arquivo de índice FASTA, facilitando o acesso rápido a regiões específicas do genoma. Finalmente, usamos o *software* GATK v4.4.0.0 (MCKENNA et al., 2010), com a ferramenta *CreateSequenceDictionary*, para criar um dicionário de sequências que fornece informações sobre os *contigs* presentes no genoma, essenciais para o funcionamento de várias ferramentas do GATK em análises genômicas.

Realizamos o alinhamento, ao genoma de referência, dos *reads* contidos em cada par de arquivos de sequenciamento (*forward* e *reverse*), utilizando o *software* bwa-mem2 v2.2.1, com a ferramenta *mem*. Para cada espécie, foi gerado um arquivo em formato SAM (*Sequence Alignment/Map*), que contém informações detalhadas sobre como cada *read* se alinha ao genoma de referência, incluindo as posições de início e fim, a qualidade do alinhamento, e outras métricas. Em seguida, os arquivos .SAM foram processados com o *software* samtools v1.12, utilizando a ferramenta *view* para conversão desses arquivos em formato binário (.BAM), e a ferramenta *sort* para ordenar os arquivos .BAM de acordo com as posições dos *reads* alinhados no genoma de referência. O formato BAM é binário e comprimido, ocupando menos espaço em disco e sendo mais rápido para ler e processar devido ao acesso aleatório eficiente aos dados. Além disso, muitos programas de bioinformática e pipelines de análise aceitam ou requerem arquivos BAM por serem mais eficientes para manipulação e análise de

grandes conjuntos de dados de sequenciamento. Já o ordenamento é importante porque muitas ferramentas de análise posteriores, como as de chamada de variantes, que utilizamos mais adiante, requerem arquivos BAM ordenados pelas posições das leituras no genoma para facilitar o acesso rápido e eficiente aos dados.

Na etapa seguinte, processamos os arquivos .BAM com o *software* GATK v4.4.0.0, usando a ferramenta *MarkDuplicates* do Picard Toolkit v3.0.0 com a opção *REMOVE_DUPLICATES=true*, a fim de marcar e remover *reads* duplicados resultantes de artefatos de PCR criados durante a preparação da biblioteca de sequenciamento. Esse passo é importante para a análise de heterozigosidade porque garante que a cobertura de sequenciamento seja representativa da amostra biológica real, reduzindo vieses e artefatos que poderiam levar a falsas chamadas de variantes. A remoção de duplicatas melhora a precisão na detecção de variantes heterozigóticas, assegurando dados de alta qualidade para cálculos mais confiáveis de heterozigosidade, além de otimizar o armazenamento e o processamento computacional (BROAD INSTITUTE, 2024).

Em seguida, utilizamos a ferramenta *bam clipOverlap* do samtools v1.12 para remover as sobreposições entre *reads* pareados nos arquivos .BAM processados. Esse passo é importante para a análise de heterozigosidade, pois evita a contagem duplicada de bases nas regiões sobrepostas, o que poderia introduzir vieses e erros na chamada de variantes (DANECEK et al., 2021). Utilizamos, então, samtools v1.12 com a ferramenta *addreplacerg*, que adiciona ou substitui grupos de leitura em arquivos .BAM, incorporando metadados importantes para a identificação e rastreamento de amostras e bibliotecas de sequenciamento. Essa etapa é necessária pois muitas ferramentas de bioinformática e pipelines de análise, como as do GATK, requerem a presença de grupos de leitura para funcionar corretamente. Finalmente, utilizamos samtools v1.12 com a ferramenta *index* para gerar, para cada arquivo .BAM, um arquivo de índice (em formato .BAI), que será necessário na próxima etapa.

Posteriormente, utilizamos o *software* GATK v4.4.0.0, com a ferramenta *HaplotypeCaller*, para identificar variantes genéticas nos arquivos BAM, produzindo – para cada espécie – um arquivo em formato VCF que detalha SNPs e *indels* detectados. Esse processo é fundamental na análise de heterozigosidade, pois é nele que obtemos os dados sobre as variantes genéticas presentes nas amostras sequenciadas.

Em seguida, utilizamos novamente o *software* GATK v4.4.0.0, com a ferramenta *GenomicsDBImport*, para importar dados de variantes genéticas de múltiplas espécies para uma base de dados GenomicsDB específica para cada cromossomo – exceto o cromossomo X, que foi removido da análise por se tratar de um estudo de heterozigosidade autossômica. Este

processo envolve a criação de *workspaces* separados para cada cromossomo, armazenando as variantes genéticas das espécies analisadas e permitindo uma integração eficiente dos dados. A importância desse procedimento reside na sua capacidade de organizar grandes volumes de dados genéticos em uma estrutura escalável, facilitando análises genômicas de larga escala, como a detecção de variantes comuns e específicas entre as espécies, bem como a análise de heteroziguidade.

Posteriormente, utilizamos o *software* GATK v4.4.0.0, com a ferramenta *GenotypeGVCFs*, para gerar arquivos em formato VCF a partir dos dados armazenados nas bases de dados GenomicsDB criadas anteriormente. Para cada cromossomo, o comando *GenotypeGVCFs* usa como referência o genoma de *D. novemcinctus* (.FNA) e os dados de variantes armazenados no *workspace* GenomicsDB correspondente. O resultado de cada execução foi um arquivo VCF comprimido (.VCF.GZ) contendo as variantes para o cromossomo específico – no caso, um arquivo .VCF.GZ para cada um dos 31 cromossomos (originalmente 32, mas retiramos o cromossomo X da análise) e 507 arquivos para sequências genômicas não-localizadas (*unlocalized genomic scaffolds*) contidas no genoma de referência. Em resumo, este processo transformou os dados de variantes armazenados em GenomicsDB em arquivos VCF utilizáveis para análises subsequentes, permitindo a identificação e a catalogação de variantes genéticas presentes em todas as espécies estudadas.

Utilizamos, então, o *software* GATK v4.4.0.0, com a ferramenta *MergeVcfs*, para combinar os arquivos VCF resultantes do passo anterior em um único arquivo VCF consolidado. Essa ferramenta tomou os arquivos VCF genotipados de cada cromossomo, agrupando-os em um único arquivo VCF, preservando todas as informações de variantes genéticas. O resultado foi um arquivo VCF abrangente contendo todas as variantes genéticas detectadas, organizado de maneira a permitir uma análise eficiente dos dados genômicos.

Em seguida, utilizamos a ferramenta VCFtools v0.1.17 (DANECEK et al., 2011), com a opção *-remove_indels*, para filtrar as variantes relacionadas a inserções e deleções, restando apenas os SNPs (*Single Nucleotide Polymorphisms*) no arquivo VCF combinado, contendo dados de todas as espécies estudadas. Esta operação lê o arquivo VCF comprimido contendo todas as informações, remove as *indels*, e gera um novo arquivo VCF contendo apenas SNPs. Após esta etapa, usamos o *software* GATK v4.4.0.0, com a ferramenta *IndexFileFeature*, para indexar o arquivo VCF resultante, o que facilita o acesso rápido e eficiente às variantes durante a etapa seguinte.

Finalmente, obtivemos os valores para o cálculo da heteroziguidade utilizando novamente a ferramenta VCFtools v0.1.17, com as opções *-vcf* e *--het*. Essa operação lê o

arquivo VCF indexado contendo apenas SNPs, e produz um arquivo de saída com os resultados para cada espécie, contendo – entre outras métricas – o número observado de sítios homozigotos (Hom_{obs}), e o número total de sítios analisados ($N_{sítios}$). Com essas informações, podemos calcular a *heterozigosidade observada* (Het_{obs}), que representa a proporção de *loci* heterozigotos no genoma de cada indivíduo, utilizando a equação:

$$Het_{obs} = \frac{N_{sítios} - Hom_{obs}}{N_{sítios}} \quad (1)$$

Todas as operações descritas nesta seção foram realizadas no MCC (Morgan Compute Cluster), da University of Kentucky.

