UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto de Ciências Biológicas

Departamento de Genética, Ecologia e Evolução

Programa de Pós-Graduação em Genética

Radarane Santos Sena

IDENTIFICAÇÃO E CARACTERIZAÇÃO DOS SATELITOMAS DE OITO ESPÉCIES DE XENARTHRA

BELO HORIZONTE

2024

Radarane Santos Sena

IDENTIFICAÇÃO E CARACTERIZAÇÃO DOS SATELITOMAS DE OITO ESPÉCIES DE XENARTHRA

Tese apresentada ao Programa de Pós-Graduação em Genética do Instituto de Ciências Biológicas da Universidade Federal de Minas Gerais, como pré-requisito obrigatório para a obtenção do título de Doutor em Genética, área de concentração "Genética Evolutiva e de Populações".

Orientadora: Profa. Dra. Marta Svartman

Coorientador: Prof. Dr. Gustavo Campos e Silva Kuhn

BELO HORIZONTE

2024

043 Sena, Radarane Santos.

Identificação e caracterização dos satelitomas de oito espécies de Xenarthra [manuscrito] / Radarane Santos Sena. – 2024.

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

Orientadora: Profa. Dra. Marta Svartman. Coorientador: Prof. Dr. Gustavo Campos e Silva Kuhn.

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

1. Genética. 2. DNA Satélite. 3. Heterocromatina. 4. Centrômero. 5. Cingulados. I. Svartman, Marta. II. Kuhn, Gustavo Campos e Silva. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 575

Ficha catalográfica elaborada pela Bibliotecária Fabiane C M Reis. CRB/6 – 2680

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

FOLHA DE APROVAÇÃO

"Identificação e Caracterização dos Satelitomas de Oito Espécies de Xenarthra"

Radarane Santos Sena

Tese aprovada pela banca examinadora constituída pelos Professores:

Marta Svartman
Universidade Federal de Minas Gerais

Edivaldo Herculano Correa de Oliveira
Universidade Federal do Pará

Guilherme Borges Dias
Uppsala University

Fernando Araújo Perini
Universidade Federal de Minas Gerais

Maria Bernadete Lovato
Universidade Federal de Minas Gerais

Belo Horizonte, 07 de fevereiro de 2024.

Documento assinado eletronicamente por **Edivaldo Herculano Correa de Oliveira**, **Usuário Externo**, em 07/02/2024, às 09:16, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Guilherme Borges Dias**, **Usuário Externo**, em 07/02/2024, às 11:35, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Fernando Araujo Perini**, **Professor do Magistério Superior**, em 07/02/2024, às 11:36, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Maria Bernadete Lovato**, **Professora do Magistério Superior**, em 07/02/2024, às 12:12, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Marta Svartman**, **Professora do Magistério Superior**, em 08/02/2024, às 10:47, conforme horário oficial de Brasília, com fundamento no art. 5º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no site https://sei.ufmg.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **3015982** e o código CRC **1BAB3A3B**.

Dedico

Aos meus pais Denise e Jover que com muito apoio e carinho não mediram esforços para que eu chegasse até esta etapa da minha vida.

AGRADECIMENTOS

Na realização do presente trabalho, contei com o apoio direto ou indireto de muitas pessoas queridas às quais sou imensamente grata:

À minha querida orientadora, professora Marta Svartman, pelos ensinamentos, pela paciência e convivência. Sou muito grata pela oportunidade que você me deu de desenvolver esse projeto sob sua orientação, pela preocupação com a minha formação profissional e com o meu crescimento pessoal.

Ao professor Gustavo Campos e Silva Kuhn, por todos os ensinamentos, paciência, atenção e ajuda.

Ao meu namorado André, pelo carinho, amor, ajuda com o R, pela enorme paciência, principalmente nos momentos finais do doutorado. Seu apoio foi muito importante para vencer mais essa etapa da minha vida.

À Ana Matioli, minha queria companheira de LCEv pós-pandemia, sua parceria, apoio, carinho e disposição foram fundamentais para aguentar a solidão do lab.

À Mirela, por todos os ensinamentos, salvamentos, incentivos, pela amizade, carinho e puxões de orelha. Sempre pude contar com sua ajuda e sou imensamente grata.

À Alice e seus queridos pais, Jussara e Sinval, pela amizade, carinho e incentivos.

Ao Braulio por seus ensinamentos, carinho, amizade e paciência.

Aos queridos amigos de laboratório que deixam a vida mais leve e divertida Rafaella, Erick e Pedro.

Aos queridos companheiros de GeneTime 2022, quando embarquei na disciplina de organização de eventos da PPG-Gen não imaginava que acabaria conhecendo pessoas tão especiais e parceiras. Formamos um belo time, sofremos, mas nos divertimos e entregamos um ótimo evento!! Elvira, Júlia, Juliana, Nayara, João, César, Carol, Izabela e a professora Fernanda Antunes obrigada!!

À Elvira, pela amizade, carinho e alegria.

Ao técnico Daniel dos Santos Filho pela convivência, disposição em ajudar, com paciência, bom humor e gentileza.

À Fundação de parques municipais e ZooBotânica, em especial à Valéria por nos enviar preciosas amostras.

Às queridas Raíssa, Daniela Izabela que com muita competência e gentileza nos ajudam.

Aos meus primos e primas, tios e tias, sobrinhos emprestados, meus avós queridos, por todo amor e incentivo.

Ao meu querido irmão, Hassan, pelo amor, amizade e torcida!

Aos meus amados pais que são meus maiores incentivadores, que nunca mediram esforços para que eu pudesse estudar e realizar nossos sonhos, porque sei que os meus sonhos fazem parte dos de vocês. Tudo que conquistei e ainda conquistarei será sempre dedicado a vocês.

Ao programa de Pós-Graduação em Genética da UFMG.

À CAPES pela bolsa de mestrado, à FAMIG e ao CNPq pelo incentivo e investimento em pesquisa.

RESUMO

Os DNAs satélites (satDNAs) são sequências altamente repetitivas, que formam longas cadeias e são os principais componentes da heterocromatina constitutiva. Seus transcritos (ncRNA) estão relacionados a funções biológicas importantes como a manutenção, organização e correto funcionamento das regiões centroméricas e teloméricas. Apesar de sua importância para os genomas, os satDNAs representam uma parcela pouco estudada dos genomas dos eutérios. Neste trabalho identificamos e caracterizamos os satDNAs mais abundantes nos genomas de oito espécies de Xenarthra usando análises *in silico* e citogenômicas. No Capítulo1, tratamos da identificação de dois satDNA nos genomas das preguiças-de-dois-dedos do gênero *Choloepus*, SATCHO1 (117pb) e SATCHO2 (2292pb). SATCHO1 foi o satDNA mais abundante nas duas espécies. Usando experimentos de hibridação *in situ* fluorescente (FISH), mapeamos essas duas sequências nos cromossomos de *C. hoffmanni*. As duas sequências colocalizaram na região centromérica de todos os cromossomos, exceto o X. SATCHO1 e SATCHO2 também possuem algumas características que sugerem alguma função centromérica como: presença de sítios de ligação da proteína centromérica B (CENP-B box), pares de pequenos dímeros e a formação de possíveis estruturas secundárias estáveis. No Capítulo 2, usando ferramentas *in silico*, identificamos 39 sequências de satDNA em seis espécies de Xenarthra: *Bradypus variegatus*, *Myrmecophaga tridactyla*, *Tamandua tetradactyla*, *Chaetophractus vellerosus*, *Dasypus novemcinctus*, e *Tolypeutes matacus*. Após uma caracterização *in silico,* verificamos que dez desses 39 satDNAs pertencem à família SATCHO1 e um pertence à família SATCHO2. Todos os satDNA das famílias SATCHO 1 e 2 possuem características sugestivas de funções centroméricas, tais como presença de CENP-B box, de pequenos pares de sequências palindrômicas e a formação de possíveis estruturas secundárias estáveis. Além disso, cinco satDNA possuem arranjos em *higher-order*, que variam de quatro a seis repetições internas de 117pb. Já no Capítulo 3, apresentamos os resultados preliminares de um estudo no qual caracterizamos e mapeamos por FISH seis satDNAs de *M. tridactyla* e *T. tetradactyla*. O satDNA MTR156, mais abundante em *M. tridactyla*, mapeou em regiões terminais de quase todos os cromossomos de *M. tridactyla* e em alguns pares de *T. tetradactyla*. Sondas teloméricas foram mapeadas em regiões pericentroméricas de seis pares autossômicos de *M. tridactyla*, em regiões de heterocromatina constitutiva evidenciadas por bandeamento CBG. Este trabalho apresenta pela primeira vez análises mais detalhadas de sequências repetitivas dos genomas de um número considerável de espécies de Xenarthra, um grupo

filogeneticamente basal entre os eutérios, com distribuição e história evolutiva eminentemente sul-americanas e ainda pouco estudado.

Palavras-chave: DNA satélite; Heterocromatina; Centrômero; RepeatExplorer; Xenarthra.

ABSTRACT

Satellite DNAs (satDNAs) are highly repetitive sequences that form long arrays and are the main components of constitutive heterochromatin. Their transcripts (ncRNA) are related to essential biological functions such as maintenance, organization, and correct functioning of the centromeric and telomeric regions. Despite their importance, satDNAs represent a poorly studied fraction of eutherian genomes. In this work we identified and characterized the most abundant satDNAs in the genomes of eight Xenarthra species using *in silico* and cytogenomic analyses. In Chapter 1, we identified two satDNAs in the genomes of the two-toed-sloths from the genus *Choloepus*, SATCHO1 (117bp) and SATCHO2 (2292bp). SATCHO1 was the most abundant satDNA in both species. Using fluorescence in situ hybridization (FISH) experiments, we mapped these two sequences onto *C. hoffmanni* chromosomes. The two sequences colocalized in the centromeric region of all chromosomes, except the X. SATCHO1 and SATCHO2 have some characteristics that suggest centromeric function, such as the presence of centromeric protein B binding sites (CENP-B box), pairs of small dimers and the formation of possible stable secondary structures. In Chapter 2, we used *in silico* tools to identify 39 satDNA sequences in the genomes of six Xenarthra species (*Bradypus variegatus*, *Myrmecophaga tridactyla*, *Tamandua tetradactyla*, *Chaetophractus vellerosus*, *Dasypus novemcinctus*, and *Tolypeutes matacus*). The *in silico* analyses allowed us to identify ten of these 39 satDNAs as belonging to the SATCHO1 family and one belonging to the SATCHO2 family. All satDNAs from the SATCHO 1 and 2 families have characteristics suggestive of centromeric functions, such as the presence of the CENP-B box, small pairs of palindromic sequences, and the formation of possible stable secondary structures. Furthermore, five satDNAs have higher-order arrangements, ranging from four to six 117bp internal repeats. In Chapter 3, we describe the cytogenomic characterization of six satDNAs present in the genomes of *M. tridactyla* and *T. tetradactyla*. SatDNA MTR156 was mapped by FISH to the terminal regions of almost all *M. tridactyla* autosomes and in some pairs of *T. tetradactyla*. Telomeric probes mapped to pericentromeric regions of six *M. tridactyla* autosomes, coinciding with constitutive heterochromatin regions revealed after CBG banding. This work presents for the first time more detailed analyses of repetitive DNA content of a considerable number of Xenarthra species, a phylogenetically basal group among eutherians, with an eminently South American distribution and evolutionary history that still poorly studied.

Keywords: Satellite DNA; Heterochromatin; Centromere; RepeatExplorer; Xenarthra.

LISTA DE FIGURAS

Introdução

Figura 1 - Filogenia dos gêneros de Xenarthra viventes, incluindo os tempos de divergência 18

Figura 2 - Representação esquemática dos DNAs repetitivos...24

Figura 3 - Representação esquemática do modelo de Evolução Combinada – Adaptação de Dover 1982...25

Resultados e Discussão

Capítulo 1

Artigo: Identification and Characterization of Satellite DNAs in Two-Toed Sloths of the Genus *Choloepus* (Megalonychidae, Xenarthra)

Figure 1 - Identification of conserved and variable satDNA segments of *Choloepus didactylus* and *C. hoffmanni* by sliding window analysis……………………………...…….33

Figure 2 - Comparative phylogenetic analyses of (a) SATCHO1 and (b) SATCHO2 sequences between *C. didactylus* and *C. hoffmanni* inferred by the Neighbor-joining method with 1000 bootstraps…………………………………………………………………….34

Figure 3 - Karyotype of *Choloepus hoffmanni* (2n = 51) after: (a) GTG-banding and (b) CBG-banding………………………………………………………………………...…35

Figure 4 - Metaphases of *Choloepus hoffmanni* after FISH using (a) SATCHO1 (red) and (b) SATCHO2 (green) as probes. (c) Merged signals from (a) and (b).......................................35

Figure 5 - CENP-B motifs identified in sequences from SATCHO1 and SATCHO2 aligned with *Homo sapiens* (HSA) and the evolutionary conserved domain (ECD)...35

Figure 6 - (a) SATCHO1 and (b) SATCHO2 sequences with dyad palindrome sequences....36

Figure 7 - The optimal secondary structure of (a) SATCHO1 and (b) SATCHO2 predicted by RNAfold..37

Supplementary Figure 1 - Uncropped agarose gel of PCR products from SATCHO1 amplified on genomic DNAs of *Bradypus variegatus* and *Myrmecophaga tridactyla*......44

Supplementary Figure 2 - Uncropped agarose gel of PCR products from SATCHO2 amplified on genomic DNAs of *Choloepus hoffmanni*, *Bradypus variegatus* and *Myrmecophaga tridactyla*......44

Capítulo 2

Manuscrito: Comparative Analyses of Satellite DNAs in Six Xenarthra Species

Figure 1 - Repetitive content for each species analyzed......50

Figure 2 - Heat map showing the satDNAs from the SATCHO1 family based on their pairwise identity and a cluster tree inferred from their pairwise distance......52

Figure 3 - Comparison between the phylogenetic relationships of extant Xenarthra species (a) (Gibb et al. 2016) and a phylogenetic tree of SATCHO1 and its family (b)......54

Figure 4 - Schematic figure with the higher-order structures of satDNAs MTR3029, CVE811, CVE812, CVE815 and TMA451......55

Figure 5 - Maximum Likelihood tree of SATCHO1 copies from *Choloepus didactylus* and *C. hoffmanni* and copies from the satDNAs with 116 bp and 117 bp in length (BVA116, BVA117, CVE116, CVE117, and TMA117)......56

Figure 6 - Small dyad palindromic sequences from a) BVA116, b) BVA117, c) BVA2291, d) MTR3029, e) CVE116, f) CVE117, g) CVE811, h) CVE812, i) CVE815, j) TMA117, k) TMA451......58

Supplementary File 3 - Sequence alignment among SATCHO1 and its subfamilies (left side) and the pairwise percentage of identity of each alignment (right side): BVA116 (a, b); BVA117 (c, d); MTR3029 (e, f); CVE116 (g, h); CVE117 (I, j); CVE811 (k, l); CVE812 (m, n); CVE815 (o, p); TMA117 (q, r); TMA451 (s, t). BVA – *Bradypus variegatus*; MTR – *Myrmecophaga tridactyla*; CVE – *Chaetophractus vellerosus*; TMA – *Tolypeutes matacus*......71

Supplementary File 5 - Cluster graph layout of satDNAs from the SATCHO1 family identified by TAREAN......74

Capítulo 3

Manuscrito: Identification and Characterization of the Most Abundant satDNAs in the Genomes of the Anteaters *Myrmecophaga tridactyla* and *Tamandua tetradactyla* (Myrmecophagidae, Xenarthra)

Figure 1 - Maximum-Likelihood tree of MTR156 copies retrieved from assembled genomes of *Myrmecophaga tridactyla* and *Tamandua tetradactyla*..82

Figure 2 - Karyotype of *M. tridactyla* (a-c) and *T. tetradactyla* (b-d) after: (a-b) GTG-banding and (c-d) CBG-banding..83

Figure 3 - Karyotypes of *M. tridactyla* (a-d) and *T. tetradactyla* (e-g) after FISH exper........85

Supplementary File 2 - Agarose gel of purified PCR products from satDNAs MR156 and MTR696 amplified from *Myrmecophaga tridactyla* genomic DNA.......................................89

Supplementary File 3 - Agarose gel of purified PCR products from satDNAs MTR1550 and MTR3029 amplified from *Myrmecophaga tridactyla* genomic DNA.....................................89

Supplementary File 4 - Agarose gel of purified PCR products from satDNAs TTE745 and dimers of the TTE35 primers after PCR with *Tamandua tetradactyla* genomic DNA...90

LISTA DE TABELAS

Introdução

Tabela 1 - Espécies de Xenarthra com genoma sequenciado e disponível no Genbank..........28

Capítulo 1

Artigo: Identification and Characterization of Satellite DNAs in Two-Toed Sloths of the Genus *Choloepus* (Megalonychidae, Xenarthra)

Table 1 - Putative satDNAs identified by RepeatExplorer 2 in *C. didactylus* and *C. hoffmanni*……………………………………...........……………………………..……32

Capítulo 2

Manuscrito: Comparative Analyses of Satellite DNAs in Six Xenarthra Species

Table 1 - Main features of the satDNAs from the SATCHO1 family……………................53

Table 2 - Identification of CENP-B box motifs in the SATCHO1 family………….........…..57

Supplementary File 1 - TAREAN basic results of satellite DNA identification in the genomes of five Xenarthra species……………………………………………....………...68

Supplementary File 2 - Matrix of pairwise identity percentage of each pair of SATCHO1 family……………………………………………………………………………………....71

Capítulo 3

Manuscrito: Identification and Characterization of the Most Abundant satDNAs in the Genomes of the Anteaters *Myrmecophaga tridactyla* and *Tamandua tetradactyla* (Myrmecophagidae, Xenarthra)

Table 1 - Must abundant satellite DNAs identified by TAREAN and their main features..…80

Supplementary File 1 - TAREAN results with all satDNAs identified……………………......89

LISTA DE ABREVIATURAS

2n Número Diploide
BLAST *Basic Local Alignment Search Tool*
Bp *Base Pairs*
BVA *Bradypus variegatus*
CBG *Padrões de bandas obtidos após tratamento com hidróxido de bário e coloração com Giemsa*
CDI *Choloepus didactylus*
CHO *Choloepus hoffmanni*
CVE *Chaetophractus vellerosus*
DNO *Dasypus novemcinctus*
DAPI 4′,6-diamidine-2′-phenylindole dihydrochloride
ECD *Evolutionarily Conserved Domain*
FISH Hibridação *in situ* Fluorescente, *Fluorescent in situ Hibridization*
FN *Fundamental number*
GTG Padrões de bandas obtidos após tratamento com tripsina e coloração com Giemsa
HOR H*igher-Order Repeat*
HAS *Homo sapiens*
IUCN *International Union for Conservation of Nature*
Kb *Kilobase*
LINEs L*ong Interspersed Nuclear Elements*
LTRs *Long Terminal Repeats*
Ma Milhões de anos
Mya *Million years ago*
NCBI *National Center for Biotechnology Information*
NF Número fundamental
NMDS *Non-metric multidimensional scaling*
Pb Pares de Base
PCR Reação em cadeia da polimerase, *polymerase chain reaction*
SatDNAs DNAs satélite, *satellite DNAs*
SINEs *Short Interspersed Nuclear Elements*
TMA *Tolypeutes matacus*

SUMÁRIO

1. **INTRODUÇÃO** 17

1.1 Xenarthra 17

1.2 DNAs Satélites 23

1.3 Citogenética dos Xenarthra – Sequências Repetitivas 26

2. **OBJETIVOS** 29

Objetivo Geral 29

Objetivos Específicos 29

3. **CAPÍTULO 1** 30

Identification and Characterization of Satellite DNAs in Two-Toed Sloths of the Genus *Choloepus* (Megalonychidae, Xenarthra) 30

4. **CAPÍTULO 2** 45

Comparative Analyses of Satellite DNAs in Six Xenarthra Species 45

5. **CAPÍTULO 3** 75

Identification and Characterization of the Most Abundant satDNAs in the Genomes of the Anteaters *Myrmecophaga tridactyla* and *Tamandua tetradactyla* (Myrmecophagidae, Xenarthra) 75

6. **CONCLUSÕES** 91

7. **REFERÊNCIAS** 92

1. INTRODUÇÃO

1.1 Xenarthra

Considerado um dos clados mais antigos entre os mamíferos atuais, a superordem Xenarthra é um táxon monofilético com bom suporte morfológico e molecular (Delsuc *et al.,* 2002; Superina e Loughry 2015; Moraes-Barros e Arteaga 2015). Estima-se que tenha surgido há cerca de 100 milhões de anos na América do Sul, onde também ocorreu sua diversificação (Delsuc 2004). Xenarthra é também considerado um dos maiores clados dentre os placentários (Viscaíno e Loughry 2008; Moraes-Barros e Arteaga 2015), graças ao vasto registro fóssil do grupo. Estima-se que durante o Paleoceno (~65Ma) as espécies de Xenarthra estavam divididas entre mais de 200 gêneros (McKenna e Bell 1997). Dentre esses animais, estavam as preguiças-gigantes, com tamanho comparável ao de elefantes, e os gliptodontes, parentes dos tatus, que são comparados em tamanho e formato ao carro Fusca. Apesar de se concentrarem na América do Sul, alguns grupos povoaram a América Central e a América do Norte, inclusive com registros fósseis de preguiças-gigantes no Alaska (Superina e Loughry 2015; Defler 2019).