4. Resultados e Discussão

4.1 Caracterização dos genomas de tatu-peba e tatu-canastra

Na **Tabela 4** estão apresentadas métricas importantes para a caracterização das montagens genômicas do tatu-peba e do tatu-canastra, como as obtidas com a ferramenta QUAST v3.2, a análise de completude com a ferramenta BUSCO v.5.4.6, e o número de proteínas preditas com a ferramenta GeMoMa v1.9, antes e depois da desfragmentação com a ferramenta RagTag v2.1.0. Após a desfragmentação, observa-se melhorias significativas em ambos os genomas, tanto em qualidade quanto em completude. No caso do tatu-peba, o número de *contigs* foi reduzido para menos da metade (de 893416 para 341898), com um aumento no número de *contigs* maiores que 50 mil bases de 1 para 50, e o maior *contig* aumentou de 52902 para 146980839 bases. O N50 aumentou drasticamente de 4375 para 74732508 bases e o L50 reduziu de 172170 para 13, indicando uma montagem muito mais contígua. A porcentagem de BUSCOs completos subiu de 27,5% para 87,2%, enquanto os BUSCOs ausentes diminuíram de 55,1% para 8,7%. O número de proteínas preditas também aumentou de 11528 para 20594. Para o tatu-canastra, o número de *contigs* diminuiu de 1225376 para 453806, os *contigs* maiores que 50 mil bases aumentaram de 3 para 41, o maior *contig* aumentou de 52668 para 80530518 bases, e o N50 aumentou de 1133 para 36956290 bases, com o L50 reduzindo de 360264 para 15. A porcentagem de BUSCOs completos aumentou de 6,2% para 26,0%, e os BUSCOs ausentes diminuíram de 85,6% para 55,7%. O número de proteínas preditas subiu de 4392 para 13592. Essas melhorias destacam a eficácia da ferramenta RagTag em aprimorar a montagem e anotação dos genomas caracterizados aqui.

Tabela 4: Características das montagens genômicas para o tatu-peba e tatu-canastra, obtidos com a ferramenta QUAST v3.2, e análise de completude com a ferramenta BUSCO v.5.4.6, bem como número de proteínas preditas com a ferramenta GeMoMa v1.9, antes e depois da desfragmentação com a ferramenta RagTag v2.1.0 utilizando o genoma de *D. novemcinctus* como referência.

	Tatu-peba	Tatu-peba	Tatu-canastra	Tatu-canastra
	pré-RagTag	pós-RagTag	pré-RagTag	pós-RagTag
#contigs	893416	341898	1225376	453806
#contigs >50k bases	1	50	3	41
maior contig (bases)	52902	146980839	52668	80530518
N50 (bases)	4375	74732508	1133	36956290
L50	172170	13	360264	15
BUSCOs Completos	27,5%	87,2%	6,2%	26,0%
BUSCOs Ausentes	55,1%	8,7%	85,6%	55,7%
#proteínas preditas	11528	20594	4392	13592

A cobertura média, estimada com a ferramenta BBMap v38.26, foi de **14,7x** para o genoma do tatu-peba, e de **9,6x** para o do tatu-canastra. Esse valor se refere ao número médio de vezes que uma base no genoma foi sequenciada, e valores baixos de cobertura significam uma chance maior de que partes do genoma não estão representadas na montagem (Bushnell, 2014). A completude do genoma pode ser verificada, em certa medida, com o uso da ferramenta BUSCO v5.4.6, pois – como os genes que fazem parte de um conjunto BUSCO para um determinado grupo, como “mamíferos”, são selecionados de conjuntos ortólogos presentes em pelo menos 90% desse grupo – há uma expectativa evolutiva de que esses genes sejam encontrados nas montagens genômicas de todas as espécies abrangidas por esse grupo. Caso não sejam identificados genes BUSCO completos nas montagens objeto de análise, ou se apenas uma porcentagem reduzida de genes BUSCO completos estiverem presentes, isso pode indicar que o sequenciamento ou a montagem falharam em capturar o conteúdo completo de genes esperado do organismo (SIMÃO et al., 2015).

Montagens genômicas são, essencialmente, conjuntos de *contigs*, ou seja, fragmentos de DNA de tamanho variável – produzidos a partir de *reads* de sequenciamento com regiões de similaridade sobrepostas, possibilitando sua agregação e formação de sequências maiores

durante o processo de clusterização. Um genoma com alto nível de fragmentação (N50 baixo, L50 elevado) pode levar a dificuldades na etapa de predição das proteínas expressas (cujos resultados estão descritos mais adiante), pois há uma clara correlação entre a fragmentação de um genoma e sua completude em termos de grupos BUSCO, onde a fragmentação tende a diminuir as estimativas de completude devido à dificuldade de identificar todos os marcadores ortólogos únicos em um genoma incompleto (MANNI et al., 2021). Esse efeito é visível na **Tabela 4**, onde podemos verificar que a desfragmentação – realizada com a ferramenta RagTag v2.1.0 e refletida principalmente pelas métricas N50 e L50 – proporciona um aumento significativo não somente na detecção de BUSCOs completos, mas também no número de proteínas preditas com a ferramenta GeMoMa v1.9. É importante notar que o processo de desfragmentação, apesar de ser realizado com auxílio de um genoma de referência (no caso, *D. novemcinctus*) não transfere sequências de nucleotídeos da referência para o genoma-alvo, limitando-se a organizar e orientar os *contigs* existentes com base na homologia de sequência com o genoma de referência (ALONGE et al., 2022).

Em nosso estudo, os resultados obtidos para os dois genomas aqui montados (**Tabela 4**) são coerentes com o fato de que o sequenciamento para o tatu-peba foi realizado em uma *lane* em plataforma Illumina HiSeq2000, gerando *reads* de 100b que somam 39,3 Gbases, enquanto o sequenciamento para o tatu-canastra foi realizado em meia *lane* em plataforma Illumina HiSeq2500, gerando *reads* de 126b que somam 21,7 Gbases (**Tabela 3**): Apesar do sequenciamento para o tatu-canastra ter sido feito em uma plataforma mais moderna e com a geração de *reads* mais longos, o número de nucleotídeos sequenciados para esta montagem genômica é aproximadamente metade do encontrado na montagem feita para o tatu-peba, levando a uma cobertura média inferior e conseqüentemente maior fragmentação, limitando a detecção de grupos BUSCO e a predição de proteínas expressas. Além disso, mesmo após a desfragmentação realizada com a ferramenta RagTag v.2.1.0, verificamos que a detecção de BUSCOs completos para o tatu-canastra permaneceu baixa (26%), e a porcentagem de BUSCOs ausentes continuou alta (55,7%). Isso fornece uma sólida indicação de que o sequenciamento e, subsequentemente, a montagem do genoma, falharam em capturar o conteúdo completo de genes esperado para a espécie. Por esse motivo, a etapa da análise de seleção positiva, a seguir, foi realizada separadamente para um grupo contendo 11 espécies de xenartros (**Tabela 3**), e para um subgrupo com 10 espécies – do qual foi excluído o tatu-canastra – pois a baixa completude desta montagem genômica poderia limitar significativamente a detecção dos ortólogos de cópia única no conjunto dos genomas estudados.

4.2. Análise de Seleção Positiva

A análise com a ferramenta BUSCO v5.4.6 revelou que alguns dos genomas baixados do GenBank (*B. variegatus*, *C. unicinctus*, *C. vellerosus* e *M. tridactyla*) apresentaram porcentagens de BUSCOs completos inferiores a 70% (**Tabela 5**).

Tabela 5: Completude das montagens genômicas de xenartros, avaliada com a ferramenta BUSCO v5.4.6, utilizando o conjunto mammalia_odb10, criado em 08 de janeiro de 2024, contendo 9226 genes ortólogos de mamíferos. Os genomas das espécies destacadas em negrito foram selecionados para desfragmentação com a ferramenta RagTag v2.1.0.

Xenartros	BUSCOs Completos (%)	Completos e Cópia Única (%)	Completos e Duplicados (%)	BUSCOs Fragmentados (%)	BUSCOs Ausentes (%)
<i>B. variegatus</i>	37,1	35,7	1,4	9,5	53,5
<i>C. unicinctus</i>	51,8	48,3	3,5	15,0	33,2
<i>C. vellerosus</i>	26,1	24,5	1,6	16,6	57,3
<i>C. didactylus</i>	95,9	89,8	6,1	0,7	3,4
<i>C. hoffmanni</i>	89,0	83,8	5,1	4,2	6,8
<i>D. novemcinctus</i>	94,1	90,9	3,3	1,2	4,7
<i>M. tridactyla</i>	67,5	65,6	1,9	11,9	20,6
<i>T. tetradactyla</i>	96,3	91,6	4,8	0,5	3,1
<i>T. matacus</i>	95,7	88,6	7,1	1,0	3,3
<i>E. sexinctus</i>	27,5	26,7	0,8	17,5	55,0
<i>P. maximus</i>	6,2	6,0	0,2	8,2	85,6

De acordo com o apresentado na seção anterior, é possível afirmar que as baixas porcentagens de grupos BUSCO detectados provavelmente se deveram aos altos níveis de fragmentação das montagens genômicas, o que deveria também repercutir na etapa de predição das proteínas expressas, feita com a ferramenta GeMoMa v1.9. Esta ferramenta extrai evidências de proteínas conhecidas (referência), mapeia essas evidências para o genoma alvo de maneira a identificar regiões que possam codificar proteínas semelhantes, e então prediz a estrutura de genes potenciais, incluindo a localização de *exons* e *introns* (KEILWAGEN et al., 2016). Assim, quanto menor o nível de fragmentação de uma montagem genômica, mais eficiente deverá ser o mapeamento das proteínas do genoma-referência sobre o genoma-alvo.

De fato, após realizarmos a desfragmentação dos genomas com a ferramenta RagTag v2.1.0 e refazermos a predição das proteínas expressas com a ferramenta GeMoMa v1.9, constatamos um aumento significativo nos números de proteínas preditas (**Tabela 6**).

Tabela 6: Número de proteínas preditas pela ferramenta GeMoMa v1.9, e aumento na quantidade de proteínas preditas como resultado da desfragmentação realizada com a ferramenta RagTag v2.1.0. Note-se que, para *C. didactylus* e *D. novemcinctus* (genomas de alta qualidade, anotados e resolvidos ao nível cromossômico), os valores de proteínas preditas correspondem aos respectivos arquivos de proteínas expressas (.faa) baixados do GenBank. Não foi necessário usar RagTag com os genomas de *C. hoffmanni*, *T. tetradactyla* e *T. matacus*, pois estes apresentaram altas porcentagens (> 85%) de BUSCOs completos.

Xenartros	BUSCOs	BUSCOs	#Proteínas	#Proteínas
	Completos (%)	Completos (%)	Preditas	Preditas
	pré-RagTag	pós-RagTag	pré-RagTag	pós-RagTag
<i>B. variegatus</i>	37,1	55,5	12725	17756
<i>C. uncinctus</i>	51,8	93,9	22365	27606
<i>C. vellerosus</i>	26,1	76,1	13328	22179
<i>C. didactylus</i>	95,9	N/A	54527	N/A
<i>C. hoffmanni</i>	88,9	N/A	29166	N/A
<i>D. novemcinctus</i>	94,2	N/A	55228	N/A
<i>M. tridactyla</i>	67,5	95,6	23297	26086
<i>T. tetradactyla</i>	96,4	N/A	27283	N/A
<i>T. matacus</i>	95,7	N/A	27999	N/A
<i>E. sexinctus</i>	27,5	87,1	11528	20594
<i>P. maximus</i>	6,2	26,0	4392	13592

Isso repercutiu diretamente na detecção de ortólogos de cópia única realizada com a ferramenta OrthoFinder v2.5.5 – que é projetada para lidar com a variabilidade na completude dos genomas, mas a precisão da inferência de ortólogos ainda depende da qualidade dos dados de entrada. Ou seja, genomas incompletos podem levar a resultados menos confiáveis, pois genes ausentes ou fragmentados podem não ser reconhecidos corretamente como ortólogos (EMMS & KELLY, 2019). No nosso estudo realizamos duas análises distintas com a

ferramenta OrthoFinder v2.5.5 para cada grupo (o grupo com as 11 espécies de xenartros, e o subgrupo com 10 espécies, sem o tatu-canastra); uma, utilizando as proteínas preditas para os genomas antes do tratamento com a ferramenta RagTag v2.1.0; e outra posterior à desfragmentação. Os resultados (**Tabela 7**) confirmam a utilidade da ferramenta no processo de determinação de ortólogos de cópia única.

Tabela 7: Número de genes ortólogos de cópia única identificados com a ferramenta OrthoFinder v2.5.5 utilizando os arquivos de proteínas preditas (.FAA) antes e após a desfragmentação dos respectivos genomas com a ferramenta RagTag v2.1.0, para os dois grupos de xenartros.

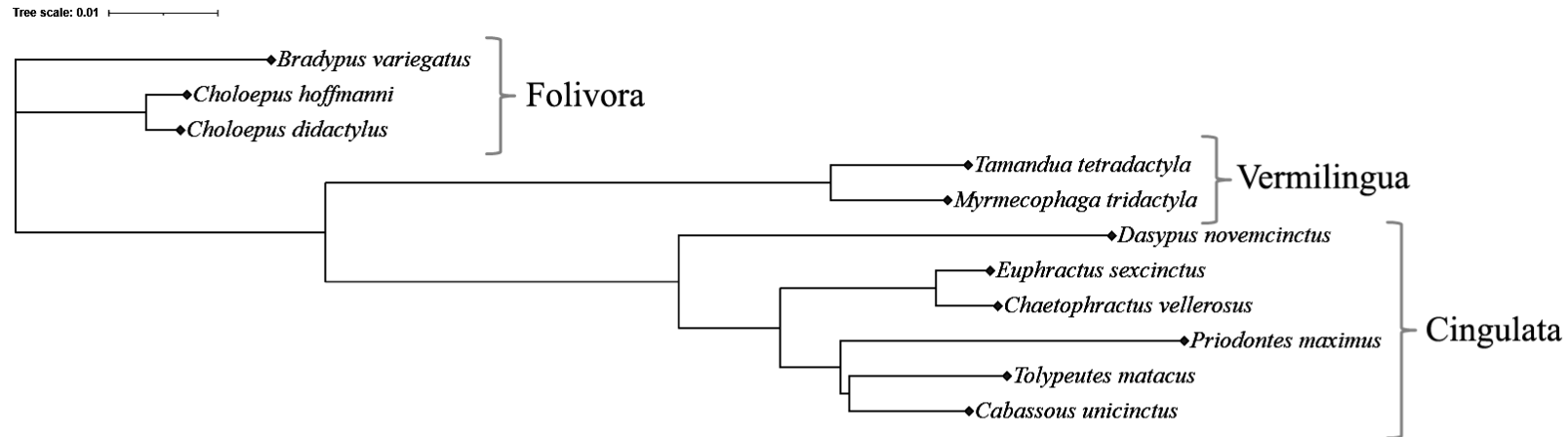
Grupos	# ortólogos de cópia única	# ortólogos de cópia única
	pré-RagTag	pós-RagTag
11 espécies	76	774
10 espécies	334	1347

Os arquivos FASTA de aminoácidos contendo os grupos de genes ortólogos de cópia única foram convertidos em formato FASTA de nucleotídeos; as então sequências de códons foram alinhadas, concatenadas e utilizadas para gerar árvores filogenéticas, que apresentaram um padrão satisfatório (**Figura 2**) para utilização nas análises de seleção positiva, a seguir.