O número de espécies de Xenarthra sofreu uma grande redução durante o Pleistoceno (O'Leary *et al*., 2013; Superina e Loughry 2015). Atualmente, a superordem está restrita a apenas 39 espécies, divididas entre duas ordens: Cingulata, que reúne 22 espécies de tatus, e Pilosa, composta por sete espécies de preguiças (Folivora) e dez espécies de tamanduás (Vermilingua) (Gardner 2008) (Figura 1). O número de espécies é provavelmente maior, já que novas espécies vêm sendo descritas. Por exemplo, com base em características morfológicas e moleculares, foram descritas seis novas espécies de tamanduá-i, genêro *Cyclopes* (Miranda *et al*., 2018), uma de tatu, *Dasypus mazzai* (Abba *et al*., 2018; Feijó *et al.*, 2018) e uma de preguiça-de-coleira (Miranda *et al.*, 2022) (https://xenarthrans.org/).

Os três clados de Xenarthra apresentam características morfológicas e ecológicas bastante distintas entre si. Essas variam desde carapaças ósseas nos tatus até pelos longos nas caudas de tamanduás. Ocupam hábitats semifossoriais, terrestres e arborícolas, e possuem alimentação variando desde insetos a pequenos mamíferos e vegetais. Apesar de apresentarem tantas características distintas, esses animais compartilham características morfofisiológicas que dão suporte ao clado. Eles possuem dentição reduzida, baixa temperatura corporal, garras, testículos internalizados e uma articulação especial que dá nome ao grupo – a articulação xenartróide (Viscaíno e Loughry 2008; Superina e Loughry 2015).

Os Xenarthra foram pouco estudados e há uma lacuna de conhecimento sobre dados populacionais, ecológicos, fisiológicos, comportamentais e morfológicos do grupo. Isso fica mais evidente quando os dados são comparados com os de outros eutérios, como os Euarchantoglires e Laurasiatheria, que possuem espécies com maior distribuição global (Viscaíno e Loughry 2008; Superina e Loughry 2015).

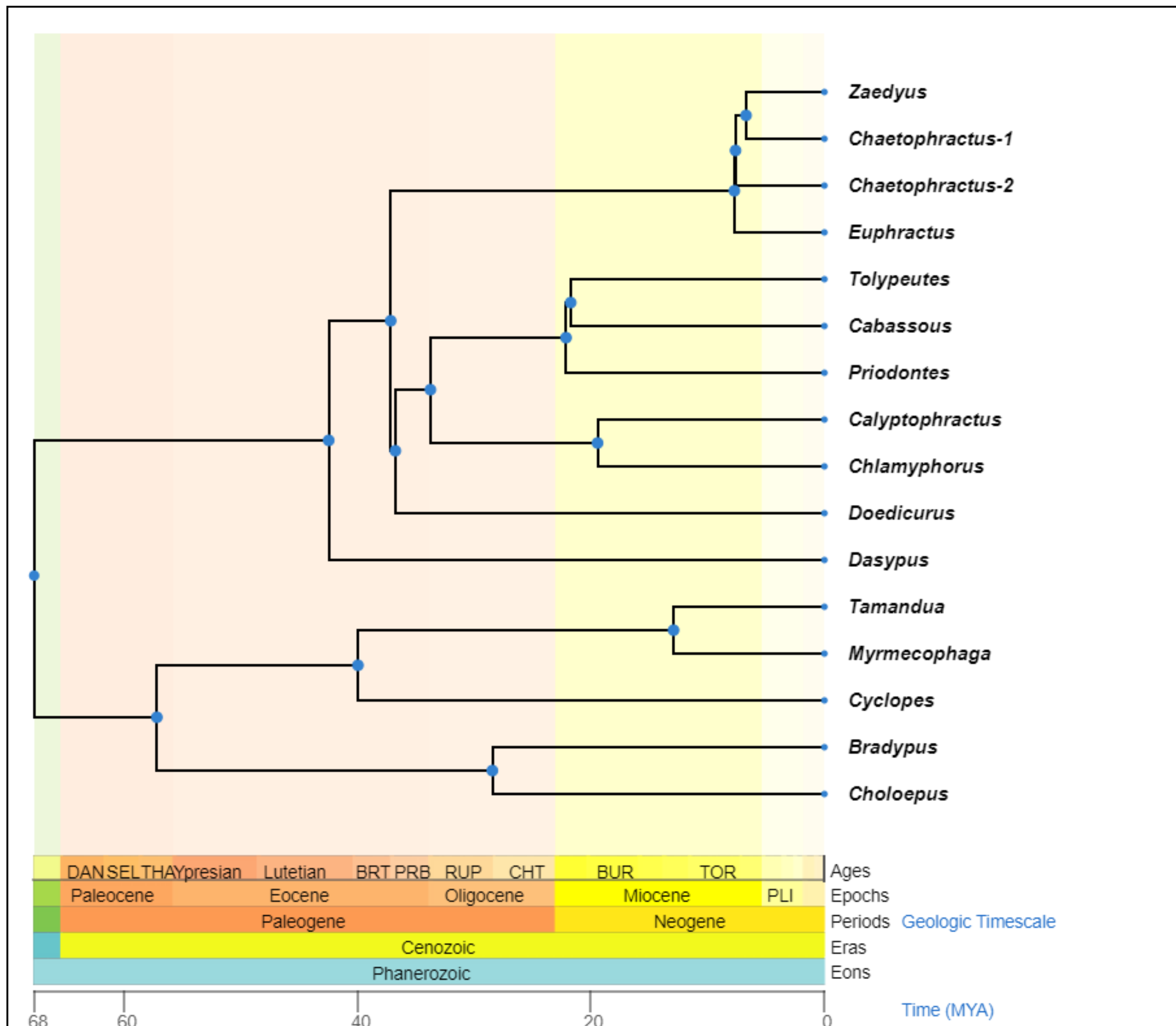


Figura 1: Filogenia dos gêneros de Xenarthra viventes, incluindo os tempos de divergência (adaptado de Gibb *et al*., 2016).

Espécies de Xenartra Estudadas na Presente Tese

Bradypodidae - *Bradypus variegatus*

A preguiça comum *B. variegatus* pertence à família Bradipodydae e apresenta a maior distribuição geográfica dentro desta família, ocorrendo em países da América do Sul e da América Central. Ocorre a partir do nível do mar até pelo menos 2.400m de altitude (Ureña *et al.*, 1986). No Brasil, ocorre nos biomas Amazônia, Cerrado e tem prevalência na Mata Atlântica. Ocorre também ao norte da Argentina, Bolívia, Colômbia, Equador, Peru, Venezuela - na América do Sul – e Nicarágua, Panamá, Costa Rica e Honduras - na América Central (Hayssen 2010; Miranda *et al.*, 2015).

Bradypus variegatus é simpátrica às espécies *B. torquatus* e *B. tridactylus*, e com as espécies de preguiças-de-dois-dedos, *Choloepus didactylus* e *C. hoffmanni* (Hayssen 2010; Miranda *et al.*, 2015). Por muitos anos, devido à sobreposição de áreas de ocorrência e variações de cor da pelagem, exemplares de *B. tridactylus* foram erroneamente classificados como *B. variegatus*. As principais diferenças morfológicas encontradas entre essas duas espécies estão na coloração da pelagem da garganta e a presença ou ausência de forames na região da nasofaringe. Enquanto em *B. tridactylus* a pelagem da garganta é amarelada/dourada e ocorre desde a raiz até as pontas do pelo, em *B. variegatus* essa pelagem é marrom, salvo por algumas populações no rio Tapajós que possuem esses pelos dourados, mas com a cor menos uniforme, com o dourado presente apenas na ponta dos pelos. *Bradypus variegatus* não possui forames distintos na região da nasofaringe anterodorsal, que estão presentes em *B. tridactylus*. *B. tridactylus*, *B. torquatus* e *B. variegatus* apresentam tamanhos semelhantes (~50cm de comprimento) e crânios mais robustos em comparação a *B. pygmaeus* (Heyssen 2010; Miranda *et al.*, 2015).

A lista vermelha de espécies ameaçadas da União Internacional para Conservação da Natureza (IUCN) categoriza *B. variegatus* como em situação menos preocupante por possuir ampla distribuição geográfica, com trânsito entre o Brasil e os países vizinhos nos quais também ocorre e por não serem identificadas grandes ameaças. Salvo exceção às regiões de ocorrência nas florestas da Colômbia e de Mata Atlântica no Brasil, devido à fragmentação do hábitat nessas regiões (Moraes-Barros *et al.*, 2022 – acesso em 06 de dezembro de 2023).

Megalonychidae – *Choloepus didactylus* e *Choloepus hoffmanni*

A família Megalonychidae é composta pelas preguiças-de-dois-dedos que são as representantes de um único gênero vivente da família – *Choloepus*. Este gênero possui duas espécies, *Choloepus didactylus* e *C. hoffmanni*, com áreas de distribuição geográficas distintas e algumas áreas de sobreposição na região Amazônica. *Choloepus didactylus* está restrita à América do Sul, ocorre em florestas tropicais úmidas da região amazônica e em florestas de altitude da região andina. *Choloepus hoffmanni*, por outro lado, possui uma distribuição disjunta, com duas populações diferentes. A primeira com ocorrência em Honduras, Nicarágua, Costa-Rica, Panamá, oeste da Venezuela, Colômbia e Equador. A segunda população encontra-se mais ao sul na região centro-norte do Peru, estados do Acre e sudoeste do Amazonas no Brasil e a região norte da Bolívia (Adam 1999; Hayssen 2011; Miranda *et al*., 2015).

Apesar das espécies de *Choloepus* apresentarem similaridades morfológicas que podem dificultar a correta identificação taxonômica, análises moleculares utilizando marcadores nucleares e mitocondriais permitem uma melhor diferenciação entre elas (Adam 1999; Hayssen 2011; Steiner *et al*., 2011). Apesar disso, normalmente são identificadas pelo padrão de coloração da pelagem e por algumas diferenças ósseas no crânio e nas vértebras. *C. didactylus* possui pelagem com coloração marrom-acinzentada, o topo da cabeça e ombros mais escuros e a face mais clara. Já *C. hoffmanni* apresenta uma coloração mais clara e dourada que *C. didactylus*, tanto no corpo quanto na face. *Choloepus hoffmanni*, geralmente possui menos vértebras cervicais e pré-caudais que *C. didactylus* e três forames penetram a região interpterigóide em *C. hoffmanni* (Wetzel e Ávila-Pires 1980; Wetzel 1985; Adam 1999; Maslin *et al*., 2007; Hayssen 2011; Miranda *et al*., 2015).

Embora a IUCN categorize as duas espécies de *Choloepus* como menor preocupação, as informações sobre tamanho e distribuição populacional das duas espécies são restritas e antigas (Chiarello *et al*., 2022; Plese *et al*., 2022 – acesso em 06 de dezembro de 2023). Essas espécies apresentam boa adaptação ao cativeiro, com exemplares distribuídos em zoológicos de diferentes países dentro e fora do continente americano (Miranda *et al*., 2015).

Mymecophagidae - *Myrmecophaga tridactyla* e *Tamandua tetradactyla*

Os tamanduás *Myrmecophaga tridactyla* (tamanduá-bandeira) e o *Tamandua tetradactyla* (tamanduá-mirim ou tamanduá-de-colete) são duas espécies representantes da família Myrmecophagidae. *Myrmecophaga tridactyla* é a maior espécie, facilmente distinguível das outras espécies de tamanduás por apresentar uma longa cauda coberta por

longos pelos. O tamanduá-bandeira possui em média 2m de comprimento, ou seja, é duas vezes maior que as duas espécies do gênero *Tamandua* (Gaudin *et al*., 2018). A coloração da pelagem também é diferente em relação às outras duas espécies da família. Enquanto *M. tridactyla* apresenta uma coloração castanho escura ou preta, as outras espécies possuem coloração marrom dourado mais clara, além de possuírem uma espécie de cinta escura, como um colete em volta pescoço e tronco anterior. *Myrmecophaga tridactyla* possui também o rostro maior e mais afilado e a caixa craniana maior que os *Tamandua*. Embora possua ampla distribuição geográfica, com áreas de ocorrência desde Honduras, na América Central, até as regiões norte da Argentina e sul da Bolívia, acredita-se que *M. tridactyla* esteja extinta na Nicarágua, Costa Rica, ao sul de Belize, norte da Argentina, sul do Brasil e na região leste do Paraguai (Gaudin *et al*., 2018).

Tamandua tetradactyla é uma espécie endêmica da América do Sul, ocorre em todo o território brasileiro, do sul ao norte do Uruguai, norte da Argentina e a leste da região andina da Colômbia, na Venezuela, em Trinidad e nas Guianas. Diferente de *T. mexicana*, *T. tetradactyla* possui orelhas mais longas, geralmente três pares de forames orbitais, em comparação aos quatro de *T. mexicana*, e um crescente incompleto na borda posterior do forame infraorbital (Wetzel 1985).

Segundo a IUCN, *M. tridactyla* é categorizada como espécie vulnerável, por apresentar perda de hábitat (Miranda *et al*., 2014b – acesso em 06 de dezembro de 2023), enquanto, *T. tetradactyla* é uma espécie de pouca preocupação (Miranda *et al*., 2014a - acesso em 06 de dezembro de 2023).

Chlamyphoridae - *Chaetophractus vellerosus* e *Tolypeutes matacus*

A família Chlamyphoridae é composta por oito gêneros e treze espécies divididos entre três subfamílias – Euphractinae, Chlamyphorinae e Tolypeutinae. *Chaetophractus vellerosus* faz parte da subfamília Euphractine, enquanto *Tolypeutes matacus* faz parte da subfamília Tolypeutinae. Dados moleculares apontam que Euphractinae e Tolypeutinae teriam se separado no Eoceno, há cerca de 33 milhões de anos (Gibb *et al*., 2015).

Chaetophractus vellerosus, também conhecido como tatu peludo-gritador, é endêmico da América do Sul com ocorrência em algumas regiões de três países – Bolívia, Paraguai e Argentina. É considerada a menor espécie e com maior pilosidade (tanto ventral quanto dorsalmente) entre os Euphractinae vivos (Carlini *et al*., 2016). A coloração do dorso varia

entre castanho variegado e castanho claro, possui orelhas maiores do que *Zaedypus pichiy* e *C. villosus*, de tamanho similar às de *C. nacionali*. *C. vellerosus* recebe o nome de tatu peludo-gritador porque grita bastante alto quando capturado, produzindo um som semelhante ao choro de uma criança (Wetzel 1982).

Tolypeutes matacus ou tatu-bola, também é uma espécie endêmica da América do Sul com ocorrência nas regiões de Cerrado e Chaco da Bolívia, na Argentina, Paraguai e no Brasil (Mato Grosso e Mato Grosso do Sul). É considerado um animal de pequeno porte que apresenta pequenas diferenças morfológicas com relação à outra espécie de tatu-bola existente – *T. tricinctus*. Esta última é endêmica do Brasil com distribuição restrita às áreas de Cerrado e Caatinga (Paglia *et al*., 2012). *T. matacus* possui quatro dedos nos membros anteriores, enquanto *T. tricinctus* possui cinco dedos e ambas as espécies possuem cinco dedos nos membros posteriores.

Chaetophractus vellerosus é considerada pela IUCN uma espécie com populações estáveis e de menor preocupação de risco de extinção (IUCN 2017 – acesso em 06 de dezembro de 2023). *Tolypeutes matacus*, apesar de ser considerada uma espécie que consegue se adaptar às mudanças de hábitat causadas pela agricultura, sofre com a caça e a perda de seu hábitat. Assim, no último levantamento realizado pela IUCN em 2014, essa pequena espécie de tatu foi considerada quase ameaçada (Noss *et al*., 2014 – acesso em 06 de dezembro de 2023).

Dasypodidae – *Dasypus novemcinctus*

A família Dasypodidae é composta por uma única subfamília com nove espécies de tatus, todas do gênero *Dasypus* (https://xenarthrans.org/). Dados morfológicos e moleculares apontam que a diversificação do grupo ocorreu há ~12Ma (Gibb *et al*., 2015). A maioria das espécies do gênero *Dasypus* ocorre apenas em países da América do Sul, e apenas *D. novemcinctus* ocorre também na América do Norte (https://xenarthrans.org/; McBee e Baker, 1982). Assim, é a espécie de Xenarthra com maior distribuição geográfica ocorrendo desde os Estados Unidos até o nordeste da Argentina, atingindo inclusive ilhas das pequenas Antilhas, Granada e Trinidad e Tobago (McBee e Baker, 1982). *D. novemcinctus* é bastante estudada por ser reservatório da bactéria *Mycobacterium leprae*, causadora da hanseníase (McBee e Baker, 1982; Zhou *et al*., 2021). Segundo a IUCN esta espécie de tatu está na lista de menor preocupação de risco de extinção (Abba e Superina, 2014 - acesso em 06 de dezembro de 2023).

1.2 DNAs Satélite

O genoma nuclear dos eucariotos é formado por sequências de cópias únicas e sequências repetitivas. Estas últimas, também denominadas DNA repetitivo, correspondem a uma parte considerável do genoma dos eucariotos, chegando a representar mais da metade do conteúdo de DNA nuclear em alguns grupos. Essas sequências podem ser encontradas dispersas pelo genoma ou repetidas em tandem (revisado em Charlesworth *et al.*, 1994; Biscotti *et al.*, 2015b). As sequências dispersas pelo genoma são divididas em duas categorias: a classe I ou retrotransposons usam um RNA intermediário durante os processos de amplificação e transposição pelo genoma, e a classe II ou transposons de DNA fazem o processo de transposição sem um intermediário de RNA (revisado em Finnegan 1989; Charlesworth *et al.*, 1994). Os principais representantes da classe I são os LINEs (*long interspersed nuclear elements*), SINEs (*short interspersed nuclear elements*) e os LTRs (*long terminal repeats*) (Charlesworth *et al.*, 1994; Wicker *et al.*, 2007; Biscotti *et al.*, 2015b) (Figura 2).