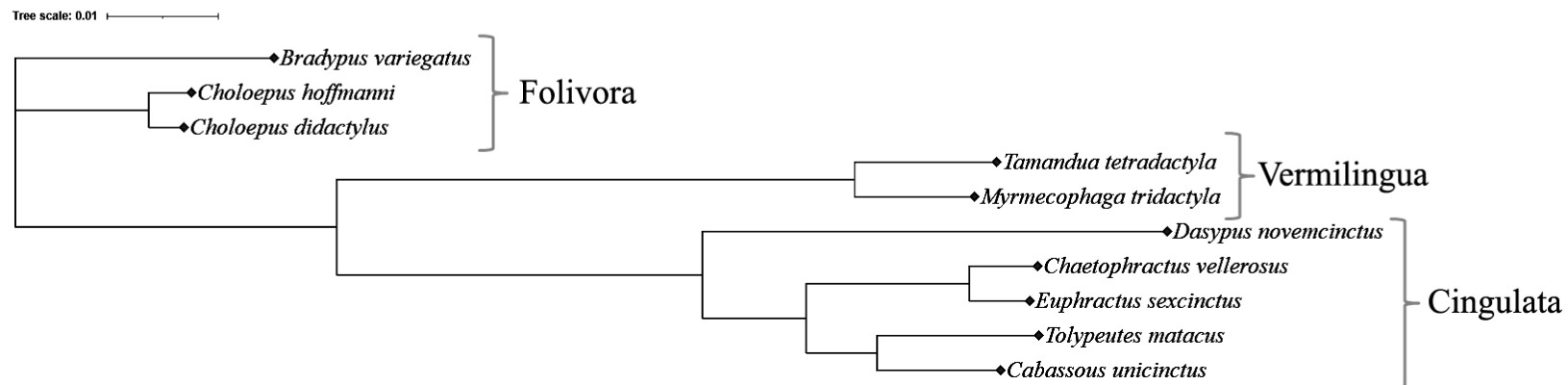
A decisão de não incluir um *outgroup* para o enraizamento da árvore filogenética no teste de seleção positiva com HyPhy v2.5.8, utilizando sua função BUSTED (*Branch-site Unrestricted Statistical Test for Episodic Diversification*), baseia-se na flexibilidade da ferramenta, que pode ser aplicada para detectar seleção diversificadora ao longo de genes inteiros em ramos específicos de uma árvore filogenética, ainda que esta não esteja enraizada. Assim, embora o enraizamento possa facilitar a interpretação da direção evolutiva, o enraizamento não é um requisito para a execução do teste BUSTED, o que simplifica a análise e elimina a necessidade de um *outgroup* que poderia introduzir variabilidade adicional desnecessária (KOSAKOVSKY POND et al., 2019).

Figura 2: Árvores filogenéticas obtidas com a ferramenta IQ-Tree2 v2.1.2, a partir de coleções de genes ortólogos de cópia única (774 genes para o grupo de 11 espécies, e 1347 para o subgrupo de 10 - com a remoção do tatu-canastra, *P. maximus*) para os xenartros abordados neste trabalho. Observa-se o agrupamento das espécies nas ordens (Folivora, Vermilingua, Cingulata) que compõem a superordem Xenarthra.

11 Espécies



10 Espécies



Cada árvore filogenética (**Figura 2**) foi utilizada, juntamente com suas respectivas coleções de alinhamentos de códons para genes ortólogos de cópia única, como parâmetros de entrada para a ferramenta HyPhy v2.5.8, função BUSTED. A detecção de seleção positiva depende da capacidade de identificar sinais de seleção em uma proporção significativa de sítios dentro dos genes analisados. Isso significa que, para a eficácia de métodos como BUSTED, é importante que haja uma proporção detectável de sítios sob seleção positiva para garantir resultados estatisticamente robustos (KOSAKOVSKY POND et al., 2019). Assim, de acordo com os parâmetros de filtragem estabelecidos – proporção de Seleção Positiva > 5%, Razão de Evidência > 10; e P-valor < 0,05 – foram identificados, no grupo com 11 espécies, 29 genes candidatos (de 774), e, no grupo com 10 espécies, 26 genes candidatos (de 1347).

Utilizamos o *software* DAMBE v7.3.32 para realizar o teste de saturação de Xia e validar genes apontados pela análise com HyPhy BUSTED como estando sob seleção positiva. A saturação de substituições ocorre quando múltiplas substituições se acumulam no mesmo *locus*, comprometendo a clareza do sinal evolutivo e a interpretação dos processos seletivos que atuaram sobre os genes. No contexto da análise de seleção positiva de ortólogos de cópia única, o teste de saturação de substituições de Xia assegura que os alinhamentos utilizados não estão saturados, garantindo que as inferências sobre seleção positiva sejam precisas e confiáveis, e confirmando a robustez dos genes candidatos identificados na análise de seleção (XIA, 2018). Assim, com o teste de saturação, foram validados, no grupo com 11 espécies, um conjunto com seis genes positivamente selecionados, e, no grupo com 10 espécies, um conjunto de 12 genes. Todos os genes validados foram anotados com a ferramenta eggNOG v5.0.2 (**Tabela 8**).

Tabela 8: Informações parciais referentes à anotação com a ferramenta eggNOG v5.0.2. Os três primeiros genes da lista são provenientes dos dois grupos (11 e 10 espécies), enquanto os três seguintes são provenientes do grupo com 11 espécies, e os demais do grupo com 10 espécies.

Gene	Nome	Descrição	PFAM	COG	#GO Terms
SCO_10.0131 / SCO_11.0075	AKNAD1	AT-hook-containing transcription factor	AKNA	S	zero
SCO_10.0331 / SCO_11.0195	PPP1R14D	protein phosphatase 1 regulatory	PP1_inhibitor	S	47
SCO_10.0549 / SCO_11.0323	SLC39A7	Solute carrier family 39	Zip	P	173
SCO_11.0112	CATSPER4	cation channel	Ion_trans	PT	172
SCO_11.0557	FOXN1	lymphoid lineage cell migration into thymus	Forkhead	K	183
SCO_11.0618	STK35	kinase 35	Pkinase	T	24
SCO_10.0241	PPBP	Pro-platelet basic protein (chemokine (C-X-C motif) ligand 7)	IL8	O	169
SCO_10.0352	CTSG	Belongs to the peptidase S1 family	Trypsin	E	204
SCO_10.0372	CPA1	Carboxypeptidase A1	Peptidase_M14,Propep_M14,ShK	O	45
SCO_10.0620	HOXC10	spinal cord motor neuron cell fate specification	DUF3528,Homeobox	K	146
SCO_10.0689	TM4SF20	Negative regulation of proteolysis	L6_membrane	S	53
SCO_10.0694	C2orf82	Domain of unknown function (DUF4690)	DUF4690	S	18
SCO_10.0717	C9orf131	protein C9orf131 homolog	FAM75	S	zero
SCO_10.1006	KRT10	Keratin, type I cytoskeletal 10	Filament	S	79
SCO_10.1177	PYCARD	Apoptosis-associated speck-like protein containing a CARD	CARD,PYRIN	S	495

Nota-se, primeiramente, que as análises com os grupos de xenartros com 11 e 10 espécies levaram à identificação de 3 genes positivamente selecionados em comum; porém, nem todos os genes sob seleção positiva do primeiro grupo foram identificados no segundo, apesar do grupo com 10 espécies ser um subgrupo do com 11 espécies. Para explicar isso, podemos argumentar que, quando uma espécie é removida da análise, o alinhamento das sequências pode mudar, levando a diferenças na forma como os sítios são alinhados e interpretados. Isso pode impactar diretamente a detecção de seleção positiva, pois alinhamentos diferentes podem gerar sinais evolutivos distintos, o que, por sua vez, afeta os resultados dos testes de seleção positiva (BOOKER et al., 2017). Além disso, a presença ou ausência de uma espécie pode alterar a árvore filogenética geral e as relações evolutivas entre as espécies restantes. Essa mudança no sinal filogenético pode afetar a detecção de seleção, já que métodos como o HyPhy BUSTED baseiam-se na estrutura da árvore filogenética para inferir pressões seletivas. A remoção de uma espécie pode tanto destacar quanto obscurecer sinais de seleção positiva, dependendo de sua história evolutiva e das relações com outras espécies (BERRIO et al., 2020).

Dentre os genes que apresentaram sinais de seleção positiva, podemos destacar o gene *KRT10*, *Keratin, type I cytoskeletal 10*. Este gene participa na codificação de queratina, uma proteína particularmente importante para o desenvolvimento de características marcantes nos tatus, como a presença de uma carapaça dura e resistente, das garras fortes usadas para cavar e forragear, e de uma pele protetora. Em termos de performance ou *fitness*, este gene pode ter sido de grande relevância para o grupo, conferindo proteção ao animal contra estresses ambientais e predadores, o que representaria, portanto, uma vantagem adaptativa. No contexto do entendimento dos processos de diversificação e evolução das espécies, o que se pode afirmar – ainda que com cautela – é que os genes positivamente selecionados podem ter conferido vantagens evolutivas aos animais pois, mesmo que a seleção positiva não garanta que todas as mudanças sejam benéficas, é razoável supor que muitas das mutações selecionadas positivamente desempenharam algum papel adaptativo (BOOKER et al., 2017). No entanto, ainda que métodos estatísticos possam sugerir seleção positiva, em diversos casos é necessário confirmar se as mutações observadas realmente conferiram vantagens adaptativas através de experimentos ou observações adicionais (BERRIO et al., 2020).