Já as repetições em tandem estão divididas entre os DNA ribossomais (DNAr) e as sequências satélites, que estão distribuídas em quatro categorias: microssatélites, minissatélites, macrossatélites e DNAs satélite (satDNAs) (revisado por Thakur *et al.*, 2021). As definições para essas categorias estão relacionadas ao tamanho das unidades de repetição (monômero), tamanho das cadeias e ao local onde essas cadeias estão localizadas nos cromossomos. Os micro e minissatélites possuem monômeros que variam de 1-10pb e 10-100pb, respectivamente. Os microssatélites (repetições simples) formam arranjos de 5-100 unidades de repetição e podem ser encontrados em qualquer região do genoma (Garrido-Ramos, 2017). Os minissatélites formam arranjos que variam de 2 até 100 monômeros por loco e podem ser localizados em diferentes regiões do genoma, mais comumente nas regiões teloméricas (Tautz 1993). Os macrossatélites formam monômeros bem maiores, que podem chegar a algumas quilobases de comprimento, ocupando grandes regiões dos genomas, expressam RNAs codificantes e não codificantes, são ricos em CG e por isso podem sofrer metilação (Dumbovic *et al.*, 2017). Os satDNAs são considerados sequências mais complexas, podem ter monômeros com algumas dezenas, centenas ou milhares de pares de base e podem formar estruturas em *higher-order* (HOR). Estruturas em HOR são formadas por múltiplos de unidades de repetição (monômeros) que se fundem formando monômeros

maiores. Estes são mais similares entre si do que as cópias que os formam. Essas múltiplas cópias são resultantes de processos de duplicação e homogeneização entre monômeros adjacentes. Os satDNAs formam grandes cadeias (>1kb), geralmente encontradas em regiões centroméricas, pericentroméricas e teloméricas, fundamentais para a integridade cromossômica (revisado por Garrido-Ramos 2017; Thakur *et al.*, 2021).

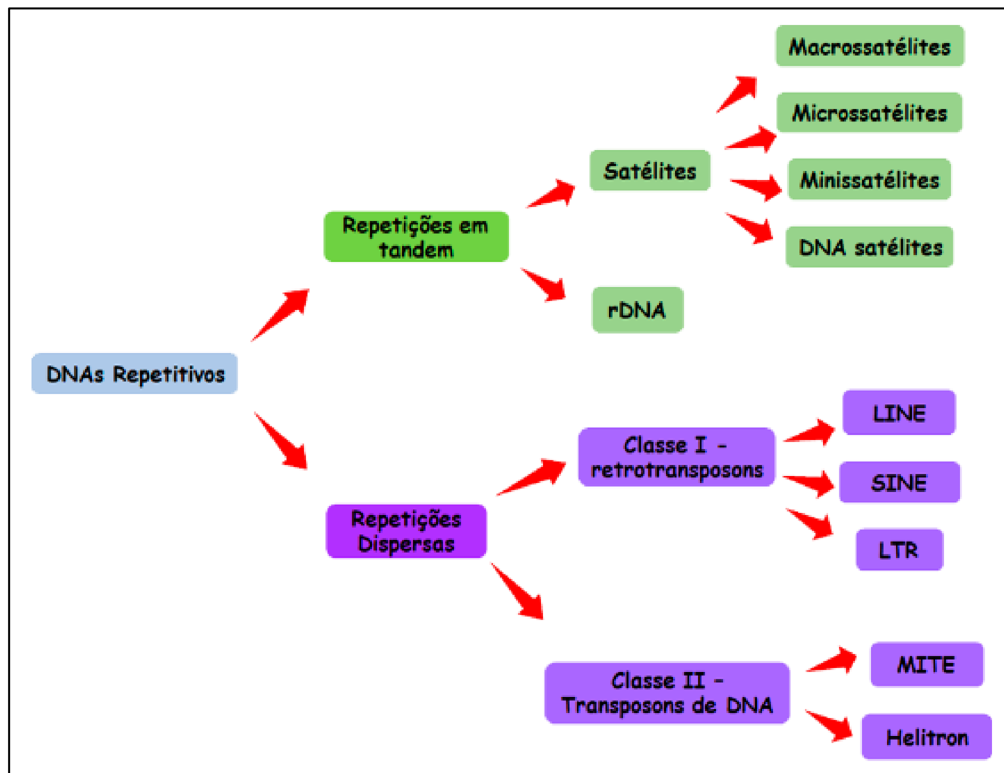


Figura 2: Representação esquemática dos DNAs repetitivos.

Os satDNAs são sequências genômicas com taxas rápidas de evolução molecular, que se homogeneizam dentro das cadeias através de mecanismos moleculares estocásticos, como a recombinação desigual, a transposição, a amplificação por círculo rolante e a conversão gênica (Dover 1982). Após a homogeneização dos monômeros dentro de um genoma, essas sequências podem se fixar dentro da população. Os processos de homogeneização e fixação dessas cadeias de repetições recebe o nome de Impulso Molecular (Dover 1982, 1986) e fazem com que as cópias de satDNA se tornem espécie-específicas, ou seja, podem ter uma alta divergência entre espécies filogeneticamente próximas, tornando-as bons marcadores taxonômicos e filogenéticos. Esse padrão de variação é conhecido como Evolução Combinada

(Dover 1982) (Figura 3). Existem famílias de satDNAs que são compartilhadas entre espécies próximas, mas podem corresponder a diferentes frações dos genomas, ocupar diferentes regiões cromossômicas, ou diferentes cromossomos das espécies.

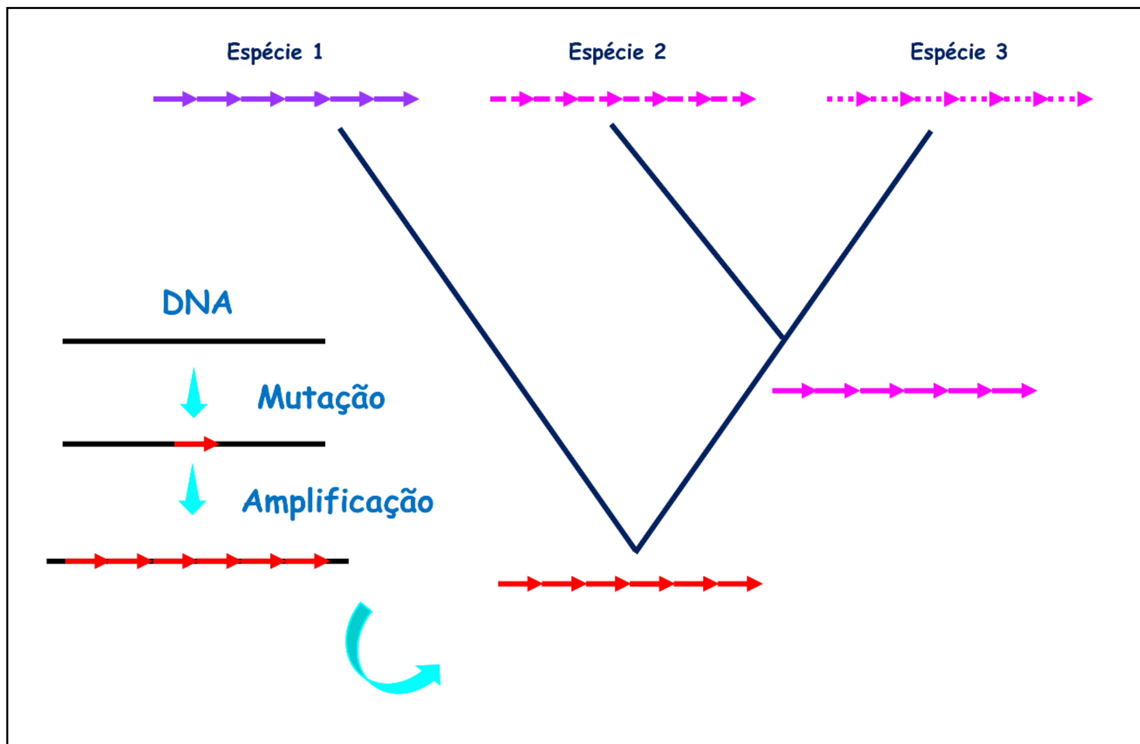


Figura 3: Representação esquemática do modelo de Evolução Combinada – Adaptação de Dover 1982. SatDNAs surgem após eventos de duplicações seguidas em tandem. As cópias em tandem sofrem mutações e podem ser homogeneizadas diferencialmente em diferentes espécies, como mostrado na figura. Consequentemente, as cópias de satDNA são mais similares dentro do que entre espécies.

Por muitos anos os satDNAs foram conhecidos como DNA lixo ou DNA egoísta (Ohno 1972; Palazzo e Gregory 2014). Acreditava-se que eram sequências sem funções nos genomas, que não eram transcritas, ou que seus transcritos não apresentavam nenhuma função (Palazzo e Gregory 2014). Porém, hoje se sabe que os satDNAs apresentam transcritos de RNAs não-codificantes (ncRNAs) estáveis, importantes para a organização, manutenção e integridade genômica, principalmente de centrômeros e telômeros (Biscotti *et al.*, 2015; revisado por Garrido-Ramos 2017; Thakur *et al.*, 2021). Também contribuem para a correta funcionalidade dessas fundamentais regiões heterocromáticas, com a formação do cinetócoro e correta segregação entre cromátides e cromossomos homólogos. Apesar de não serem

traduzidos, os ncRNAs desempenham outras funções biológicas importantes como: auxiliar na expressão e atividade de determinados genes, contribuir com modificações epigenéticas - notadamente em situações de estresse – e auxiliar na formação de estruturas centroméricas e teloméricas (revisado por Biscotti *et al*., 2015a; 2015b; Garrido-Ramos 2017; Thakur *et al*., 2021).

Os satDNAs são as sequências repetitivas menos estudadas (revisado por Thakur *et al*., 2021). Sua natureza altamente repetitiva e suas longas cadeias dificultam sua montagem após o sequenciamento dos genomas, principalmente quando as *reads* produzidas são curtas. Os cromossomos humanos são um dos exemplos mais marcantes. Apesar do Projeto Genoma Humano ter mais de 30 anos e alto investimento, apenas recentemente foi possível sequenciar um cromossomo humano de ponta a ponta. As tecnologias de sequenciamento e montagem de genomas não eram capazes de identificar e montar corretamente as regiões centroméricas dos cromossomos, em virtude da presença de satDNAs nos centrômeros, como o satDNA centromérico alfa dos cromossomos humanos. Apenas em 2020, utilizando a tecnologia Nanopore, capaz de sequenciar com alta qualidade *reads* com centenas de kilobases, Miga *et al*., conseguiram sequenciar e montar o cromossomo X humano de telômero a telômero. As tecnologias de sequenciamento de nova geração e de análises *in silico* permitiram progressos na identificação e estudo dos satelitomas de vários genomas (revisado por Garrido-Ramos 2017; Thakur *et al*., 2021). Os satelitomas são o conjunto dos satDNAs presentes num determinado genoma, essa denominação foi criada por Ruiz-Ruano *et al*. (2016). Há também programas como o satMINER, o RepeatExplorer e sua ferramenta TAREAN, que foram desenvolvidos para identificar satDNAs e têm sido amplamente usados para identificar satelitomas em genomas de animais e plantas (Novák *et al*., 2013; 2017; Ruiz-Ruano 2016; Pita *et al*., 2017; Silva *et al*., 2017; da Silva *et al*., 2020; Mora *et al*., 2020; Valeri *et al*., 2021).

1.3 Citogenética dos Xenarthra – Sequências Repetitivas

O estudo citogenético dos Xenarthra é ainda bastante deficiente e geralmente restrito à simples descrição dos cariótipos com coloração convencional, sem nenhum padrão de bandeamento (Viscaíno e Loughry 2008; Svartman 2012). A variação do número diploide é bem expressiva, indo de 2n=38 em *Tolypeutes matacus* até 2n=67 em *Choloepus didactylus*

(Jorge *et al.*, 1977; Steiner *et al.*, 2011, respectivamente). A família Megalonychidae apresenta cariótipos complexos, com variação do número diploide em *C. hoffmanni* de 2n=49 a 53, e em *C. didactylus,* de 2n=52 a 67. Foram descritas fêmeas X0, cromossomos B e translocações entre o cromossomo Y e alguns autossomos (Corin-Frederic 1969; Sonta *et al*., 1980; Jorge 1981; Jorge *et al*., 1978, 1985a; Dobigny *et al*., 2005; Benirschke 2006; Svartman *et al*., 2006; Steiner *et al*., 2011; Yu *et al*., 2012). Entre as espécies de *Bradypus*, apenas *B. pygmeus* não teve seu cariótipo descrito (Jorge 1982; 1985a; Jorge e Pinder 1990; Goldschmidt and Almeida 1993; Anderson e Handley 2001; 2002). A variação cariotípica observada entre as outras espécies de *Bradypus* é menos significativa: *B. tridactylus* possui 2n=52 (NF=64), *B. variegatus* possui 2n=54-55 (NF=56-58) e *B torquatus* possui 2n=50 e NF=64. Entre os tamanduás, as espécies do gênero *Tamandua* possuem 2n=54, com exceção de um indivíduo de *Tamandua* sp. analisado por Pereira *et al*., (2004), que apresentou um cariótipo com 2n=56. *Myrmecophaga tridactyla* possui 2n=60 (Rossi *et al*., 2014) e *Cyclopes didactylus* possui 2n=64 (Jorge *et al*., 1985). Das sete espécies de tatus da família Dasypodidae, pelo menos três (*D. hybridus*, *D. novemcinctus* e *D. septemcinctus*) já tiveram seus cariótipos descritos e apresentaram o mesmo número diploide (2n=64), porém com perceptível variação morfológica entre os cariótipos (Jorge e Pereira 2008). Já os tatus da família Chlamyphoridae possuem uma grande variação do número diploide (2n=38-62), incluindo o menor número diploide conhecido entre os Xenarthra, presente em *T. matacus* (2n=38). A outra espécie do gênero *Tolypeutes* (*T. tricinctus*) ainda não teve seu cariótipo descrito.

Das 39 espécies de Xenarthra atualmente descritas, apenas *B. torquatus*, *B. tridactylus*, *B. variegatus*, *C. dydactylus*, *C. hoffmanni*, *D. novemcinctus*. *Euphractus sexcinctus* e *T. tetradactyla* tiveram seus cariótipos analisados por pintura cromossômica com sondas de cromossomos humanos individuais (Svartman *et al*., 2006; Yang *et al*., 2006; Liu *et al*., 2011; Azevedo *et al*., 2012). Essa técnica é bastante útil para elucidar cariótipos ancestrais de grupos e para a compreensão de relações filogenéticas entre as espécies de clados.

Apenas oito espécies de Xenarthra tiveram seus genomas sequenciados (Tabela 1). Os estudos de identificação e caracterização de sequências repetitivas, sobretudo de DNAs satélite, também é bastante reduzido neste grupo. Foram identificados transposons do tipo LINE e SINE nas espécies *B. tridactylus*, *C. hoffmanni*, *D. novemcinctus*; *E. sexcinctus*, *M. tridactyla* e *T. tetradactyla* (Waters *et al*., 2004; Churakov *et al*., 2005; Nishihara *et al*., 2007;

Bao e Jurka, 2010). Apenas *Dasypus novemcinctus* e *Euphractus sexcinctus* tiveram satDNAs identificados em seus genomas (Alkan *et al*., 2011; Liu *et al*., 2011). Em *D. novemcinctus*, Alkan *et al*. (2011) identificaram o satélite centromérico 173 *in silico*, mapeado nas regiões centroméricas de todos os cromossomos. A identificação de duas possíveis sequências satélites em *E. sexcintus* foi feita através de ensaios enzimáticos, seguidos de amplificação, sequenciamento e mapeamento cromossômico. As duas sequências hibridaram nas regiões centroméricas dos cromossomos 3, 4, 7, 8, 10–13, 16–25 e nas regiões de heterocromatina constitutiva dos braços curtos e/ou longos dos cromossomos 3, 4, 7 e 19 (Liu *et al*., 2011).

Tabela 1: Espécies de Xenarthra com genoma sequenciado e disponível no Genbank.

Subordem	Família	Espécie	Número de Acesso
Folívora	Bradypodidae	*Bradypus variegatus*	GCA_004027775.1
	Maegalonychidae	*Choloepus didactylus*	GCA_004027855.1
		Choloepus hoffmanni	GCA_000164785.2
Vermilíngua	Myrmecophagidae	*Myrmecophaga tridactyla*	GCA_004026745.1
		Tamandua tetradactyla	GCA_004025105.1
Cingulata	Dasypodidae	*Dasypus novemcinctus*	GCF_000208655.1
	Chlamyphoridae	*Chaetophractus vellerosus*	GCA_004027955.1
		Tolypeutes matacus	GCA_004025125.1

2. OBJETIVOS

Objetivo Geral

Identificar e caracterizar de sequências *de novo* de satDNAs em genomas de espécies da superordem Xenarthra a fim de contribuir para a compreensão da evolução cromossômica e genômica deste grupo e de mamíferos placentários em geral.

Objetivos Específicos

1. Identificar e caracterizar os satelitomas dos genomas sequenciados e disponíveis de oito espécies de Xenarthra: *Bradypus variegatus*, *Choloepus didactylus*, *Choloepus hoffmanni*, *Myrmecophaga tridactyla*, *Tamandua tetradactyla*, *Chaetophractus vellerosus* e *Tolypeutes matacus* - usando as pipelines RepeatExplorer2 e TAREAN;
2. Mapear as famílias mais comuns de satDNAs identificadas pela pipeline RepeatExplorer2 e sua ferramenta TAREAN em cromossomos das espécies estudadas usando hibridação *in situ* fluorescente (FISH);
3. Avaliar quais famílias de satDNAs podem ser usadas como marcadores taxonômicos e filogenéticos em Xenarthra;
4. Verificar a possível relação entre satDNAs e centrômeros funcionais.

3. CAPÍTULO 1

Identification and Characterization of Satellite DNAs in Two-Toed Sloths of the Genus *Choloepus* (Megalonychidae, Xenarthra)

Radarane Santo Sena[1]; Pedro Heringer[1]; Mirela Pelizaro Valeri[1]; Valéria do Socorro Pereira[2]; Gustavo Campos Silva Kuhn[1]; Marta Svartman[1]*

[1]Laboratório de Citogenômica Evolutiva, Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

[2]Fundação de Parques Municipais e Zoobotânica, Belo Horizonte, MG, Brazil

Running Title: Satellite DNAs of two-toed sloths

*Corresponding author
E-mail: svartmanm@ufmg.br
Telephone number: 5531-34092612
Fax number: 5531-34092568

Abstract

Choloepus, the only extant genus of the Megalonychidae family, is composed of two living species of two-toed sloths: *Choloepus didactylus* and *C. hoffmanni*. In this work, we identified and characterized the main satellite DNAs (satDNAs) in the sequenced genomes of these two species. SATCHO1, the most abundant satDNA in both species, is composed of 117 bp tandem repeat sequences. The second most abundant satDNA, SATCHO2, is composed of ~ 2,292 bp tandem repeats. Fluorescence *in situ* hybridization in *C. hoffmanni* revealed that both satDNAs are located in the centromeric regions of all chromosomes, except the X. In fact, these satDNAs present some centromeric characteristics in their sequences, such as dyad symmetries predicted to form secondary structures. PCR experiments indicated the presence of SATCHO1 sequences in two other Xenarthra species: the tree-toed sloth *Bradypus variegatus* and the anteater *Myrmecophaga tridactyla*. Nevertheless, SATCHO1 only showed a satDNA pattern in the genus *Choloepus*. Our results reveal interesting features of the satDNA landscape in *Choloepus* species with the potential to aid future phylogenetic studies in Xenarthra and mammalian genomes in general.

Keywords: Satellite DNAs, Heterochromatin, CENP-B, Centromere

www.nature.com/scientificreports

scientific reports

OPEN

Identification and characterization of satellite DNAs in two-toed sloths of the genus *Choloepus* (Megalonychidae, Xenarthra)

Radarane Santos Sena[1], Pedro Heringer[1], Mirela Pelizaro Valeri[1], Valéria Socorro Pereira[2], Gustavo C. S. Kuhn[1] & Marta Svartman[1]✉

Choloepus*, the only extant genus of the Megalonychidae family, is composed of two living species of two-toed sloths: *Choloepus didactylus* and *C. hoffmanni*. In this work, we identified and characterized the main satellite DNAs (satDNAs) in the sequenced genomes of these two species. SATCHO1, the most abundant satDNA in both species, is composed of 117 bp tandem repeat sequences. The second most abundant satDNA, SATCHO2, is composed of ~2292 bp tandem repeats. Fluorescence in situ hybridization in *C. hoffmanni* revealed that both satDNAs are located in the centromeric regions of all chromosomes, except the X. In fact, these satDNAs present some centromeric characteristics in their sequences, such as dyad symmetries predicted to form secondary structures. PCR experiments indicated the presence of SATCHO1 sequences in two other Xenarthra species: the tree-toed sloth *Bradypus variegatus* and the anteater *Myrmecophaga tridactyla*. Nevertheless, SATCHO1 is present as large tandem arrays only in *Choloepus* species, thus likely representing a satDNA exclusively in this genus. Our results reveal interesting features of the satDNA landscape in *Choloepus* species with the potential to aid future phylogenetic studies in Xenarthra and mammalian genomes in general.

A significant part of eukaryotic genomes, ~30% in some plants to more than 50% in some insects and mammals, is composed of tandemly organized highly repetitive sequences, known as satellite DNAs (satDNAs) (reviewed in Ref.[1]). In general, satDNAs differ from other tandemly repetitive sequences by their organization, which consists of long arrays that can extend up to megabases in length. SatDNAs are major components of the constitutive heterochromatin present in fundamental chromosome structures, such as centromeres and telomeres (reviewed in Refs.[1,2]).

They also have been shown to be important components of chromosome organization, pairing, and segregation. For instance, their transcripts have been reported to participate in centromeric activity and genomic regulation[3–5]. Some satDNAs also have protein binding motifs such as the CENP-B motif which, together with the CENP-B protein, is known to be involved in kinetochore structuring by helping the assembly of the CENP-A protein in mammals[6–8]. Both the CENP-B protein and the CENP-B box motif are largely conserved in mammalian centromeres, but despite this broad conservation, the role of the CENP-B proteins is still poorly understood (reviewed in Ref.[8]).

Moreover, around 50% of some studied satDNAs have short inverted repeat (short dyad symmetry) sequences within their monomers, which have been reported as essential to chromatin structure and/or function[4,7,9–11]. Short dyad symmetry sequences have been identified in satellite DNA-free centromeres and in centromeric satDNAs which lack CENP-B boxes[7]. Those dyad symmetries are predicted to adopt non-B-form DNA structures such as cruciform, hairpins, triplexes, and single-stranded DNA, which are commonly identified in functional centromeres[4,7].

It is important to note that functional centromere sequences (those associated with CENP-A) are restricted to relatively short segments of DNA nested within megabase arrays of pericentromeric satDNAs, each of them having different epigenetic compositions[1,11]. Although pericentromeric satDNAs are involved in centromere maintenance and stability, the factors determining their boundaries and intrinsic differences with functional centromeric sequences are not fully known[1].

[1]Laboratório de Citogenômica Evolutiva, Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil. [2]Fundação de Parques Municipais e Zoobotânica, Belo Horizonte, MG, Brazil. ✉email: svartmanm@ufmg.br

Satellite DNA	C. didactylus		C. hoffmanni	
	SATCHO1	SATCHO2	SATCHO1	SATCHO2
Satellite confidence	High	Low	High	Low
Satellite probability	0.986	0.0425	0.992	0.0426
Consensus size	117 bp	2292 bp	117 bp	2292 bp
Genome proportion	13%	0.62%	2.6%	0.23%
AT content	58.97%	54.97%	58.97%	54.97%

Table 1. Putative satDNAs identified by RepeatExplorer2 in *C. didactylus* and *C. hoffmanni*.

SatDNAs are important components in the evolution of eukaryotic genomes. They can evolve three times faster than intergenic regions, which often results in significant differences between sequences, even among closely related species (reviewed in Ref.[1]). This rapid evolution is thought to be a consequence of mechanisms such as unequal crossing-over, gene conversion and replication slippage[12], which are all related with the process known as molecular drive, described by Dover[13]. Because new mutations are constantly spread by molecular drive, intraspecific satDNA arrays are often composed of very similar tandemly repeated sequences that have the potential to be used as species-specific markers.

The study of repetitive DNAs has been significantly advanced with the introduction of next-generation sequencing technologies and high-throughput in silico analyses of genomes (reviewed in Ref.[1]). One of the tools used in these studies is RepeatExplorer, a pipeline that identifies repetitive DNA sequences de novo in genomes, using the raw reads without the need of a reference library of known repetitive sequences[14]. This pipeline performs graph-based clustering analyses, identifying read similarities by comparing pairwise reads all-to-all, before grouping them into clusters.

Xenarthra is a basal eutherian group which originated and diversified entirely in South America[15,16]. With 31 recognized extant species, this superorder is divided into two orders: Cingulata, represented by armadillos; and Pilosa, composed by anteaters (Vermilingua) and sloths (Folivora)[17]. Despite its importance as a basal placental group, Xenarthra has been poorly studied in comparison with other mammals, mostly because of their strict geographic distribution and collection difficulty because of their natural behavior. Hence, more information about their ecology and genetics is essential to a better characterization of the group[16].

Studies on the repetitive DNA fraction of Xenarthra genomes have been mostly restricted to the identification of retrotransposon families. For instance, LINE (Long Interspersed Element) and SINE (Small Interspersed Element) families have been described in six species: the sloths *Choloepus hoffmanni* and *Bradypus tridactylus*[18,19], the anteaters *Tamandua tetradactyla* and *Myrmecophaga tridactyla*[18,20], and the armadillos *Dasypus novemcinctus* and *Euphractus sexcinctus*[18,21]. Currently, the only Xenarthra species with an identified satDNA sequence is the armadillo *D. novemcinctus*[22], which has a satDNA with ~ 173 bp monomers. Mapping by fluorescence in situ hybridization (FISH) revealed that this satDNA was present in the centromeres of all chromosomes in this species[22].

Two-toed sloths are the only extant representatives of the Megalonychidae family, composed by the single living genus *Choloepus*[23], with two species: *C. didactylus* and *C. hoffmanni*. Both species inhabit the tropical forests of South and Central America with a small overlap area of occurrence in the Amazon forest in Peru, southwestern Amazonas state and Acre state in Brazil. These two species can be differentiated mainly by morphological characters, such as pelage color[24], osteological features[25], the mitochondrial *COI* and *Cyt-b* genes, and the nuclear gene *Enamelin*[26,27]. Cytogenetic analyses of *Choloepus* have been mostly based on simple karyotypic descriptions without banding patterns[26,28–34]. These studies revealed a complex and confusing karyotypic scenario with significant variation in diploid numbers in *C. didactylus* (2n = 52–67) and less variation in *C. hoffmanni* (2n = 49–53), with translocations between the Y chromosome and different autosomes, occurrence of X0 females, and unpaired chromosomes described as B chromosomes.

In this work we identified and characterized the most abundant satDNA sequences from the *C. didactylus* and *C. hoffmanni* genomes using in silico methods. In addition, we mapped these sequences in the chromosomes of *C. hoffmanni*. This is the first study to identify, characterize and map satDNAs in sloths, revealing interesting aspects of the centromeric and repetitive fraction of their genomes.

Results

In silico identification and analysis of satDNAs. The RepeatExplorer2 analysis identified two abundant putative satDNAs in the *C. didactylus* and *C. hoffmanni* genomes, which we named SATCHO1 and SATCHO2 (Supplementary data 1 and 2) (Table 1). The analysis indicated differences in the proportion of satDNAs in the two species: the satDNA content represents > 13% of the *C. didactylus* genome, whereas this value is approximately 3% in *C. hoffmanni*.

SATCHO1, the most abundant satDNA sequence in both species, has ~ 117 bp monomers, low levels of inter-repeat nucleotide variability (~ 3% on average) and AT content of ~ 59%. This satDNA represents 13% of the *C. didactylus* and 2.6% of the *C. hoffmanni* genomes. SATCHO2 is the second most abundant satDNA and has ~ 2292 bp monomers, inter-repeat nucleotide variability of ~ 24% on average and AT content of ~ 55%. It corresponds to 0.62% and 0.23% of the *C. didactylus* and *C. hoffmanni* genomes, respectively. Although SATCHO1 and SATCHO2 sequences are abundant in both genomes, we did not identify similar sequences in any

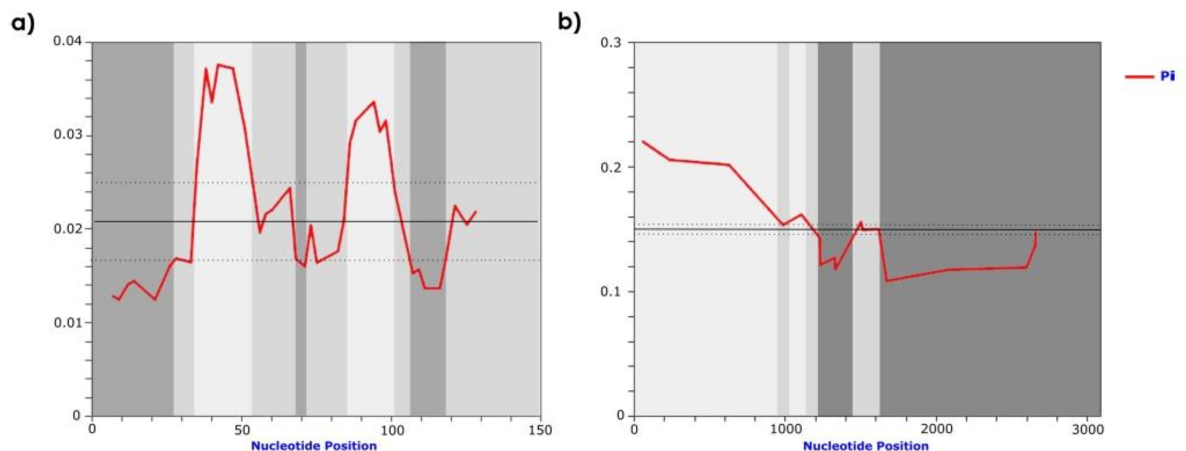


Figure 1. Identification of conserved (dark grey) and variable (light grey) satDNA segments of *C. didactylus* and *C. hoffmanni* by sliding window analysis. Sliding window of 10 bp for (**a**) SATCHO1 and (**b**) SATCHO2. Nucleotide diversity (Pi) is indicated by the red line, average nucleotide diversity is indicated by the solid line, and average diversity ± 2 SD is indicated by the dotted line.

other species on Repbase or in searches against all sequences from the non-redundant nucleotide collection in Genbank (accessed in 03/01/2020).

The analysis of nucleotide variability along both satDNAs revealed the presence of conserved regions within their monomers (Fig. 1), even though satDNAs are expected to evolve neutrally, revealing regions under potential selective constraints.

Phylogenetic and NMDS analyses. In order to infer the interspecific similarity between copies of SATCHO1 and SATCHO2 in *C. didactylus* and *C. hoffmanni*, we constructed a Neighbor-Joining tree using a sample of copies from each satDNA. The resulting tree showed that satDNA copies from both species are very similar and did not segregate into species-specific branches for SATCHO1 and SATCHO2 sequences (Fig. 2a,c).

We also estimated the pairwise distance values of the same set of sequences to generate NMDS ordinations for their Euclidean distances. The results also did not reveal any clear topological segregation between copies from each species (Fig. 2b,d). Nevertheless, each satDNA appeared to evolve under distinct evolutionary rates, as evidenced by their heterogeneous distribution across the NMDS ordinations.

Overall, both analyses indicate that the satDNA sequences from *C. didactylus* and *C. hoffmanni* have not diverged enough to segregate into species-specific clusters.

Chromosome mapping of SATCHO1 and SATCHO2. The *C. hoffmanni* individual we studied presented a karyotype with a diploid number 2n = 51. GTG-banding allowed the identification of all chromosome pairs and of an odd chromosome, which we identified as a B chromosome (Fig. 3a). The CBG-banding revealed the presence of constitutive heterochromatin in the centromeric regions of all chromosomes, except the X (Fig. 3b).

Our specimen has the same karyotype described earlier by Svartman et al.[34] for *C. hoffmanni* (2n = 50), from which it differs by the presence of the extra odd chromosome and by an inversion in pair 3 (metacentric in our specimen and acrocentric in the one previously described).

SATCHO1 and SATCHO2 were both FISH mapped in the centromeric regions of all *C. hoffmanni* chromosomes, except the X (Fig. 4), coinciding with the constitutive heterochromatin revealed after CBG-banding (Fig. 3b). This finding suggests that both satDNAs could play a functional role in the centromeres of *C. hoffmanni*.

Centromeric features of SATCHO1 and SATCHO2. Because SATCHO1 and SATCHO2 were located in the centromeric regions of *C. hoffmanni* chromosomes, we searched for putative CENP-B box-like motifs within these satDNA sequences. These motifs are typical of mammalian centromeric sequences and are thought to associate with kinetochore proteins[6,35,36]. We found that SATCHO1 has a motif with 5 of the 9 conserved nucleotides present in the evolutionary conserved domain (ECD) box (TTCGNNNNANNCGGG)[22,37], having 73% of overall similarity with its canonical structure and sharing 59% sequence similarity with the human CENP-B box (Fig. 5). Interestingly, this putative CENP-B box-like motif from SATCHO1 overlaps with the conserved region identified by DnaSP analysis on its distal portion (Fig. 1a). In the SATCHO2 sequence, we identified two segments separated by ~ 140 bp which together form a putative CENP-B box-like motif (Fig. 5). These segments however constitute a broken motif and are thus unlikely to compose a functional sequence.

We also found some small palindromic sequences with 4–5 bp on both satDNAs (Fig. 6). As we have mentioned, these dyad symmetries have the potential to form secondary DNA structures which are commonly found on functional centromeric sequences. Indeed, the analysis of nucleic acid folding prediction showed that several

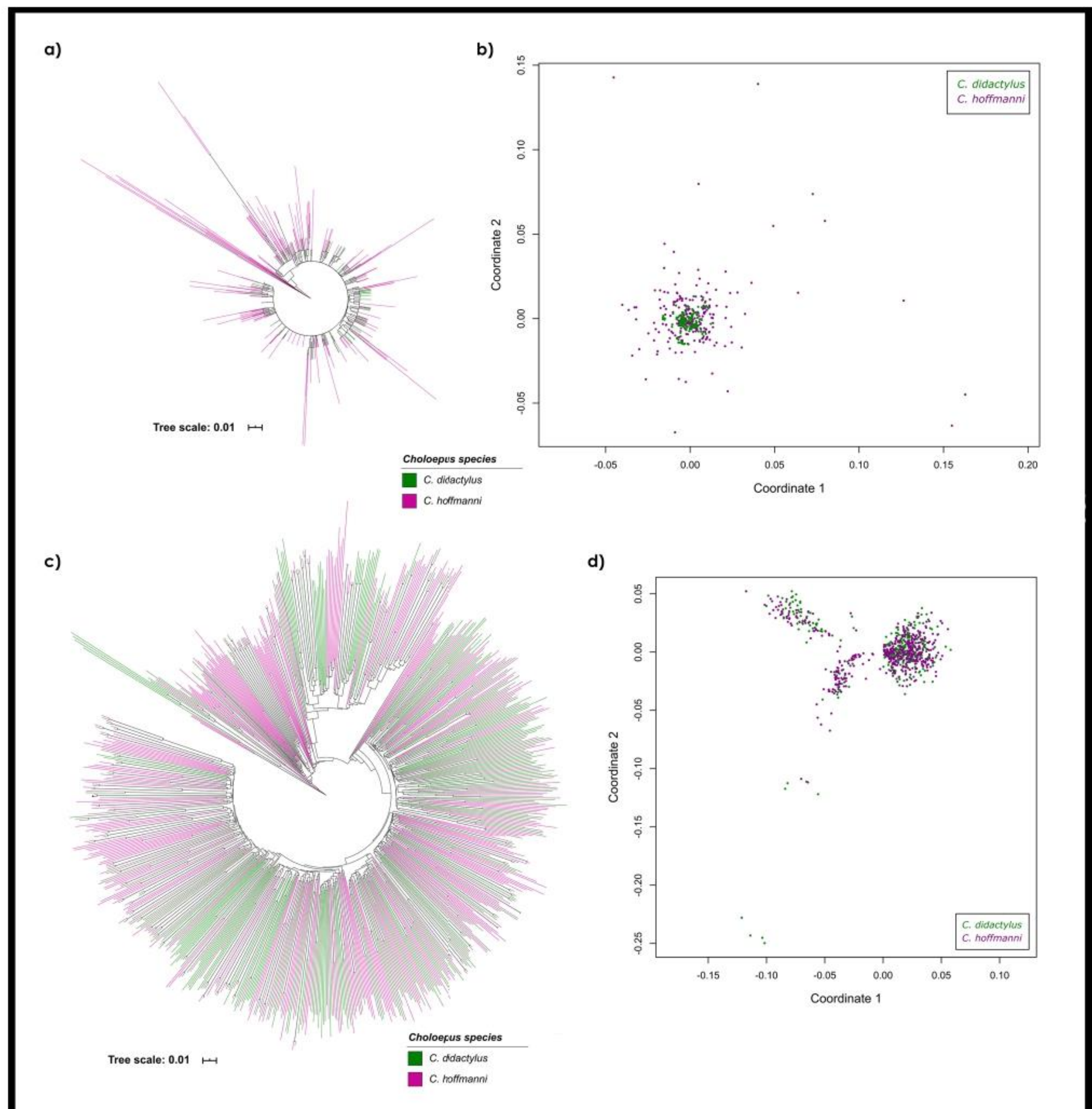


Figure 2. Comparative phylogenetic analyses of (**a**) SATCHO1 and (**c**) SATCHO2 sequences between *C. didactylus* and *C. hoffmanni* inferred by the Neighbor-Joining method with 1000 bootstraps. Minimum bootstrap support is 50%. Non-metric Multidimensional Scaling (NMDS) of evolutionary divergence among (**b**) SATCHO1 and (**d**) SATCHO2 sequences between *C. didactylus* and *C. hoffmanni*. The ordinations in (**b**) and (**d**) represent Euclidian distances for four dimensions. Each color represents sequences from one *Choloepus* species: *C. didactylus* (green) and *C. hoffmanni* (magenta).

segments within SATCHO1 and SATCHO2 have the potential to form stable DNA secondary structures (Fig. 7). These results indicate that both satDNAs contain structural hallmarks of functional centromeric sequences.

SATCHO1 and SATCHO2 in other Xenarthra. In order to verify if the satDNAs identified in *Choloepus* are also present outside the genus, we conducted Blastn searches against assembled Xenarthra genomes. For SATCHO1 we got hits in multiple contigs of the assembled genomes of *B. variegatus*, *M. tridactyla* and *T. tetradactyla*. However, the maximum number of tandemly repeated copies retrieved in a single contig was 42 on *B. variegatus*, and 3 on *M. tridactyla* and *T. tetradactyla*. In contrast, searches on both *Choloepus* species returned hundreds of contigs including considerable results, with some of them having up to 295 tandemly repeated copies of SATCHO1.