Feitas as devidas ressalvas quanto à necessidade de observações adicionais, um outro gene positivamente selecionado que poderia ser destacado é o *HOXC10*, *Homeobox C10*, que atua na regulação do desenvolvimento anatômico, especialmente na especificação das células motoras na medula espinhal, particularmente nas divisões lombares (MGI, 2024). Em xenartros,

este gene pode ter desempenhado um papel importante nas adaptações locomotoras. Por exemplo, tatus são escavadores proficientes, usando membros anteriores poderosos para cavar tocas, o que requer coordenação motora precisa e força muscular, possibilitadas pelo desenvolvimento adequado dos neurônios motores controlados por HOXC10. Tamanduás possuem membros adaptados para cavar e abrir formigueiros e cupinzeiros, enquanto preguiças dependem de musculatura e controle motor específicos para se locomoverem entre galhos. HOXC10 regula a formação e função dos neurônios motores, podendo ter contribuído diretamente para essas capacidades motoras distintas, conferindo vantagens adaptativas que permitiram aos xenartros explorar seus nichos ecológicos de maneira eficaz (MGI, 2024).

Ainda entre os genes listados na **Tabela 8**, alguns podem ter conferido vantagens adaptativas significativas aos xenartros no que diz respeito aos seus sistemas imunes. O gene FOXP1, que regula a migração de células linfóides para o timo, é fundamental para o desenvolvimento do sistema imunológico. Um timo bem desenvolvido é crucial para a maturação dos linfócitos T, que desempenham um papel central na resposta imune adaptativa. Nos xenartros, essa adaptação pode ter sido especialmente vantajosa para lidar com a diversidade de patógenos encontrados em seus habitats variados, que incluem solos ricos em microorganismos. Outro gene relevante é o PPBP, que codifica uma proteína básica proplaquetária e atua na sinalização e recrutamento de células imunes. Esse gene pode ter contribuído para uma resposta inflamatória eficiente, essencial para a cicatrização de feridas e defesa contra infecções. Considerando que muitos xenartros, como os tatus, frequentemente escavam e podem se ferir, e que tamanduás – que consomem grandes quantidades de formigas e cupins – precisam lidar com as defesas químicas desses insetos, e que preguiças – que vivem nas copas das árvores e têm um metabolismo lento – são suscetíveis a infecções de pele e parasitas, uma resposta inflamatória robusta seria uma vantagem significativa para a sobrevivência. O gene PYCARD também desempenha um papel importante na imunidade, sendo envolvido na regulação da apoptose e resposta inflamatória. A capacidade de controlar a morte celular programada e regular a inflamação é importante para manter a homeostase celular e a defesa contra infecções. Essas especificidades dos sistemas imunes dos xenartros, como a maturação eficiente dos linfócitos T, respostas inflamatórias robustas e controle eficaz da apoptose, poderiam ter sido favorecidas pela seleção positiva nesses genes, proporcionando uma vantagem adaptativa crucial para a sobrevivência e sucesso dessas espécies em seus respectivos nichos ecológicos.

Alguns dos genes identificados estão relacionados a processos digestivos e podem ter conferido vantagens adaptativas aos xenartros, considerando suas peculiaridades em termos de habitat e alimentação. O gene CTSG, que pertence à família das peptidases S1, e o gene CPA1, que codifica a carboxipeptidase A1, estão diretamente envolvidos na digestão de proteínas. A carboxipeptidase A1 é uma enzima exopeptidase que remove aminoácidos das extremidades das proteínas (CATASÚS et al., 1992), enquanto as peptidases S1, como a CTSG, desempenham papel na clivagem de ligações peptídicas internas (TURK, 2006). Essas enzimas são cruciais para a digestão eficiente de proteínas, possivelmente introduzindo uma vantagem adaptativa em xenartros – que possuem dietas variadas – permitindo-lhes maximizar a extração de nutrientes dos alimentos disponíveis em seus habitats específicos.

Apesar da regressão dentária ser uma característica marcante dos xenartros, nossa análise de seleção positiva não identificou genes dentários como alvo de seleção positiva dentro do grupo. A principal razão para isso é que a regressão dentária nos xenartros é um fenômeno associado à seleção relaxada e à deriva genética, ao invés da seleção positiva. A seleção positiva tipicamente envolve mutações que conferem uma vantagem adaptativa, levando ao aumento da frequência dessas mutações na população. Essas mutações geralmente resultam em novas funções ou na melhoria de funções existentes, refletindo-se em uma razão dN/dS (taxa de substituições não-sinônimas para sinônimas) maior que 1. No caso da regressão dentária dos xenartros, o processo ocorre através da acumulação de mutações inativadoras nos genes dentários, como inserções de *frameshift*, deleções, códons de parada prematuros e mutações nos locais de *splicing* (EMERLING et al., 2022). Essas mutações resultam na perda de função dos genes, um processo que é característico da seleção relaxada. Quando uma característica se torna menos necessária devido a mudanças no ambiente ou na dieta, a pressão seletiva para manter essa característica diminui. Isso permite que mutações inativadoras se acumulem e, eventualmente, se fixem na população através da deriva genética. Portanto, as análises de seleção positiva, que buscam sinais de aumento na frequência de mutações benéficas, não são adequadas para detectar a perda de função dentária. Em vez disso, a pseudogenização dos genes dentários nos xenartros foi identificada por meio de análises que revelaram padrões de seleção relaxada e acumulação de mutações inativadoras compartilhadas entre diferentes espécies. Assim, embora a regressão dentária seja uma característica notável dos xenartros, a natureza do processo evolutivo subjacente explica a ausência de detecção de seleção positiva em nossos estudos.

4.3. Análise de Heterozigosidade Autossômica

Após a obtenção do arquivo VCF contendo os dados filtrados (*indels* removidos, e mantendo-se apenas os SNPs) de todos os sítios variantes das 10 espécies de xenartros em relação ao genoma-referência (*D. novemcinctus*), a ferramenta VCFtools v0.1.17 foi utilizada para produzir os resultados descritos na **Tabela 9**, utilizando a equação (1) para calcular as heterozigosidades observadas (H_{OBS}) para todas as espécies, representadas sob a forma de gráfico na **Figura 3**.

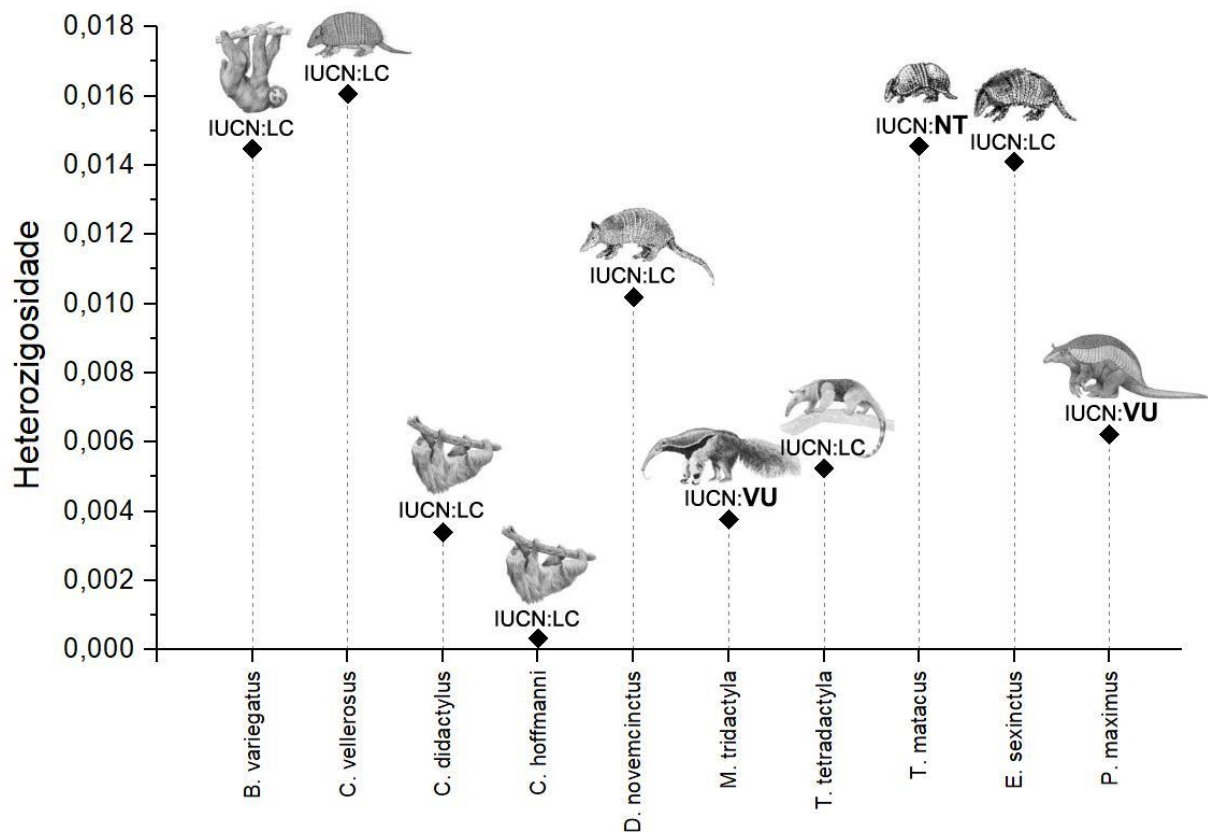
Tabela 9: Resultados do teste de heterozigosidade autossômica.