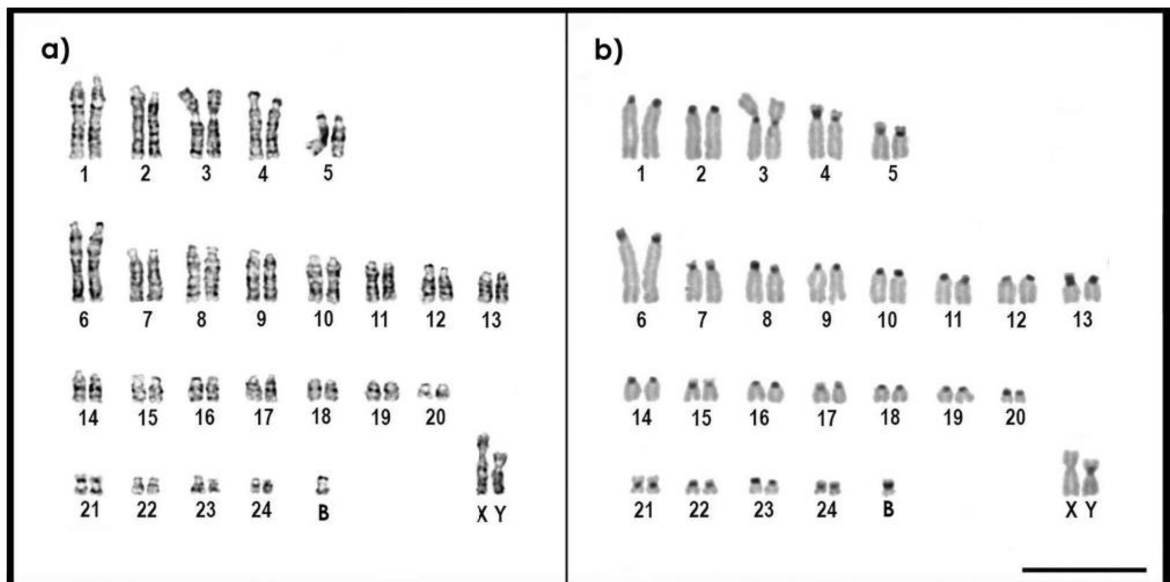


Figure 3. Karyotype of *Choloepus hoffmanni* (2n = 51) after: (**a**) GTG-banding and (**b**) CBG-banding. Bar = 10 µm.

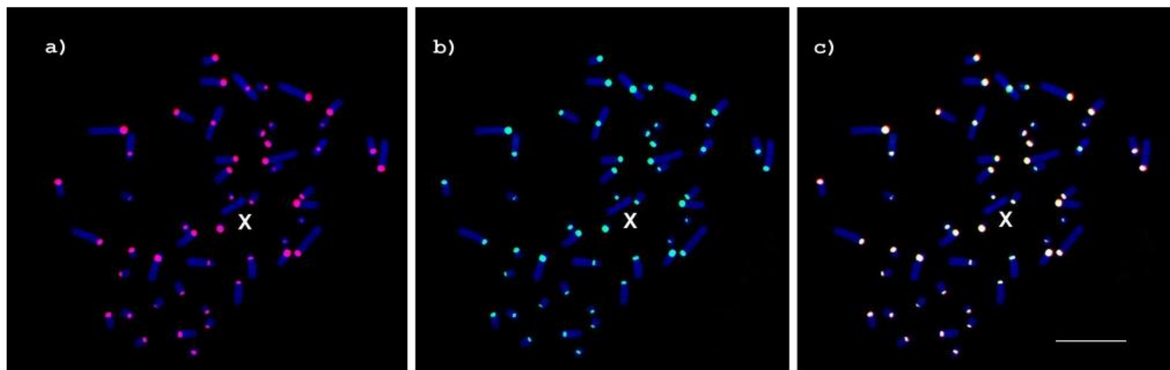


Figure 4. Metaphases of *Choloepus hoffmanni* after FISH using (**a**) SATCHO1 (red) and (**b**) SATCHO2 (green) as probes. (**c**) Merged signals from (**a**) and (**b**). Chromosomes were counterstained with DAPI. Note the signals in the centromeric regions of all chromosomes, except the X. Bar = 10 µm.

```
SATCHO1 C T T T T T T G A A A G C A G T C
SATCHO2 A T T C A C T G G A A A C G G A A
HSA     C T T C G T T G G A A A C G G G A
ECD     N T T C G N N N N A N N C G G G N
```

Figure 5. CENP-B motifs identified in sequences from SATCHO1 and SATCHO2 aligned with *Homo sapiens* (HSA) and the evolutionary conserved domain (ECD). In red: conserved bases compared with HSA. In yellow: conserved bases compared with ECD. The CENP-B motif found in SATCHO2 is composed of two different fragments separated by ~ 140 bp, as indicated by the green symbol.

Blastn searches on different assembled Xenarthra genomes using SATCHO2 as a query returned hits in multiple contigs only in *B. variegatus*. However, we only found up to three tandemly repeated copies in this species. In the genus *Choloepus* however, Blastn searches retrieved hundreds of contigs with hits, and up to 60 tandemly arranged copies in a single contig. Interestingly, although SATCHO1 and SATCHO2 have a centromeric localization, we did not find contigs including both satDNA sequences in none of our Blast searches.

Figure 6. (**a**) SATCHO1 and (**b**) SATCHO2 sequences with dyad palindrome sequences. Each palindrome pair is represented by the blue color and the direction arrow above them.

We also performed PCR experiments using the SATCHO1 and SATCHO2 primers in the genomic DNAs of the three-toed sloth *B. variegatus* and the giant anteater *M. tridactyla*. SATCHO1 homologous sequences were amplified from both species (Supplementary Fig. 1), which was confirmed by cloning and sequencing. The two sequenced copies from *B. variegatus* showed an average of ~ 2% nucleotide divergence from the *Choloepus* SATCHO1 consensus sequence, whereas the two copies of *M. tridactyla* presented an average of ~ 1% nucleotide divergence. FISH with the SATCHO1 probe in *M. tridactyla* chromosomes did not produce any signal (data not shown). The SATCHO2 sequence did not amplify by PCR with the genomic DNAs of neither *B. variegatus* nor *M. tridactyla* (Supplementary Fig. 2).

These results suggest that, although SATCHO1 and SATCHO2 are present outside the genus *Choloepus*, these sequences are not distributed as abundant long arrays of tandem repeats in other Xenarthra genera, in which they should not be classified as satDNAs.

Discussion

In this work we identified two novel centromeric satDNAs in the genomes of *C. didactylus* and *C. hoffmanni*, which could potentially have a centromeric function. Although both species have the same satDNAs, the results from RepeatExplorer revealed some marked differences in the genome proportion of these sequences in each species (Table 1). It is important to note that both species have approximately the same genome size (~ 3.3 Gb) as indicated by their sequencing projects (*C. didactylus* accession: GCA_004027855.1, *C. hoffmanni* accession: GCA_000164785.2). Despite the possibility that these observed differences reflect a real interspecific variation, it is also likely that they constitute artifacts derived from distinct values of genome coverage and/or sequencing platforms used for each species (Illumina HiSeq 2000 for *C. hoffmanni*, and Illumina HiSeq 2500 for *C. didactylus*). Although it is currently not possible to rule out any of these possibilities, the high sequence similarity and comparable number of Blastn results in both satDNAs between species indicate that a real large difference in abundance is unlikely. Indeed, a recent study demonstrated that different sequencing platforms, or even different versions of the same platform, have their own biases in representing the true proportion of highly abundant repeats[38].

Our phylogenetic and NMDS results revealed that both satDNAs do not segregate into different branches in a species-specific manner. That result was unexpected, considering that satDNAs usually evolve rapidly through the process of molecular drive, which also tends to produce a high degree of intra-species sequence homogeneity[13]. Hence, this high level of sequence identity could be explained by one or more of the following hypotheses: (i) *C. didactylus* and *C. hoffmanni* share a very recent common ancestor; (ii) they display a slow rate of molecular evolution; (iii) they went through a recent process of hybridization; (iv) or that these satDNAs sequences have been conserved by selective pressures. Regarding the first possibility, previous molecular data from different studies showed considerable variation in the estimated divergence between *C. didactylus* and *C. hoffmanni*. For instance, using the mitochondrial gene *Cyt-b*, the split of the two *Choloepus* species was estimated at ~ 18.7 Mya with Bayesian inference and ~ 5.8 Mya with Median Joining Network[27]. Gibb et al.[39] estimated the split varying from 3.5 to 16.7 Mya, based on mitogenomic shotgun data with Bayesian and maximum likelihood phylogenetic inferences. Hence, these estimates of divergence times argue against the hypothesis of a very recent common ancestor of *C. didactylus* and *C. hoffmanni*. In relation to the second hypothesis, *Choloepus* species have been shown to display a relatively slow rate of molecular evolution when compared to other Xenarthra groups[39], although the reason for that is not fully understood. However, even considering that a slower rate of molecular evolution could partially explain the high sequence identity found between these satDNAs, it does not seem likely that sequences evolving neutrally would keep this level of conservation after several million years. As to the third possibility, it is worth mentioning that the two *Choloepus* species inhabit some overlapping areas of the Amazon forest and there is no precise information about the collecting areas of most specimens studied[27]. Hence, the chance of interspecific hybridization cannot be ruled out. Finally, the hypothesis that SATCHO1 and SATCHO2 could have been conserved by selective pressures is currently more difficult to evaluate beyond the evidence we provided for a putative centromeric function, as its likelihood also depends on the exclusion of the first three possibilities. Nevertheless, considering all the evidence provided here and elsewhere, we suggest that the sequence conservation of these satDNAs between *C. didactylus* and *C. hoffmanni* likely derive from a combination of selective pressures and a slow rate of molecular evolution.

More importantly, our results revealed that both satDNAS are located in the centromeric regions of all *C. hoffmanni* chromosomes, except the X (Fig. 4a). It has been suggested that the most abundant tandem repeat

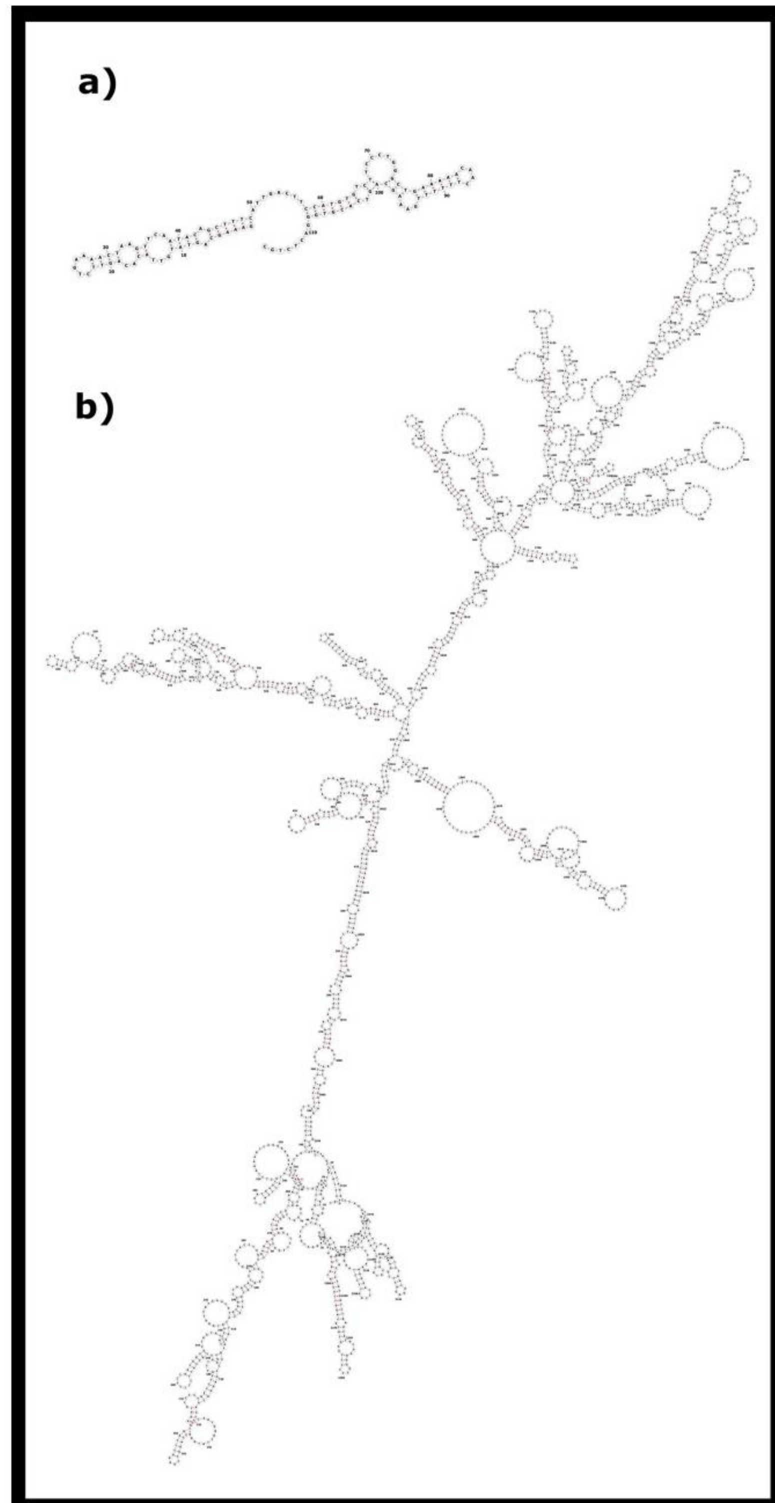


Figure 7. The optimal secondary structure of (**a**) SATCHO1 and (**b**) SATCHO2 predicted by RNAfold.

in a given genome likely corresponds to its centromeric sequence[40], a feature that was observed for SATCHO1 in *C. hoffmanni*, and presumably also in *C. didactylus*. Although *C. hoffmanni* had the two satDNAs mapped to centromeric regions, the resolution of our results does not enable us to determine how they are distributed along the centromeric heterochromatin and if this distribution varies among chromosomes. As we have mentioned, it is also not possible to determine if one or both satDNAs are part of the functional centromere. Further analyses using long sequencing reads, chip-seq with CENP-A antibodies and immuno-fiber FISH experiments would be important to address these issues.

In addition, we found conserved regions in SATCHO1 and SATCHO2, which include motifs sharing similarities with CENP-B box-like sequences (Fig. 5). Although the CENP-B box-like motif of SATCHO2 is disrupted by an intruding sequence, and thus is probably non-functional, its presence indicates that this large satDNA might have been previously involved in centromeric activity during the evolution of *Choloepus*. It is also possible that SATCHO2 currently has a secondary centromeric function, unrelated with the activity carried out by satDNAs containing CENP-B box-like sequences. In any case, the conservation of such regions in these satDNAs suggests that they could be under some sort of selective constraint. The fact that SATCHO1 and SATCHO2 also have an enrichment of symmetric sequences capable of forming non-B DNA forms and secondary structures argues for their putative centromeric function, as these nucleotide arrangements are thought to interact with centromere components[7,11].

Taken together, our data suggest a putative functional role for these satDNAs, which would explain their centromeric localization in *C. hoffmanni* and remarkable conservation in both *Choloepus* species. Similar results were reported in rodents of the genus *Peromyscus*, in which the centromeric satDNA PMsat was found in the centromeres of seven species[41]. Similarly to our results, PMsat monomers presented small sequence variation and shared similarities with the human CENP-B box-like motif. Based on these observations, the authors suggested that PMsat may have played some biological role which led to its maintenance in *Peromyscus*[41].

Another interesting finding of our study is that SATCHO2 is composed by ~ 2292 bp monomers, an uncommonly large size for a satDNA sequence. Most satDNAs identified in plants and animals showed monomer lengths around 150–180 bp and 300–360 bp, respectively[42,43]. There is a limited number of species in which satDNAs with monomers ranging from 1 kb to ~ 2 kb have been reported. That is the case of some whales[44], South American monkeys[45], banana[46], non-domestic Bovidae[47], and the field bean[48]. SatDNA monomers larger than 2 kb have been identified in bovines[49] and in the ant *Monomorium subopacum*[50]. To our knowledge, the only examples of monomers significantly larger than SATCHO2 were reported in Bovidae: the satDNA 1.709 (*SATIV*) with ~ 3.8 kb and the satDNA 1.711b with ~ 2.6 kb[49,51].

Finally, several studies have demonstrated that satDNAs, especially those found in centromeres, are associated with Robertsonian translocations, the main chromosome rearrangements related to Bovidae genome evolution[5,52–55]. It would be interesting to investigate if there is also a link between satDNAs and chromosome rearrangements in Xenarthra, as the number of available genomes of this group will certainly increase in the near future.

Materials and methods

Identification and analysis of satDNA sequences. In order to identify the most abundant satDNA sequences in the genomes of *Choloepus* species we performed a graph-based clustering analysis of sequence reads using the pipeline RepeatExplorer2[14]. The analysis was performed in a set of 357,044 random sampled reads (~ 1.19% genome coverage) from the *C. didactylus* genome (accession: SRX4501348) and 789,160 random sampled reads (~ 2.6% genome coverage) from the *C. hoffmanni* genome (accession: SRX282195). Identified satDNA consensus sequences were used as queries in searches conducted on Repbase[56] and GenBank (https://www.ncbi.nlm.nih.gov/genbank/) in order to detect similarities with previously described sequences. To analyze the satDNA copies directly in the species genomes, we retrieved a sample of each satDNA sequences from the *C. didactylus* (accession: PVKG000000000.1) and *C. hoffmanni* (accession: ABVD00000000.2) assembled genomes available on GenBank using Blastn searches with default parameters[57]. The software DnaSP 6.12.03[58] was used to identify DNA polymorphisms and nucleotide diversity along the satDNA sequences, by applying a window size of 10 bp (SATCHO1 and SATCHO2) and a step size of 2 bp for SATCHO1 and 3 bp for SATCHO2. Windows that exhibited standard deviation (S.D.) values ≥ 2, from the average variability, were considered highly variable, while those with values ≤ 2 S.D. were considered conserved.

We searched putative CENP-B box-like motifs (CTTCGTTGGAAACGGGA)[36] on the SATCHO1 and SATCHO2 monomer sequences using the alignment algorithm MUSCLE[59] on MEGA7[60]. We also searched for dyad symmetries in the satDNA sequences using the EMBOSS palindrome software[61] with a minimum palindrome length of 4 bp and maximum gap between elements of 20 bp. We used the RNAfold web server (https://rna.tbi.univie.ac.at/)[62] to search for optimal secondary structure with minimum free energy on the SATCHO1 and SATCHO2 sequences.

Pairwise evolutionary distances within each satDNA sequence from *C. didactylus* and *C. hoffmanni* were estimated using MEGA7[60]. The values were used to obtain non-metric multidimensional scaling (NMDS) ordinations with the R package Vegan[63], representing Euclidian distances in four dimensions. We used Rstudio v1.1.442[64] to conduct the NMDS analysis and plotting of the results. We constructed a phylogeny of the sequences using the Neighbor-Joining method with 1000 replicates on MEGA7[60]. The phylogenetic tree was edited using iTOL4.4.1 (https://itol.embl.de/)[65].

Biological samples. Chromosome preparations and genomic DNAs were obtained from cultured fibroblasts of *C. hoffmanni* and *M. tridactyla* male individuals. Tissue and blood samples from both specimens were obtained from Fundação de Parques Municipais e Zoobotânica de Belo Horizonte/MG, Brazil, under a license

from IBAMA (Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis) conceded to M. Svartman (Process Sisbio 28422-5). The *C. hoffmanni* individual came from an unknown location in Rondônia estate, Brazil, and the *M. tridactyla* individual was apprehended by IBAMA in Esmeraldas, Minas Gerais, Brazil, but its origin is unknown. We also used the genomic DNA from a male *B. variegatus* captured in Teófilo Otoni, Minas Gerais, Brazil.

Cell cultures and chromosome spreads were obtained according to Stanyon and Galleni[66] and genomic DNAs were obtained with the Wizard Genomic Purification kit (Promega).

Molecular analysis. The identified satDNAs were amplified by polymerase chain reaction (PCR) from *C. hoffmanni* genomic DNA with the following primers designed from the consensus sequences generated on RepeatExplorer: SATCHO1-F (AGTTGTTTTTCAGCCCAGGG) and SATCHO1-R (CACGTGGGACTC TGCGAAAG); SATCHO2-F (TCTCACCCGGATCTGAACCT) and SATCHO2-R (GGATACGGGGGTTTG AAGCA). The thermocycling conditions were as follows: 95 °C-5 min, 30 cycles: 95 °C-1 min; 53.4 °C-1 min; 72 °C-1 min; final elongation: 72 °C-10 min. The PCR products were extracted from a 1% agarose gel, purified with Wizard SV Gel and PCR Clean-up System kit (Promega), and cloned into a plasmid vector pGEM-T-Easy cloning kit (Promega). The recombinant plasmids were sequenced with the ABI 3730 platform (Applied Biosystems). The sequences obtained have GenBank accession numbers: MT505303–MT505310.

Banding patterns and fluorescence in situ hybridization (FISH). The GTG- and CBG-banding of *C. hoffmanni* chromosomes were performed according to Seabright[67] and Sumner[68], respectively. FISH was performed using the cloned satDNA sequences as probes after they were labeled by nick-translation with digoxigenin-11-dUTP (DIG-Nick Translation mix, Roche Applied Science) for SATCHO1 and biotin-16-dUTP (Biotin-Nick Translation mix, Roche Applied Science) for SATCHO2. The probes (~ 150 ng in 50% formamide/2xSSC) were denatured for 10 min at 98 °C. Chromosomes were dehydrated in ethanol series (70%, 90%, 100%) and denatured in 70% formamide/2xSSC for 2 min at 75 °C. The hybridization was performed overnight at 37 °C. Post-hybridization washes comprised two 5 min incubations in 2xSSC at 45 °C. Immunodetection was performed with anti-digoxigenin conjugated with rhodamine (Roche Applied Science) for SATCHO1 and avidin conjugated with FITC (Roche Applied Science) for SATCHO2. Chromosomes were counterstained with DAPI 1:500 in Slowfade (Invitrogen). The analysis was performed under a Zeiss Axioimager 2 epifluorescence microscope adapted with a CCD camera and image acquisition was performed with the AxioVision (Zeiss) software (Carl Zeiss MicroImaging, Jena, Germany).

Verification of SATCHO1 and SATCHO2 in other Xenarthra species. To verify the presence of the identified satDNAs in other Xenarthra species, we conducted Blastn searches on all assembled Xenarthra genomes except *Choloepus*, using SATCHO1 and SATCHO2 consensus sequences as queries. We also performed PCRs with genomic DNAs from *B. variegatus* and *M. tridactyla* using the same set of primers and conditions applied to amplify SATCHO1 and SATCHO2 in *C. hoffmanni*. The genomic DNA of *C. hoffmanni* was used as a positive control. PCR products from *B. variegatus* and *M. tridactyla* were cloned, sequenced (accession numbers: MT505305–MT505308), and used as probes for FISH under the same conditions described above.

Data availability

The datasets generated during and/or analyzed in the current study are available in the GenBank repository (https://www.ncbi.nlm.nih.gov/genbank/) under accession numbers: MT505303–MT505310.

Received: 23 July 2020; Accepted: 19 October 2020
Published online: 05 November 2020

References

1. Garrido-Ramos, M. A. Satellite DNA: An evolving topic. *Genes (Basel).* **8**, 230 (2017).
2. Lower, S. S., McGurk, M. P., Clark, A. G. & Barbash, D. A. Satellite DNA evolution: Old ideas, new approaches. *Curr. Opin. Genet. Dev.* **49**, 70–78 (2018).
3. Kuhn, G. C. S. Satellite DNA transcripts have diverse biological roles in *Drosophila*. *Heredity* **115**, 1–2. https://doi.org/10.1038/hdy.2015.12 (2015).
4. Talbert, P. B. & Henikoff, S. Transcribing centromeres: Noncoding RNAs and kinetochore assembly. *Trends Genet.* **34**, 587–599 (2018).
5. Escudeiro, A. *et al.* Bovine satellite DNAs—A history of the evolution of complexity and its impact in the Bovidae family. *Eur. Zool. J.* **86**, 20–37 (2019).
6. Masumoto, H., Nakano, M. & Ohzeki, J. I. The role of CENP-B and α-satellite DNA: De novo assembly and epigenetic maintenance of human centromeres. *Chromosome Res.* **12**, 543–556 (2004).
7. Kasinathan, S. & Henikoff, S. Non-B-form DNA is enriched at centromeres. *Mol. Biol. Evol.* **35**, 949–962 (2018).
8. Gamba, R. & Fachinetti, D. From evolution to function: Two sides of the same CENP-B coin?. *Exp. Cell Res.* https://doi.org/10.1016/j.yexcr.2020.111959 (2020).
9. Martínez-Balbás, A. *et al.* Satellite DNAs contain sequences that induce curvature. *Biochemistry* **29**, 2342–2348 (1990).
10. Fitzgerald, D. J., Drydens, G. L., Bronsonl, E. C. & Anderson, J. N. Conserved patterns of bending in satellite and nucleosome positioning DNA. *J. Biol. Chem.* **269**, 21303–21314 (1994).
11. Plohl, M., Meštrović, N. & Mravinac, B. Satellite DNA evolution. *Genome Dyn.* **7**, 126–152 (2012).
12. Ruiz-Ruano, F. J., López-León, M. D., Cabrero, J. & Camacho, J. P. M. High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Sci. Rep.* **6**, 1–14 (2016).
13. Dover, G. Molecular drive: A cohesive mode of species evolution. *Nature* **299**, 111–117 (1982).
14. Novák, P., Neumann, P., Pech, J., Steinhaisl, J. & MacAs, J. RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics* **29**, 792–793 (2013).

15. Delsuc, F., Vizcaíno, S. F. & Douzery, E. J. P. Influence of tertiary paleoenvironmental changes on the diversification of South American mammals: A relaxed molecular clock study within xenarthrans. *BMC Evol. Biol.* **4**, 1–13 (2004).
16. Superina, M. & Loughry, W. J. Why do xenarthrans matter?. *J. Mammal.* **96**, 617–621 (2015).
17. Gardner, A. L. *Mammals of South America, Volume 1- Marsupials, Xenarthrans, Shrews, and Bats.* (University of Chicago Press, Chicago, 2008).
18. Waters, P. D., Dobigny, G., Pardini, A. T. & Robinson, T. J. LINE-1 distribution in Afrotheria and Xenarthra: Implications for understanding the evolution of LINE-1 in eutherian genomes. *Chromosoma* **113**, 137–144 (2004).
19. Bao, W. & Jurka, J. Origin and evolution of LINE-1 derived "half-L1" retrotransposons (HAL1). *Gene* **465**, 9–16 (2010).
20. Nishihara, H., Kuno, S., Nikaido, M. & Okada, N. MyrSINEs: A novel SINE family in the anteater genomes. *Gene* **400**, 98–103 (2007).
21. Churakov, G., Smit, A. F. A., Brosius, J. & Schmitz, J. A novel abundant family of retroposed elements (DAS-SINEs) in the nine-banded armadillo (*Dasypus novemcinctus*). *Mol. Biol. Evol.* **22**, 886–893 (2005).
22. Alkan, C. *et al.* Genome-wide characterization of centromeric satellites from multiple mammalian genomes. *Genome Res.* **21**, 137–145 (2011).
23. Engelmann, G. F. The phylogeny of the Xenarthra. In *The Evolution and Ecology of Armadillos, Sloths and Vermilinguas* (ed. Montgomery, G.G.) 51–64 (Smithsonian Institution, Washington, DC, 1985).
24. Hayssen, V. *Choloepus hoffmanni* (Pilosa: Megalonychidae). *Mamm. Species* **43**, 37–55 (2011).
25. Wetzel, R. & Avila-Pires, F. Identification and distribution of the recent sloths of Brazil (Edentata). *Rev. Bras. Biol.* **40**, 831–836 (1980).
26. Steiner, C. C., Houck, M. L. & Ryder, O. A. Species identification and chromosome variation of captive two-toed sloths. *Zoo Biol.* **30**, 623–635 (2011).
27. Ruiz-García, M., Chacón, D., Plese, T., Schuler, I. & Shostell, J. M. Mitogenomics phylogenetic relationships of the current sloth's genera and species (Bradypodidae and Megalonychidae). *Mitochondrial DNA Part A DNA Mapp. Seq. Anal.* **29**, 281–299 (2018).
28. Corin-Frederick, J. Les formules gonosomiques dites aberrantes chez les mammifères euthériens. *Chromosoma* **27**, 268–287 (1969).
29. Jorge, W., Meritt, D. J. & Bernirschke, K. Chromosome studies in Edentata. *Cytobios* **18**, 157–172 (1978).
30. Jorge, W. *Estudo cromossômico de algumas espécies da ordem Edentata* (UNESP, São Paulo, 1981).
31. Jorge, W., Orsi-Souza, A. T. & Best, R. The somatic chromosomes of Xenarthra. In *The Evolution and Ecology of Armadillos, Sloths, and Vermilinguas* (ed. Montgomery, G. G.) 121–129 (Smithsonian Institution, Washington, DC, 1985).
32. Sonta, S. I., Hayata, I., Sasaki, M. & Kondo, N. Karyotype and sex determining in the two-toed sloth, *Choloepus didactylus*. *Chromosome Inf. Serv.* **28**, 15–17 (1980).
33. Dobigny, G. *et al.* Low rate of genomic repatterning in Xenarthra inferred from chromosome painting data. *Chromosome Res.* **13**, 651–663 (2005).
34. Svartman, M., Stone, G. & Stanyon, R. The ancestral Eutherian karyotype is present in Xenarthra. *PLoS Genet.* **2**, 1006–1011 (2006).
35. Earnshaw, W. C. & Rothfield, N. Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma. *Chromosoma* **91**, 313–321 (1985).
36. Muro, Y. *et al.* Centromere protein B assembles human centromeric α-satellite DNA at the 17-bp sequence, CENP-B box. *J. Cell Biol.* **116**, 585–596 (1992).
37. Stitou, S., Díaz De La Guardia, R., Jiménez, R. & Burgos, M. Isolation of a species-specific satellite DNA with a novel CENP-B-like box from the North African rodent *Lemniscomys barbarus*. *Exp. Cell Res.* **250**, 381–386 (1999).
38. Flynn, J. M., Long, M., Wing, R. A. & Clark, A. G. Evolutionary dynamics of abundant 7 bp satellites in the genome of *Drosophila virilis*. *Mol. Biol. Evol.* **37**, 1362–1375 (2020).
39. Gibb, G. C. *et al.* Shotgun mitogenomics provides a reference phylogenetic framework and timescale for living xenarthrans. *Mol. Biol. Evol.* **33**, 621–642 (2016).
40. Melters, D. P. *et al.* Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biol.* **14**, 1–20 (2013).
41. Smalec, B. M., Heider, T. N., Flynn, B. L. & O'Neill, R. J. A centromere satellite concomitant with extensive karyotypic diversity across the *Peromyscus* genus defies predictions of molecular drive. *Chromosome Res.* **27**, 237–252 (2019).
42. Schmidt, T. & Heslop-Harrison, J. S. Genomes, genes and junk: The large-scale organization of plant chromosomes. *Trends Plant Sci.* **3**, 195–199 (1998).
43. Henikoff, S., Ahmad, K. & Malik, H. S. The centromere paradox: Stable inheritance with rapidly evolving DNA. *Science* **293**, 1098–1102 (2001).
44. Árnason, Ú., Grétarsdóttir, S. & Widegren, B. Mysticete (baleen whale) relationships based upon the sequence of the common cetacean DNA satellite. *Mol. Biol. Evol.* **9**, 1018–1028 (1992).
45. Fanning, T. G., Seuánez, H. N. & Forman, L. Satellite DNA sequences in the New World primate *Cebus apella* (Platyrrhini, Primates). *Chromosoma* **102**, 306–311 (1993).
46. Hřibová, E. *et al.* Repetitive part of the banana (*Musa acuminata*) genome investigated by low-depth 454 sequencing. *BMC Plant Biol.* **10**, 1–10 (2010).
47. Kopecna, O. *et al.* Tribe-specific satellite DNA in non-domestic Bovidae. *Chromosome Res.* **22**, 277–291 (2014).
48. Ávila Robledillo, L. *et al.* Satellite DNA in *Vicia faba* is characterized by remarkable diversity in its sequence composition, association with centromeres, and replication timing. *Sci. Rep.* **8**, 1–11 (2018).
49. Skowronski, J., Plucienniczak, A. & Bednarek, A. Bovine l.709 satellite recombination hotspots and dispersed repeated sequences. *J. Mol. Biol.* **177**, 399–416 (1984).
50. Lorite, P., Carrillo, J. A., Aguilar, J. A. & Palomeque, T. Isolation and characterization of two families of satellite DNA with repetitive units of 135 bp and 2.5 kb in the ant *Monomorium subopacum* (Hymenoptera, Formicidae). *Cytogenet. Genome Res.* **105**, 83–92 (2004).
51. Taparowsky, E. J. & Gerbi, S. A. Structure of 1.711 lb gm/cm^3 bovine satellite DNA: Evolutionary relationship to satellite I. *Nucleic Acids Res.* **10**, 5503–5515 (1982).
52. Jobse, C. *et al.* Evolution and recombination of bovine DNA repeats. *J. Mol. Evol.* **41**, 277–283 (1995).
53. Modi, W. S., Gallagher, D. S. & Womack, J. E. Evolutionary histories of highly repeated DNA families among the Artiodactyla (Mammalia). *J. Mol. Evol.* **42**, 337–349 (1996).
54. Adega, F., Guedes-Pinto, H. & Chaves, R. Satellite DNA in the karyotype evolution of domestic animals—Clinical considerations. *Cytogenet. Genome Res.* **126**, 12–20 (2009).
55. Escudeiro, A., Adega, F., Robinson, T. J., Heslop-Harrison, J. S. & Chaves, R. Conservation, divergence, and functions of centromeric satellite DNA families in the bovidae. *Genome Biol. Evol.* **11**, 1152–1165 (2019).
56. Jurka, J. *et al.* Repbase update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.* **110**, 462–467 (2005).
57. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
58. Rozas, J. *et al.* DnaSP v6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* **34**, 3299–3302 (2017).
59. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
60. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular evolutionary genetics analysis Version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).

61. Rice, P., Longden, L. & Bleasby, A. EMBOSS: The European molecular biology open software suite. *Trends Genet.* **16**, 276–277 (2000).
62. Gruber, A. R., Lorenz, R., Bernhart, S. H., Neuböck, R. & Hofacker, I. L. The Vienna RNA websuite. *Nucleic Acids Res.* **36**, 70–74 (2008).
63. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).
64. RStudio Team. RStudio: Integrated Development for R. [Computer sotware]. RStudio, Inc. https://www.rstudio.com/ (2020).
65. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259 (2019).
66. Stanyon, R. & Galleni, L. A rapid fibroblast culture technique for high resolution karyotypes. *Bolletino di Zool.* **58**, 81–83 (1991).
67. Seabright, M. A rapid banding technique for human chromosomes. *Lancet* **2**, 971–972 (1971).
68. Sumner, A. T. A simple technique for demonstrating centromeric heterochromatin. *Exp. Cell Res.* **75**, 304–306 (1972).

Acknowledgements

R.S.S., P.H. and M.P.V. received doctoral fellowships from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). G.C.S.K. and M.S. are recipients of productivity fellowships from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq—processes 308386/2018-3 and 310433/2018-5, respectively).

Author contributions

R.S.S., P.H. and M.P.V. conducted the bioinformatics and data analyses, designed the study and drafted the manuscript; R.S.S. and M.P.V. carried out the cytogenetic and molecular analyses; R.S.S. prepared all figures; V.S.P. provide the tissue and blood samples; G.C.S.K. and M.S. conceived and coordinated the study, and helped drafting the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-76199-8.

Correspondence and requests for materials should be addressed to M.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020

Supplementary Information for:

Identification and characterization of satellite DNAs in two-toed sloths of the genus *Choloepus* (Megalonychidae, Xenarthra)

Radarane Santos Sena[1]; Pedro Heringer[1]; Mirela Pelizaro Valeri[1]; Valéria do SocorroPereira[2]; Gustavo C.S. Kuhn[1]; Marta Svartman[1]*

[1]Laboratório de Citogenômica Evolutiva, Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais,Belo Horizonte, MG, Brazil

[2]Fundação de Parques Municipais e Zoobotânica, Belo Horizonte, MG, Brazil

Running Title: Satellite DNAs of two-toed sloths

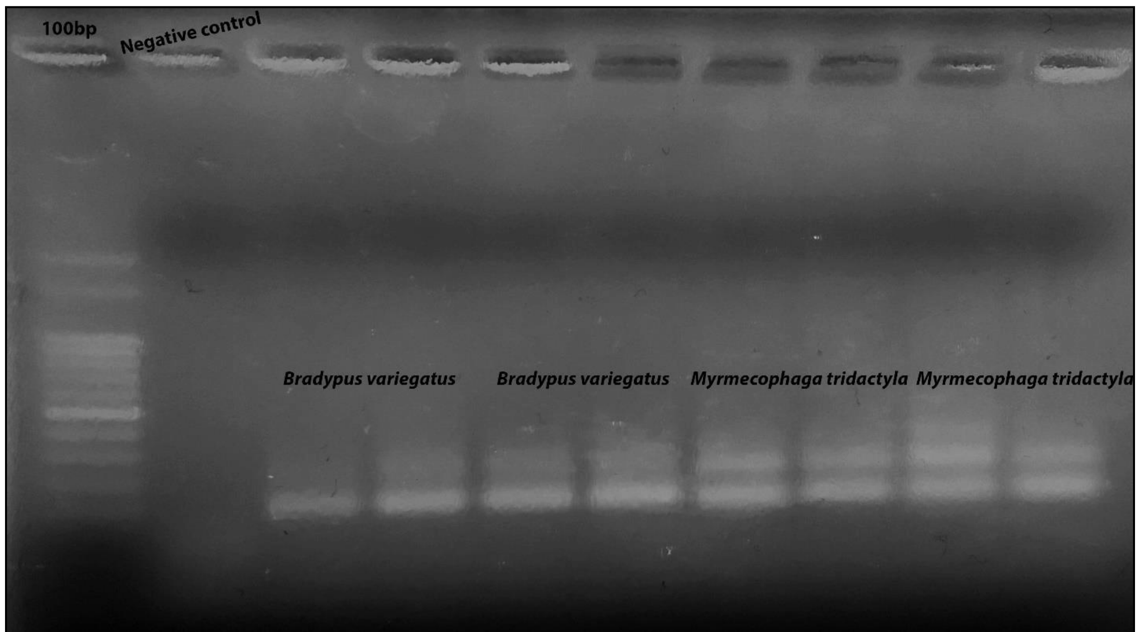
*Corresponding author
E-mail: svartmanm@ufmg.br
Telephone number: 5531-34092612
Fax number: 5531-34092568

Supplementary Data1:

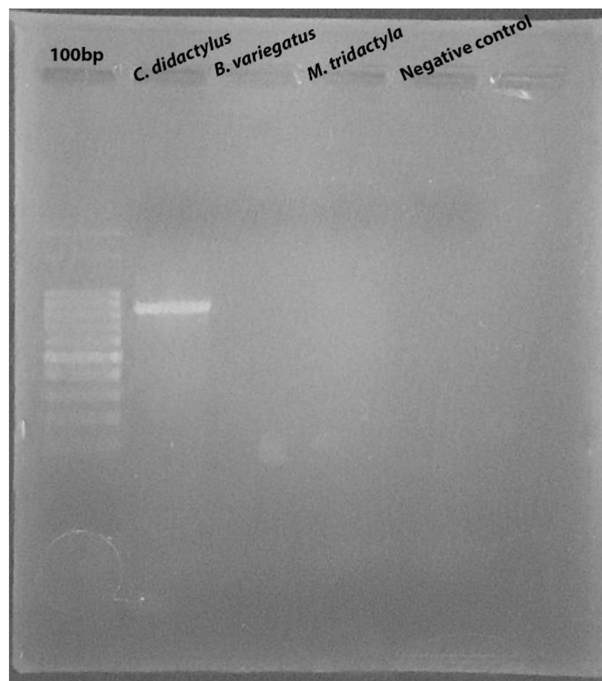
>SATCHO1
GAAAGCAGTATGTTACACAGTTCTGAAAACTAAGTTCAATACAGCTTTCATTGACTTCCATGTGTCTTCCCTGGGCTGAAAAAC
AACTTTTTTGAAAGCAGTCACGTGGGACTCTGC

Supplementary Data2:

>SATCHO2
TGGCCTGTGCAGTCAGGAGTAGGCATATGAGGATGATTCTGCAGCCCTGAGGATGACATGCCCCCGGGTGGTTGGTGAGATTTT
TGACTCTCCCGTTAGTGCGCTGCACTGGTATGCCCGAGGGCAACTTCCAGTGGGTTGTGGGCTAGGCCGAGGCTTAGTGAAGTG
CCCCTGCCCTCGCTTGGGCCTGCCATTCTTTTCCAATGGCACTGATATTACCAGTGGATCGTAGACATGCAGCCTAGCACCCCG
AATCTCACTTAGGAGAGCTCACCGTGAGGCATGCCCATGTTCAGTCAATGAACACTCTAGAAGTCAGGGAAATTGGGACAACTG
TGACGTGTCTGTCCTCTCCACTGTTTATACCCGGCATGATGATTGGTTCATGGCTCCCTTAATCGTGGGATTTCTTGCTCTATT
GGTGCTTGCCGGATCTCTCCTTCTGGGACTAAAGGCAGAGTGATAACCAAGGTGGTCCCGTGATGCAGAGACCACATGACATGT
GGACAGGCTGATGTTTTCTGGGTTGAGATTGCTGCAGGCACGGTTGTGGACAGGGACGGAGACTCCCAGAGGGCCGGCACGAAG
GTGGACACTGAATATTTTTCTTGCTCTTTCCAGTCATTTCCATTGCCATTAGGGAATTCCCTGTTCTGGGAGCCAACATACTCT
GGCTCTGTCATTGAGTGTGATCCACAAGGAACTGTAAAGGATCACCCCGAGTCCCCCTCTGACCTGCTGCTGGGCATGGCAGAG
ATTTACCACAGGCGTGGCCCCTGAACTGTTCCACATCTGAAAGAAGATTTCAGCAAAGGGAGAGTTCTTCCCCGAAGAATTTAG
TTCCTGAATTTGCTGATGCTCAAGGGAGTATGCCCAAGTGGAACTGACTGTGACAATTCTTCTCCTATCCCATGAAAGCAGGTC
AGTGTAATCTCACCCGGATCTGAACCTAGGCCTTCATAAGGATGGAAGCCAACACTTAGGAAGTGTGTGTTTATATAATTCATT
TCTTCACCGTGTTAACATTGTTTCACATTCCTTGTATTAAAAGGTTGCCTTTAAGAACTGGACATTTGTCAGTAAAAGAATGTC
AATTTCTTTAGTTATTCTTGTTGTATTTCCCTTCACAACAGAGACGTTCTCTTAGGTATTCACTGGAACATTACCACAATCTTT
AAATTTACCTGGTCGTAGTACAAATATGTCGCCTAGGGACCTGGTCCACACCTCAACAATTAGCCACTCATGACCTATAGACCA
AGCAAAATGTTAATGGCAGTAATAATGTGAGAAATATAGAACGGAAAAAACAAGACAGAGTGTTAGGTCTCACGTGAAACATTC
CGTTTTCTTCAGGTGTATTTATTCTCTTTCCCGTGAATATACATCAATTATTTCCTTCACCTGTATAAATTAGAAAACTGCTGA
ACCATCTTTAAAAACTATTCCCCAAATTCTGTTTTATTCAGAGATTCTTCCTGCTGATATTGACTTTATGAGCTTGAGGTGAGA
TAGGCCGATAATTATTTTGCCCGCACTGCGGTTATTTTTTAAATGAAAAGTTTAAATTTTGTTTCTTTCAGTCTATTGAATTCA
TTATACATATTCTCTGTAAGCTACAAAAACACACAAACATCAAAAAGTCAAAACTTTTCATTAAAGCAAAGCAGAGTTTAAGAA
AAGCAAATAACCTATAGTAACTACTTTGCTTCCTCATTGCTCAAGTTTGCATTCATGTGTTATCTTCAAGGCATTTGTTTAAAG
ATTTCTAATATTATAAACAATACTCGTCCTACAGTCCATTTAGTTCCTTTTGGGTACTTTTTCTTGCTGCACAGACCTGTTTCC
CTTTTCCGTCTCTTGTCATCTAGTTACAGTAACACTGCTTCAAACCCCCGTATCCAGTTTCAGATAGAGGCTTTTCTGCTTTGC
CTAGTGTGCAGACTCAGCTAAAGAGCAGGGAATGTGTTACTGAATTTTTCCCCTTACCCCACGGAGAAGTCGAAGTGCCCACGT
GTGCCCAATAGGCGGCAGGAAATCCCAAGTTAACAAAAGTCAGTGAAAAATTCCCACTGGGGCTCGAACCGGTGGAGGAGACAG
CGTCAGCACAGGCTCCAAAGCGGTTTAGAATTCTAGCAGGGGCTCGGAATGCGCATGCGCAAACGACGCTGTACCTCATCTGTG
TCCCAAAGGAATCTGGTTCCTAAGCAACAGACCTCTAGCCCAACGGCTGCCCTGTGGTCTGAACATGTAAGGGGCTCCCCAAGT
GTGCAGGTCACGGTGGAGGTGTTG



Supplementary Figure 1. Uncropped agarose gel of PCR products from SATCHO1 amplified on genomic DNAs of *Bradypus variegatus* and *Myrmecophaga tridactyla*.



Supplementary Figure 2. Uncropped agarose gel of PCR products from SATCHO2 amplified on genomic DNAs of *Choloepus hoffmanni*, *Bradypus variegatus* and *Myrmecophaga tridactyla*.

4. CAPÍTULO 2

Comparative Analyses of Satellite DNAs in Six Xenarthra Species

Radarane Santo Sena[1]; Gustavo Campos Silva Kuhn[1]; Marta Svartman[1]*
[1]Laboratório de Citogenômica Evolutiva, Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

Running Title: Satellite DNAs of Six Xenarthra Species

*Corresponding author
E-mail: svartmanm@ufmg.br
Telephone number: 5531-34092612
Fax number: 5531-34092568

Abstract

Using *in silico* analysis, we identified the repetitive DNA content in the genomes of six Xenarthra species: *Bradypus variegatus*, *Myrmecophaga tridactyla*, and *Tamandua tetradactyla*, from the order Pilosa, and *Chaetophractus vellerosus, Dasypus novemcinctus,* and *Tolypeutes matacus*, from the order Cingulata. LINEs were the most abundant class of repetitive DNA. Nevertheless, our focus was to identify and characterize the most abundant satellite DNA sequences (satDNAs), mainly the satDNAs from the SATCHO1 and SATCHO2 families. These two families of centromeric satDNA from the genus *Choloepus* were recently described in *C. didactylus* and *C. hoffmanni.* We found 39 different satDNAs in the six species analyzed and identified ten candidates for the SATCHO1 and one for the SATCHO2 subfamily. We did not identify any satDNA that shares homology with SATCHO1 in the *Tamandua tetradactyla* genome. Performing a phylogenetic analysis with copies of satDNAs with ~116 bp and ~117 bp retrieved from the five species and the two *Choloepus* species, we found that only *B. variegatus* sequences have a complete species-specific pattern of homogenization. All satDNAs from the SATCHO1 family have putative centromeric features such as small dyad symmetries, putative CENP-B box, secondary structure, and five satDNAs have a higher-order (HOR) organization with a subunit of ~117 bp. We also identified 22 uncommonly large satDNAs that may be a feature of Xenarthra genomes. Our results show that SATCHO1 is an important satDNA family present in several Xenarthra species, which may participate in centromeric function. Our results shed some light in the understanding of Xenarthra genome size and evolution.

Keywords: Satellite DNAs; Centromere; RepeatExplorer2; Megasatellite; Higher-Order Structure (HOR)

Introduction

Repetitive DNA elements make up a substantial fraction of eukaryotic genomes, may occupy >50% of genome content, and are known for increasing the genome size of some eukaryotic species, such as the human and maize genomes (reviewed in Garrido-Ramos, 2017). According to their organization, they are divided into two main categories: dispersed and tandemly organized repetitions. Satellite DNAs (satDNAs) are highly tandemly organized sequences that can build very long arrays, up to megabases in length, located at the constitutive heterochromatin of essential chromosome regions, like centromeres and telomeres (reviewed in Garrido-Ramos, 2017; Lower et al., 2019; Thakur et al., 2021).

The satDNA sequences can also be found in a complex organization called higher-order structure (HOR), constituted by monomers comprised of multiple simple repetitions of two or more adjacent units (M. Garrido-Ramos, 2017; Thakur et al., 2021). SatDNAs with HOR organization have been studied in different mammals. Some primates including humans have the centromeric alpha satDNA, the most studied HOR satellite in mammals (Sullivan et al., 2017). In addition to higher-order structures, satDNAs as the centromeric alpha may also form secondary structures, short-inverted repeats (short dyad symmetries), and long non-coding products, which have been related with functional centromeres (Kasinathan & Henikoff, 2018). In mammals, centromeric satDNA sequences also show a 17 bp motif binding-domain for the centromeric protein CENP-B. The CENP-B protein promotes the correct assembly of the kinetochore structure during cell divisions in some species. The CENP-B motif is a conserved sequence in mammalian genomes, called the most evolutionary conserved domain (ECD - nTTCGnnnnAnnCGGGn) (Muro et al., 1992; Stitou et al., 1999; Alkan et al., 2011).

SatDNAs are among the fastest evolving components of the eukaryotic genomes and, in some cases, can be used as taxonomic markers (Liao 1999; Kuhn et al. 2012). The primary mode of satDNA variation is through Concerted Evolution, which results in high intra-specific DNA identity between repeats but high interspecific divergence. The development of sequencing and assembling technologies and new bioinformatic tools, such as RepeatExplorer2 and TAREAN, allowed advances in *de novo* identification of repetitive elements, including satDNAs (Novák et al., 2020). TAREAN is a computational tool located in the RepeatExplorer2 pipeline and developed to detect satDNA sequences present in

unassembled raw short reads. TAREAN searches for clusters of reads with graphs in a circular shape, which characterize satDNAs.

Xenarthra is a monophyletic clade placed on the base of the mammalian phylogenetic tree. It is endemic from South and Central America and comprises thirty-nine extant species divided into two orders: Pilosa with ten anteater (Vermilingua) and seven sloth (Folivora) species; and Cingulata with twenty-two species of armadillos (https://www.xenarthrans.org/). Xenarthrans represent a notorious contribution to paleontological records by providing a massive catalog of extinct species, which comprises some of the largest terrestrial mammals that inhabited the planet, like the giant sloths and glyptodons (Superina & Loughry, 2015). They occupy different habitats, including fossorial and arboreal species, and show variable body features, like a rigid carapace or short and long coats (Delsuc et al., 2001, 2004; Superina & Loughry, 2015). Another relevant feature about Xenarthra is their genome size, which, together with those of Afrotheria, is considered one of the largest among mammals, with ~4.5pg and ~5.5pg, respectively (Redi & Capanna, 2012).

Xenarthrans are still poorly studied, including their repetitive DNA content (Superina & Loughry, 2015). Only three species had their most abundant satDNAs analyzed. The nine-banded armadillo *Dasypus novemcinctus* had its ~173 bp centromeric satDNA identified and characterized by Alkan et al. (2011). The two extant species of sloths of the genus *Choloepus*, *C. didactylus* and *C. hoffmanni*, had their most abundant satDNAs recently identified, described, and mapped to centromeric regions: SATCHO1 with ~117 bp and SATCHO2 with ~2292 bp (Sena et al., 2020). Other repetitive elements, like the transposable elements LINE (Long Interspersed Nuclear Element) and SINE (Small Interspersed Nuclear Element), were identified and studied in six Xenarthra species: the armadillos *D. novemcinctus* and *Euphractus sexcinctus* (Waters et al., 2004; Churakov, 2005), the anteaters *Myrmecophaga tridactyla* and *Tamandua tetradactyla* (Waters et al., 2004; Nishihara et al., 2007), and the sloths *Bradypus variegatus* and *C. hoffmanni* (Waters et al., 2004; Bao & Jurka, 2010). These two transposons are the most abundant repetitive elements in copy number and genome percentage in mammalian genomes. In average, LINEs occupy ~22.6% and SINEs ~10.5% of mammalian genomes (Osmanski et al., 2023). SINEs are non-autonomous elements and depend on the presence of functional LINEs, which is also associated with the origin or amplification of some satDNAs (Plohl et al., 2012).

From a total of 39 recognized extant Xenarthra species, only eight had their genomes sequenced and made available on databases (Genbank access in 10/03/2023). In this work we used *in silico* methods to analyze the repetitive DNA fraction of the sequenced genomes of six species, in which we identified for the first time the most abundant satDNAs: the three-toed sloth *B. variegatus*, the anteaters *M. tridactyla* and *T. tetradactyla*, and the armadillos *Chaetophractus vellerosus*, and *T. matacus*. In addition, we reanalyzed the *D. novemcinctus* genome to identify its repetitive DNA content.

Material and Methods

Identification and Characterization of SatDNAs

We ran the RepeatExplorer2 pipeline and its tool TAREAN to identify the repetitive content and the most abundant satDNA sequences in six Xenarthra genomes available on Genbank (Benson et al., 2012) (https://www.ncbi.nlm.nih.gov/genbank/). We analyzed raw paired-end Illumina reads from *Bradypus variegatus* (SRX10458626), *Myrmecophaga tridactyla* (SRX8010125), *Tamandua tetradactyla* (SRX4562107), *Chaetophractus vellerosus* (SRX4562080), *Dasypus novemcinctus* (SRX9176964) and *Tolypeutes matacus* (SRX4562109). We performed the analyses using default options and did not specify a sample size for the runs. The software randomly chose and analyzed 2,723.002 reads of *B. variegatus* (~7% of the genome proportion (GP)), 5,693.729 reads of *M. tridactyla* (~13% of GP), 4,938.038 reads of *T. tetradactyla* (~12% of GP), 3,716.672 reads of *C. vellerosus* (~8.5% of GP), 5,945.428 reads of *D. novemcinctus* (~10.7% of GP), and 4,968.572 reads of *T. matacus* (~10% of the GP).

After that, we used the monomer consensuses of the putative satDNAs identified by TAREAN from each analyzed genome as queries in the Repbase CENSOR (Kohany et al., 2006) to check if these sequences had been already annotated as repetitive DNAs. Using default options, we also searched the level of nucleotide sequence identity of the identified satDNAs in BLASTn searches in the Genbank nucleotide collection (accessed on 10/01/2023).

Searching SatDNAs from the SATCHO1 Family We focused on the sequences that could be closely related to SATCHO1, a centromeric satDNA with ~117 bp long repeat units previously identified in the *Choloepus didactylus* and *C. hoffmanni* genomes (Sena et al., 2020).

Three of the six species analyzed by TAREAN (*B. variegatus*, *C. vellerosus*, and *T. matacus*) had satDNAs with 116 bp or/and 117 bp in length. We then used the alignment tool present in the software Geneious to verify the nucleotide identity among the satDNAs with 117 bp and SATCHO1. We also searched for other satDNAs that could be related to SATCHO1 among the satDNAs with more than 117 bp identified by TAREAN.

To retrieve copies of the 117 bp satDNAs of *B. variegatus*, *C. vellerosus*, and *T. matacus*, we used the consensuses of each 117 bp satDNAs as queries in BLASTn searches against whole genome shotgun contigs from these three species genomes. Moreover, to get contigs with more tandemly repeated sequences in these BLASTn searches, we set artificial queries containing at least 50 tandem repeats. We also used the "Program Selection" optimized for 'somewhat similar sequences' with max number of target sequences set in 1000 contigs.

We then used the 'Dotplot' tool of the Geneious software to manually isolate 116 bp and 117 bp satDNA copies from the BLASTn results (hits). For each species, we retrieved at least a hundred copies randomly chosen from 1,000 hits. After that, we constructed a phylogeny using the Maximum Likelihood method with 1,000 replicates on MEGA X (Kumar et al., 2018). For this analysis we included a hundred copies of SATCHO1 from each *C. didactylus* and *C. hoffmanni*, which were also randomly chosen. We edited the phylogenetic tree using the iTOLLv.6 tool (Letunic & Bork, 2021).

Searching for Centromeric Hallmarks

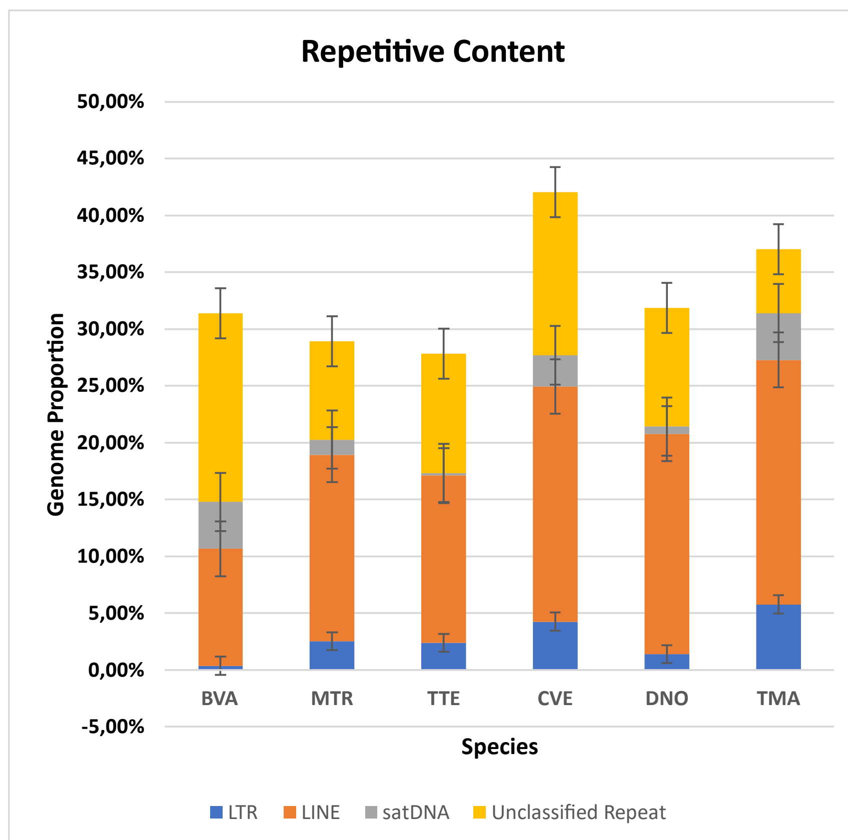
We investigated the presence of putative CENP-B box-like motifs, the 17 bp region conserved among mammals, in the SATCHO1 family candidate sequences. For this analysis, we used the alignments algorithm MUSCLE (Edgar, 2004) on MEGA X. To identify the presence of optimal secondary structures with minimum free energy (MFE) in those sequences we ran RNAfold web server (https://rna.tbi.univie.ac.at/). We searched for small dyad symmetries in the satDNAs of interest by running the EMBOSS palindrome web explorer (https://www.bioinformatics.nl/cgi-bin/emboss/palindrome) with a minimum palindrome length of 4 bp, maximum length of 100 bp, and maximum gap between elements of 20 bp.