Xenartros	#Sítios homozigotos observados (Hom _{OBS})	#Sítios observados totais (N _{sítios})	Heterozigosidade Observada (Het _{OBS})
<i>B. variegatus</i>	238934657	242443833	0,01447
<i>C. vellerosus</i>	238523113	242418151	0,01607
<i>C. didactylus</i>	241699832	242524923	0,00340
<i>C. hoffmanni</i>	242485511	242570015	0,00035
<i>D. novemcinctus</i>	240155857	242627025	0,01019
<i>M. tridactyla</i>	241632170	242545145	0,00376
<i>T. tetradactyla</i>	241250043	242523720	0,00525
<i>T. matacus</i>	238892737	242419477	0,01455
<i>E. sexinctus</i>	239047500	242468099	0,01411
<i>P. maximus</i>	240788282	242295362	0,00622

Os resultados do estudo de heterozigosidade autossômica indicaram que *M. tridactyla* e *P. maximus*, ambos considerados “Vulneráveis” pela IUCN, apresentam heterozigosidade relativamente baixa, sugerindo uma diversidade genética reduzida que pode estar associada a eventos demográficos e genéticos históricos relacionados ao seu atual status de vulnerabilidade. Em contraste, *T. matacus*, classificado como “Quase Ameaçado”, exibe um valor de heterozigosidade relativamente alto, comparável ao de algumas espécies classificadas como

“Menos Preocupantes”. *C. hoffmanni*, apesar de ser “Menos Preocupante”, apresenta a menor heteroziguidade entre todas as espécies estudadas, tais dados podem indicar baixa variabilidade genética, possivelmente devido a um histórico de *inbreeding* ou uma população pequena e isolada. Esses dados são importantes para direcionar estratégias de conservação, destacando espécies com menor diversidade genética que podem necessitar de atenção imediata para evitar um declínio populacional maior (SUPPLE & SHAPIRO, 2018).

Figura 3: Gráfico da heteroziguidade autossômica observada para as onze espécies de xenartros abordadas, incluindo o *status* atualmente atribuído a cada espécie pela IUCN (LC: Pouco Preocupante; NT: Quase Ameaçada; VU: Vulnerável).



Estratégias de conservação para *M. tridactyla*, *P. maximus* e *C. hoffmanni* (apesar do *status* deste último ser considerado “Pouco Preocupante”) poderiam incluir a identificação de regiões genômicas com alta e baixa diversidade genética para entender questões de adaptação local e a seleção purificadora. Além disso, a análise de variantes genéticas potencialmente deletérias pode ajudar a informar sobre o impacto da endogamia e a necessidade de intervenções – como a translocação de indivíduos para promover o resgate genético. A presença de alelos adaptativos é crucial para a sobrevivência a longo prazo e deve ser considerada no planejamento

de ações de conservação, como a translocação de indivíduos para aumentar a diversidade genética e reduzir a carga mutacional de alelos deletérios (VAN DER VALK et al., 2024).

Além disso, *C. vellerosus*, *C. didactylus*, *M. tridactyla*, *T. tetradactyla*, e *T. matacus* são espécies cujos *reads* de sequenciamento foram depositados no SRA pelo San Diego Zoo. Neste contexto, seria importante saber se estes espécimes se tratam de animais silvestres ou nascidos em cativeiro. Tal dado pode ser importante para inferir se estes achados se devem às questões de manejo para produzir as gerações em cativeiro, que comumente apresentam o efeito fundador e/ou aumento da endogamia com perda de heterozigosidade, devido ao baixo número de reprodutores utilizados em condições ex-situ; ou se às pressões que o meio ambiente impõe às populações dessas espécies na natureza, incluindo eventos históricos que resultaram na baixa diversidade observada para algumas das espécies.

5. Conclusões

Este trabalho, focado nas espécies *E. sexcinctus* (tatu-peba) e *P. maximus* (tatu-canastra), cujas montagens genômicas são inéditas, possibilitou a descrição de dois genomas para o grupo dos tatus. As montagens genômicas inéditas de *E. sexcinctus* e *P. maximus* somam-se às nove já existentes para xenartros, aumentando a representatividade genômica de um clado com apenas 31 espécies viventes e uma história evolutiva de milhares de espécies extintas.

A caracterização genômica revelou desafios na montagem dos genomas, especialmente para *P. maximus*, cuja baixa cobertura e elevada fragmentação limitaram a completude do conteúdo gênico e, conseqüentemente, a predição de proteínas expressas. A ferramenta RagTag v2.1.0 desempenhou um papel notável na desfragmentação das montagens, possibilitando um aumento significativo na predição de proteínas expressas e subsequente detecção de genes ortólogos de cópia única.

Além da caracterização dos genomas, os dados de diversidade genética e de genes sob seleção positiva, à luz dos *status* de conservação das espécies abordadas, adicionam importantes contribuições aos xenartros cujos genomas e dados de sequenciamento se encontram atualmente disponíveis publicamente, descrevendo características genômicas, filogenômicas e de heterozigosidade autossômica dessas espécies.

A análise de seleção identificou 15 genes sob seleção positiva que podem ter conferido vantagens evolutivas aos xenartros, e a diferença nos genes identificados entre os grupos de 11 e 10 espécies destacou a influência da composição do grupo no processo de detecção de seleção positiva.

Os resultados da heterozigosidade autossômica mostraram variações significativas entre as espécies. O tamanduá-bandeira (*M. tridactyla*) e o tatu-canastra (*P. maximus*), classificados como "Vulneráveis" pela IUCN, apresentaram baixa heterozigosidade autossômica, indicando diversidade genética reduzida. Em contraste, o matoco (*T. matacus*), classificado como "Quase Ameaçado", exibiu alta heterozigosidade.

As análises realizadas neste estudo, incluindo montagem genômica, seleção positiva e heterozigosidade autossômica, fornecem dados detalhados sobre a diversidade genética e evolução dos xenartros, contribuindo para a compreensão dessas espécies e suas necessidades de conservação.

6. Referências Bibliográficas

ALJANABI, S. Universal and rapid salt-extraction of high-quality genomic DNA for PCR-based techniques. *Nucleic Acids Research*. [S. l.]: Oxford University Press (OUP), 15 nov. 1997. DOI 10.1093/nar/25.22.4692.

ALONGE, M.; LEBEIGLE, L.; KIRSCH, M.; JENIKE, K.; OU, S.; AGANEZOV, S.; WANG, X.; LIPPMAN, Z. B.; SCHATZ, M. C.; SOYK, S. Automated assembly scaffolding using RagTag elevates a new tomato system for high-throughput genome editing. *Genome Biology*. [S. l.]: Springer Science and Business Media LLC, 15 dez. 2022. DOI 10.1186/s13059-022-02823-7.

ANDREWS, Simon. 2010. FastQC: a quality control tool for high throughput sequence data. Disponível em: <<http://www.bioinformatics.babraham.ac.uk/projects/fastqc>>

AYA-CUERO, C.; RODRÍGUEZ-BOLAÑOS, A.; SUPERINA, M. Population density, activity patterns, and ecological importance of giant armadillos (*Prionomys maximus*) in Colombia. *Journal of Mammalogy*. [S. l.]: Oxford University Press (OUP), 3 maio 2017. DOI 10.1093/jmammal/gyx006.

BANKEVICH, A.; NURK, S.; ANTIPOV, D.; GUREVICH, A. A.; DVORKIN, M.; KULIKOV, A. S.; LESIN, V. M.; NIKOLENKO, S. I.; PHAM, S.; PRJIBELSKI, A. D.; PYSHKIN, A. V.; SIROTKIN, A. V.; VYAHHI, N.; TESLER, G.; ALEKSEYEV, M. A.; PEVZNER, P. A. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *Journal of Computational Biology*. [S. l.]: Mary Ann Liebert Inc, maio 2012. DOI 10.1089/cmb.2012.0021.

BERRIO, A.; HAYGOOD, R.; WRAY, G. A. Identifying branch-specific positive selection throughout the regulatory genome using an appropriate proxy neutral. *BMC Genomics*. [S. l.]: Springer Science and Business Media LLC, 13 maio 2020. DOI 10.1186/s12864-020-6752-4.

BOLGER, Anthony M.; LOHSE, Marc; USADEL, Bjoern. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*, [s.l.], v. 30, n. 15, p.2114-2120, 1 abr. 2014. Oxford University Press (OUP). <http://dx.doi.org/10.1093/bioinformatics/btu170>.

BROAD INSTITUTE. Picard Toolkit. Disponível em: <http://broadinstitute.github.io/picard/>. Acesso em: 26 fev. 2024.

BUSHNELL, Brian. 2014. BMap short read aligner, and other bioinformatic tools. Disponível em: <<https://sourceforge.net/projects/bbmap/>>

CAPELLA-GUTIÉRREZ, Salvador; SILLA-MARTÍNEZ, José M.; GABALDÓN, Toni. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*. [S. l.]: Oxford University Press (OUP), 8 jun. 2009. DOI 10.1093/bioinformatics/btp348.

CARTER, T. S.; SUPERINA, M.; LESLIE, D. M., Jr. *Priodontes maximus* (Cingulata: Chlamyphoridae). *Mammalian Species*. [S. l.]: Oxford University Press (OUP), 5 maio 2016. DOI 10.1093/mspecies/sew002.

CATASÚS, L.; VILLEGAS, V.; PASCUAL, R.; AVILÉS, F. X.; WICKER-PLANQUART, C.; PUIGSERVER, A. cDNA cloning and sequence analysis of human pancreatic procarboxypeptidase A1. *Biochemical Journal*. [S. l.]: Portland Press Ltd., 1 out. 1992. DOI 10.1042/bj2870299.

CHEN, Shifu; ZHOU, Yanqing; CHEN, Yaru; GU, Jia. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. [S. l.]: Oxford University Press (OUP), 1 set. 2018. DOI 10.1093/bioinformatics/bty560.

DANECEK, Petr; AUTON, Adam; ABECASIS, Goncalo; ALBERS, Cornelis A.; BANKS, Eric; DEPRISTO, Mark A.; HANDSAKER, Robert E.; LUNTER, Gerton; MARTH, Gabor T.; SHERRY, Stephen T.; MCVEAN, Gilean; DURBIN, Richard. The variant call format and VCFtools. *Bioinformatics*. [S. l.]: Oxford University Press (OUP), 7 jun. 2011. DOI 10.1093/bioinformatics/btr330.

DANECEK, Petr; BONFIELD, James K; LIDDLE, Jennifer; MARSHALL, John; OHAN, Valeriu; POLLARD, Martin O; WHITWHAM, Andrew; KEANE, Thomas; MCCARTHY, Shane A; DAVIES, Robert M; LI, Heng. Twelve years of SAMtools and BCFtools. *GigaScience*. [S. l.]: Oxford University Press (OUP), 29 jan. 2021. DOI 10.1093/gigascience/giab008.

DESBIEZ, A. L. J.; KLUYBER, D. The Role of Giant Armadillos (*Priodontes maximus*) as Physical Ecosystem Engineers. *Biotropica*. [S. l.]: Wiley, 19 ago. 2013. DOI 10.1111/btp.12052.

DELSUC, F.; CTZEFLIS, F. M.; STANHOPE, M. J.; DOUZERY, E. J. P. The evolution of armadillos, anteaters and sloths depicted by nuclear and mitochondrial phylogenies:

implications for the status of the enigmatic fossil Eurotamandua. Proceedings of the Royal Society of London. Series B: Biological Sciences. [S. l.]: The Royal Society, 7 ago. 2001. DOI 10.1098/rspb.2001.1702.

DI BLANCO, Y. E.; DESBIEZ, A. L. J.; DI FRANCESCANTONIO, D.; DI BITETTI, M. S. Excavations of giant armadillos alter environmental conditions and provide new resources for a range of animals. Journal of Zoology. [S. l.]: Wiley, 29 maio 2020. DOI 10.1111/jzo.12782.

EMMS, David M.; KELLY, Steven. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biology. [S. l.]: Springer Science and Business Media LLC, 14 nov. 2019. DOI 10.1186/s13059-019-1832-y.

FÁVERO MASSOCATO, G.; DESBIEZ, A. L. J. Presença e importância do tatu-canastra, *Priodontes maximus* (Kerr, 1792), na maior área protegida do leste do Estado de Mato Grosso do Sul, Brasil. Edentata: The Newsletter of the IUCN/SSC Anteater, Sloth and Armadillo Specialist Group. [S. l.]: IUCN, 1 dez. 2017. DOI 10.2305/iucn.ch.2017.edentata-18-1.4.en.

FISCHER, Eliana Michelle Paviotti. Diversidade genética de *Euphractus sexcinctus* (Linnaeus, 1758) em áreas de Cerrado e Pantanal. 2016. Tese (Doutorado) - Pós-Graduação em Ecologia e Recursos Naturais, Universidade Federal de São Carlos, São Carlos, 2016.

FONTES, B. L.; DESBIEZ, A. L. J.; MASSOCATO, G. F.; SRBEK-ARAUJO, A. C.; SANAIOTTI, T. M.; BERGALLO, H. G.; FERREGUETTI, Á. C.; NOIA, C. H. R.; SCHETTINO, V. R.; VALLS, R.; MOREIRA, D. de O.; GATTI, A.; MENDONÇA, E. de S.; BANHOS, A. The local extinction of one of the greatest terrestrial ecosystem engineers, the giant armadillo (*Priodontes maximus*), in one of its last refuges in the Atlantic Forest, will be felt by a large vertebrate community. Global Ecology and Conservation. [S. l.]: Elsevier BV, dez. 2020. DOI 10.1016/j.gecco.2020.e01357.

GASPAROTTO, V. P. de O.; ATTIAS, N.; MIRANDA, F. R.; SORESINI, G. C. G.; DA COSTA CANENA, A.; MOURÃO, G. Chemical Immobilization of Free-ranging Yellow Armadillos (*Euphractus sexcinctus*) for Implantation of Intra-abdominal Transmitters. Journal of Wildlife Diseases. [S. l.]: Wildlife Disease Association, out. 2017. DOI 10.7589/2016-12-262.

GUREVICH, A.; SAVELIEV, V.; VYAHHI, N.; TESLER, G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. [S. l.]: Oxford University Press (OUP), 19 de fevereiro 2013. DOI 10.1093/bioinformatics/btt086.

HUERTA-CEPAS, J.; SZKLARCZYK, D.; HELLER, D.; HERNÁNDEZ-PLAZA, A.; FORSLUND, S. K.; COOK, H.; MENDE, D. R.; LETUNIC, I.; RATTEI, T.; JENSEN, L. J.; VON MERING, C.; BORK, P. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. Nucleic Acids Research. [S. l.]: Oxford University Press (OUP), 12 nov. 2018. DOI 10.1093/nar/gky1085.

INTERNATIONAL UNION FOR CONSERVATION OF NATURE. The IUCN Red List of Threatened Species. Disponível em: <<https://www.iucnredlist.org/>>. Acesso em: 12 de fevereiro, 2024.

JABER, J.; MORALES BORDON, D.; ARENCIBIA, A.; CORBERA, J.; CONDE-FELIPE, M.; AYALA, M.; ENCINOSO, M. Correlation between Cross-Sectional Anatomy and Computed Tomography of the Normal Six-Banded Armadillo (*Euphractus sexcinctus*) Nasal Cavity and Paranasal Sinuses. *Animals*. [S. l.]: MDPI AG, 8 abr. 2024. DOI 10.3390/ani14071135.

KALYAANAMOORTHY, S.; MINH, B. Q.; WONG, T. K. F.; VON HAESELER, A.; JERMIIN, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nature Methods*. [S. l.]: Springer Science and Business Media LLC, 8 maio 2017. DOI 10.1038/nmeth.4285.

KEILWAGEN, Jens; HARTUNG, Frank; PAULINI, Michael; TWARDZIOK, Sven O.; GRAU, Jan. Combining RNA-seq data and homology-based gene prediction for plants, animals and fungi. *BMC Bioinformatics*. [S. l.]: Springer Science and Business Media LLC, 30 maio 2018. DOI 10.1186/s12859-018-2203-5.

KOSAKOVSKY POND, S. L.; POON, A. F. Y.; VELAZQUEZ, R.; WEAVER, S.; HEPLER, N. L.; MURRELL, B.; SHANK, S. D.; MAGALIS, B. R.; BOUVIER, D.; NEKRUTENKO, A.; WISOTSKY, S.; SPIELMAN, S. J.; FROST, S. D. W.; MUSE, S. V. HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. (K. Crandall, org.). *Molecular Biology and Evolution*. [S. l.]: Oxford University Press (OUP), 27 ago. 2019. DOI 10.1093/molbev/msz197.

LETUNIC, I.; BORK, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*. [S. l.]: Oxford University Press (OUP), 22 abr. 2021. DOI 10.1093/nar/gkab301.

LÖYTYNOJA, Ari. Phylogeny-aware alignment with PRANK. *Methods in Molecular Biology*. [S. l.]: Humana Press, 23 ago. 2013. DOI 10.1007/978-1-62703-646-7_10.

MANNI, M.; BERKELEY, M. R.; SEPPEY, M.; SIMÃO, F. A.; ZDOBNOV, E. M. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. (J. Kelley, org.). *Molecular Biology and Evolution*. [S. l.]: Oxford University Press (OUP), 28 jul. 2021. DOI 10.1093/molbev/msab199.

MCKENNA, Aaron; HANNA, Matthew; BANKS, Eric; SIVACHENKO, Andrey; CIBULSKIS, Kristian; KERNYTSKY, Andrew; GARIMELLA, Kiran; ALTSHULER, David; GABRIEL, Stacey; DALY, Mark; DEPRISTO, Mark A. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Research*. [S. l.]: Cold Spring Harbor Laboratory, 19 jul. 2010. DOI 10.1101/gr.107524.110.

MGI. Hoxc10 MGI Mouse Gene Detail - MGI:96192 - Homeobox C10. Disponível em: <https://www.informatics.jax.org/marker/MGI:96192>. Acesso em: 1 mar. 2024.

MINH, B. Q.; SCHMIDT, H. A.; CHERNOMOR, O.; SCHREMPF, D.; WOODHAMS, M. D.; VON HAESELER, A.; LANFEAR, R. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. (E. Teeling, org.). *Molecular Biology and Evolution*. [S. l.]: Oxford University Press (OUP), 3 fev. 2020. DOI 10.1093/molbev/msaa015.

MURRELL, B.; WEAVER, S.; SMITH, M. D.; WERTHEIM, J. O.; MURRELL, S.; AYLWARD, A.; EREN, K.; POLLNER, T.; MARTIN, D. P.; SMITH, D. M.; SCHEFFLER, K.; KOSAKOVSKY POND, S. L. Gene-Wide Identification of Episodic Selection. *Molecular Biology and Evolution*. [S. l.]: Oxford University Press (OUP), 19 fev. 2015. DOI 10.1093/molbev/msv035.

PITMAN, R. L. Habitat Use and Activity of the Giant Armadillo (*Priodontes maximus*): Preliminary Data from Southeastern Peru. Unpublished, 2004. DOI 10.13140/RG.2.2.22666.47049.

PRIKRYL, Martin. 2000. WinSCP: WinSCP is a popular free SFTP and FTP client for Windows. Disponível em: <<https://github.com/martinprikryl>>

SAMBROOK J, RUSSEL DW (2001). *Molecular Cloning: A Laboratory Manual*. 4. ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

SIMÃO, F. A.; WATERHOUSE, R. M.; IOANNIDIS, P.; KRIVENTSEVA, E. V.; ZDOBNOV, E. M. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. [S. l.]: Oxford University Press (OUP), 9 jun. 2015. DOI 10.1093/bioinformatics/btv351.

SOUSA, P. C.; SANTOS, E. A. A.; SILVA, A. M.; BEZERRA, J. A. B.; SOUZA, A. L. P.; LIMA, G. L.; OLIVEIRA, M. F.; SILVA, A. R. Identification of ultrastructural and functional damages in sperm from six-banded armadillos (*Euphractus sexcinctus*) due to cryopreservation. *Pesquisa Veterinária Brasileira*. [S. l.]: FapUNIFESP (SciELO), ago. 2016. DOI 10.1590/s0100-736x2016000800015.

SUPERINA, M.; PAGNUTTI, N.; ABBA, A. M. What do we know about armadillos? An analysis of four centuries of knowledge about a group of South American mammals, with emphasis on their conservation. *Mammal Review*. [S. l.]: Wiley, 15 jul. 2013. DOI 10.1111/mam.12010.

SUPERINA, M.; LOUGHRY, W. J. Why do Xenarthrans matter?: Table 1. *Journal of Mammalogy*. [S. l.]: Oxford University Press (OUP), 24 jun. 2015. DOI 10.1093/jmammal/gyv099.

SUPPLE, M. A.; SHAPIRO, B. Conservation of biodiversity in the genomics era. *Genome Biology*. [S. l.]: Springer Science and Business Media LLC, 11 set. 2018. DOI 10.1186/s13059-018-1520-3.

TATHAM, Simon. 2000. PuTTY: a Telnet and SSH client for Windows. Disponível em: <<https://www.chiark.greenend.org.uk/~sgtatham/putty/>>

TURK, B. Targeting proteases: successes, failures and future prospects. *Nature Reviews Drug Discovery*. [S. l.]: Springer Science and Business Media LLC, set. 2006. DOI 10.1038/nrd2092.

VAN DER VALK, T.; JENSEN, A.; CAILLAUD, D.; GUSCHANSKI, K. Comparative genomic analyses provide new insights into evolutionary history and conservation genomics of gorillas. *BMC Ecology and Evolution*. [S. l.]: Springer Science and Business Media LLC, 26 jan. 2024. DOI 10.1186/s12862-023-02195-x.

VASIMUDDIN, Md.; MISRA, Sanchit; LI, Heng; ALURU, Srinivas. Efficient Architecture-Aware Acceleration of BWA-MEM for Multicore Systems. 2019 IEEE International Parallel and Distributed Processing Symposium (IPDPS). [S. l.]: IEEE, maio 2019. DOI 10.1109/ipdps.2019.00041.

VIZCAÍNO, S. F. The teeth of the “toothless”: novelties and key innovations in the evolution of xenarthrans (Mammalia, Xenarthra). *Paleobiology*. [S. l.]: Cambridge University Press (CUP), 2009. DOI 10.1666/0094-8373-35.3.343.

XIA, X.; LEMEY, P. Assessing substitution saturation with DAMBE. *The Phylogenetic Handbook*. [S. l.]: Cambridge University Press, 26 mar. 2009. DOI 10.1017/cbo9780511819049.022.

XIA, X. DAMBE7: New and Improved Tools for Data Analysis in Molecular Biology and Evolution. (S. Kumar, org.). *Molecular Biology and Evolution*. [S. l.]: Oxford University Press (OUP), 14 abr. 2018. DOI 10.1093/molbev/msy073.