Results

RepeatExplorer2 and TAREAN Results

RepeatExplorer 2 and TAREAN pipelines results showed that the repetitive DNA content corresponds to at least ~25% of the total genome of the six species analyzed. Most repetitive sequences were LINEs (Figure 1), corresponding to more than 10% of each genome proportion. The species with the highest repetitive DNA content was the armadillo *C. vellerosus*, in which it corresponded to more than 45% of the genome. In contrast, *T. tetradactyla* had the lowest repetitive DNA content, corresponding to ~25% of the genome.

Figure 1: Repetitive content for each species analyzed.



Species analyzed: *Bradypus variegatus* (BVA), *Myrmecophaga tridactyla* (MTR), *Tamandua tetradactyla* (TTE), *Cheatophractus vellerosus* (CVE), *Dasypus novemcinctus* (DNO), and *Tolypeutes matacus* (TMA).

The pipelines also identified 39 putative satDNAs among the repetitive DNA content of the six species. Seven putative satDNAs in the *B. variegatus* genome, four with high and three with low confidence (HC, high confidence; LC, low confidence), corresponding to approximately ~4.1% of the *B. variegatus* genome. Eight satDNAs in *M. tridactyla*, six with

HC, and two with LC. These eight putative satDNAs correspond to approximately ~2.6% of the *M. tridactyla* genome. TAREAN found two satDNAs in *T. tetradactyla* with LC, matching with ~0.35% of the *T. tetradactyla* genome. In *C. vellerosus*, TAREAN identified eleven satDNAs, nine with HC and two with LC, corresponding to approximately ~2.7% of the *C. vellerosus* genome. TAREAN identified, with HC, the same satDNA (DNO173) detected by Alkan et al. (2011) in the *D. novemcinctus* genome and representing ~0.63% of the genome content. Finally, in *T. matacus*, ten satDNAs were identified, three with HC and seven with LC. These ten sequences comprise approximately ~4.1 % of the *T. matacus* genome.

Supplementary File 1 summarizes the main TAREAN results on satDNAs from the six Xenarthra species analyzed. TAREAN considers the indexes C and P to identify the putative satDNAs with HC or LC. In general, an AT content above 50% agrees with true satDNAs, and most satDNAs identified herein by TAREAN are AT-rich. Supplementary File 1 presents the consensus sequences of satDNAs identified by TAREAN.

Results from sequence similarity searches on Repbase Censor and in the GenBank nucleotide collection showed that some satDNAs that we identified are absent on the Repbase collection, significant similarity was not found for them in the nucleotide collection of GenBank, or these satDNAs absent on Repbase showed some hits of similarities with other Xenarthra species (Supplementary File 1). Others are composed of fragments of repetitive elements, especially of different LINEs (Supplementary File 1).

Identification and Characterization of SatDNAs from the SATCHO1 and SATCHO2 Families

We screened our collection of satDNAs identified by TAREAN for those that share sequence homology with SATCHO1, using pairwise sequence comparisons in the Geneious software. According to the nucleotide alignments, the following clusters share a pairwise percentage of identity ≥55% with SATCHO1: 1 and 48 in *B. variegatus*; 58 in *M. tridactyla*; 8, 9, 36, 58, and 77 in *C. vellerosus*; 1 and 41 in *T. matacus* (Figure 2). We named these satDNAs according to their monomer length and species abbreviation: *B. variegatus* (BVA116 and BVA117), *M. tridactyla* (MTR3029), *C. vellerosus* (CVE812, CVE116, CVE117, CVE811, and CVE815), and *T. matacus* (TMA451 and TMA117) (Figure 2; Table 1).

The satDNAs with 116 bp and 117 bp had the highest identity with SATCHO1 (Figure 2; Table 1). The satDNAs BVA116 and BVA117 showed ~84% of identity between their sequences. In *C. vellerosus*, the CVE116 and CVE117 showed ~68% of identity between their sequences (Figure 2; Table 1).

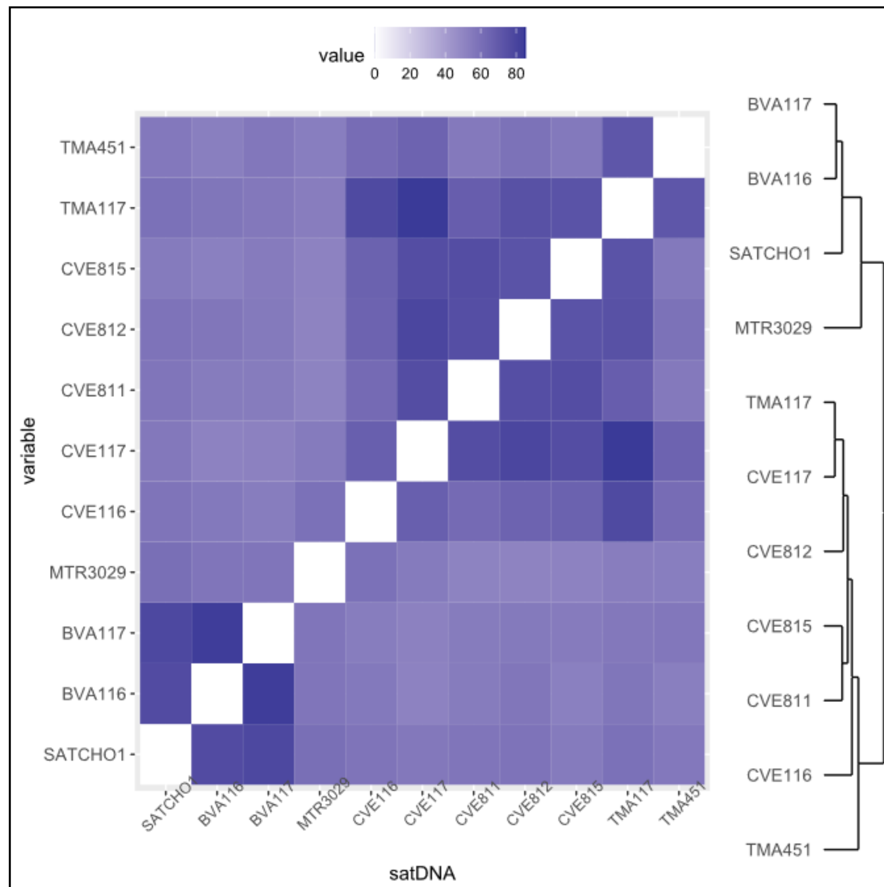


Figure 2: Heat map showing the satDNAs from the SATCHO1 family based on their pairwise identity and a cluster tree inferred from their pairwise distance.

The satDNAs BVA116 and BVA117 from the sloth *B. variegatus* share more identical sites with SATCHO1 (>76% of identity) (Table 1). The only satDNA in *M. tridactyla* with some identity with SATCHO1 was MTR3029 (~60.5% of identity) (Table 1). These results suggest that all these satDNAs (BVA116, BVA117, MTR3029, CVE116, CVE117, CVE811, CVE812, CVE815, TMA117, and TMA451) are members of the same SATCHO1 satDNA family (Figure 2; Table 1). Unexpectedly, we did not find satDNAs sharing homology with SATCHO1 in *T. tetradactyla*.

Table 1: Main features of the satDNAs from the SATCHO1 family.

Species	Monomer size bp	Internal Repetition	Genome Proportion %	AT Content	Repbase	Pairwise % identity with SATCHO1
***Bradypus variegatus* (BVA)**	116	no	1.700	61,21%	Run didn´t produce any hit	76.9
	117	no	0.130	58,97%	Run didn´t produce any hit	78.6
***Myrmecophaga tridactyla* (MTR)**	3029	yes	0.025	52,53%	Two small fragments of TEs	60.5
Chaetophractus vellerosus (CVE)	116	no	0.940	59,48%	Run didn´t produce any hit	57.8
	117	no	0.160	60,68%	Run didn´t produce any hit	56.3
	812	yes	1.000	59,73%	Run didn´t produce any hit	58.3
	811	yes	0.061	55,61%	Run didn´t produce any hit	57.5
	815	yes	0.024	57,55%	Run didn´t produce any hit	55.1
***Tolypeutes matacus* (TMA)**	117	no	1.400	65,81%	Run didn´t produce any hit	61.2
	451	yes	0.410	64,52%	Run didn´t produce any hit	55.1

We constructed a Maximum Likelihood tree with SATCHO1 consensus sequences from *C. hoffmanni*, *C. didactylus*, *B. variegatus*, *M. tridactyla*, *C. vellerosus*, and *T. matacus*. The resulting tree agrees with the phylogenetic relationship between species, as proposed by Gibb et al. (2016) that used mitogenomic data (Figure 3).

We then investigated the presence of internal repeats within the satDNA repeats with more than 117 bp. We found that some of these satDNAs have a higher-order repeats (HOR) organization, made by multiples of the basic repeat unit of ~117 bp. MTR3029 has nine ~117 bp units, but they are not in tandem, and five of these units are fragments of SATCHO1 (Figure 4). The satDNAs CVE811, CVE812, and CVE815 have their repetition units (monomers) constituted by six 117 bp sequences organized in tandem (Figure 4). The pairwise comparison among these three satDNAs showed that their sequences share ~73% of nucleotide identity (Supplementary File 2). Thus, although TAREAN identified them as three different satDNAs, with no supercluster similarity, they likely belong to the same satDNA family (SATCHO1) in *C. vellerosus* and could be regarded as intra-specific variants from the

same SATCHO1 family. After checking their proportion and organization pattern on the assembled genomes and TAREAN contigs, it seems that CVE811 is the most abundant variant. The nucleotide identity among their internal 117 bp units varied: ~61.2% in CVE811, ~60.2% in CVE812, and ~61.4% in CVE815, while the HOR units share more than ~83% of pairwise identity (Figure 4). The TMA451 sequence had its monomer organized into three 117 bp units plus an incomplete unit with 110 bp at the beginning of the sequence. These four units have ~58.4% nucleotide identity (Figure 4). Searching for this sequence in contigs of the *T. matacus* assembled genome, we did not find multiple copies from TMA451, and some of these copies are in a tandem array. However, they are truncated, smaller, or larger than TMA451.

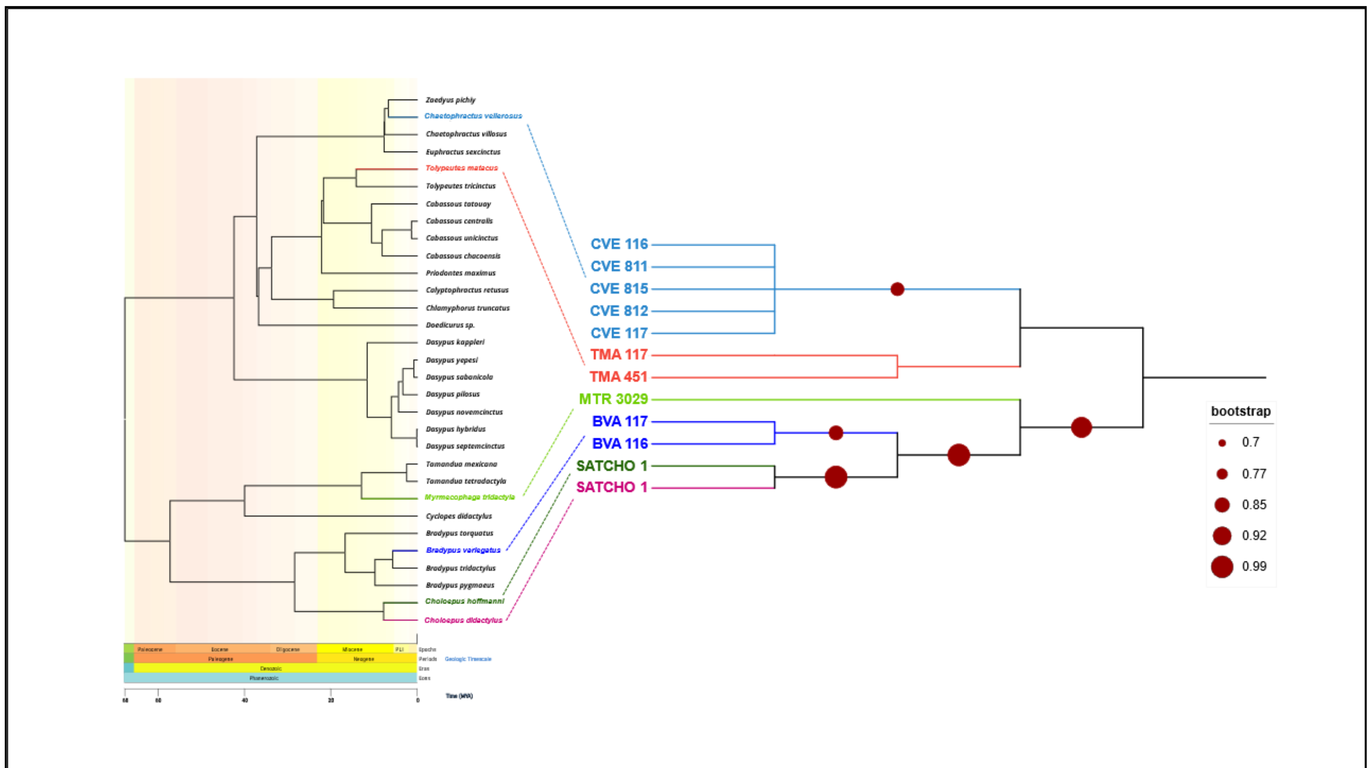


Figure 3: Comparison between the phylogenetic relationships of extant Xenarthra species (a) (Gibb *et al.* 2016) and a phylogenetic tree of SATCHO1 and its family (b). The phylogenetic tree of the SATCHO1 family was inferred by the Maximum Likelihood method with 1000 bootstraps and minimum bootstrap support is 70%.

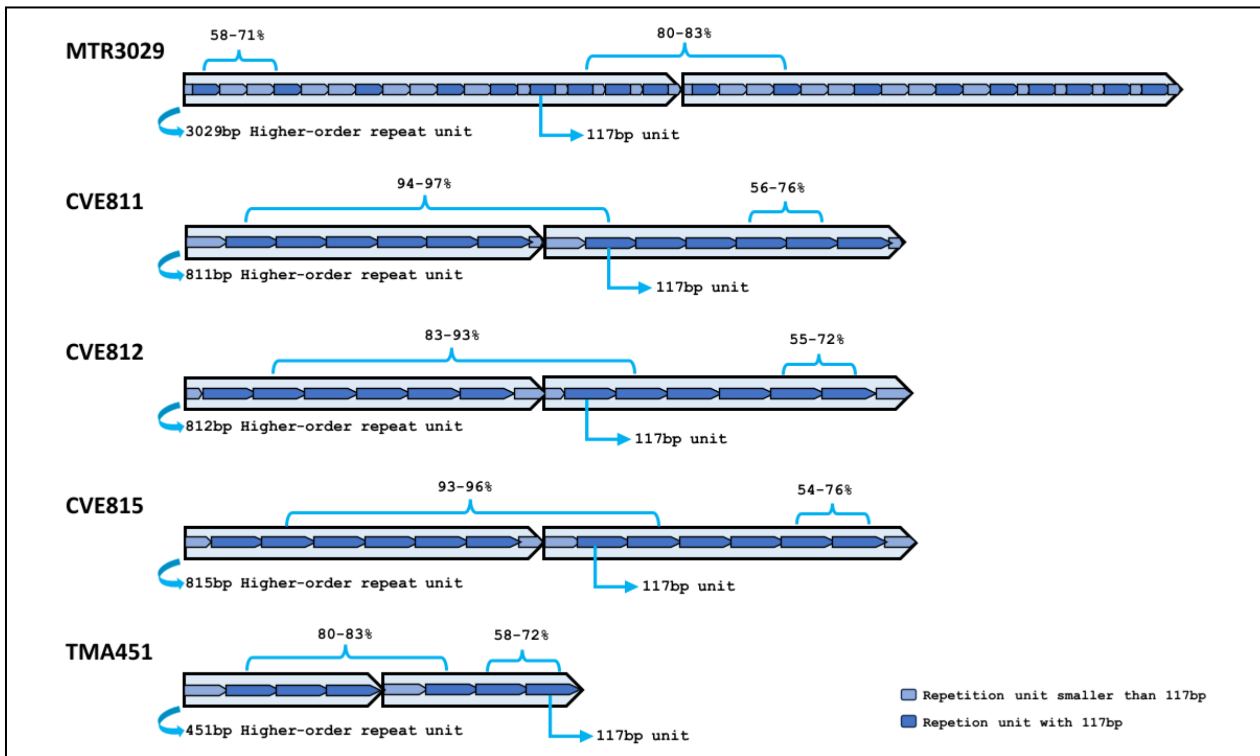


Figure 4: Schematic figure with the higher-order structures of satDNAs MTR3029, CVE811, CVE812, CVE815 and TMA451. The repetitive units (in dark blue) represent the 117 bp sequences which are repeated in tandem and together form the satDNAs monomer sequences.

After the identification of the putative satDNAs that belong to the SATCHO1 family, we constructed a maximum likelihood tree with some copies of the 116 bp, 117 bp satDNAs (SATCHO1, BVA116, BVA117, CVE117, CVE117, and TMA117) retrieved from the assembled genomes of *C. didactylus*, *C. hoffmanni*, *B. variegatus*, *C. vellerosus*, and *T. matacus*. The resulting tree (Figure 5) shows satDNA copies forming two groups: one comprising the sloths (*B. variegatus* and the two species of *Choloepus*) and the other with the two species of armadillos (*C. vellerosus* and *T. matacus*). Additionally, the sloths group also divided into two. The copies of *B. variegatus* split into a different branch from those of *Choloepus*. The sequences from the two *Choloepus* species are very similar to each other and do not separate into species-specific branches, as already described (Sena et al., 2020).

In *Bradypus variegatus*, there is a more complete homogenization that follows the classical Concerted Evolution pattern in which the satDNA repeats share a higher intraspecific than interspecific identity. The copies from the armadillos *C. vellerosus* and *T. matacus* are very similar to each other but less similar than the copies of the two *Choloepus* species are between them. The armadillos sequences showed some degree of divergence. They are not entirely mixed and still have an incomplete homogenization, as the copies did

not appear divided into species-specific branches. These results suggest that the SATCHO1 family is divided into four variant groups (Figure 5). The divergence varies from none, partial or total in each of these subgroups, indicating that these variants may be localized in different chromosomes, a situation that allows local genomic homogenization for satDNA variants (Kuhn et al., 2012). In armadillos, the CVE117 and TMA117 may be undergoing concerted evolution.

The bulk of satDNAs identified herein showed repeat units longer than 2,000 bp, but only BVA2291 showed homology to SATCHO2. They share ~69% of pairwise identity. We did not find/retrieve copies of BVA2291 from the *B. variegatus* assembled genome because the majority of the contigs were shorter than 2,291 bp.

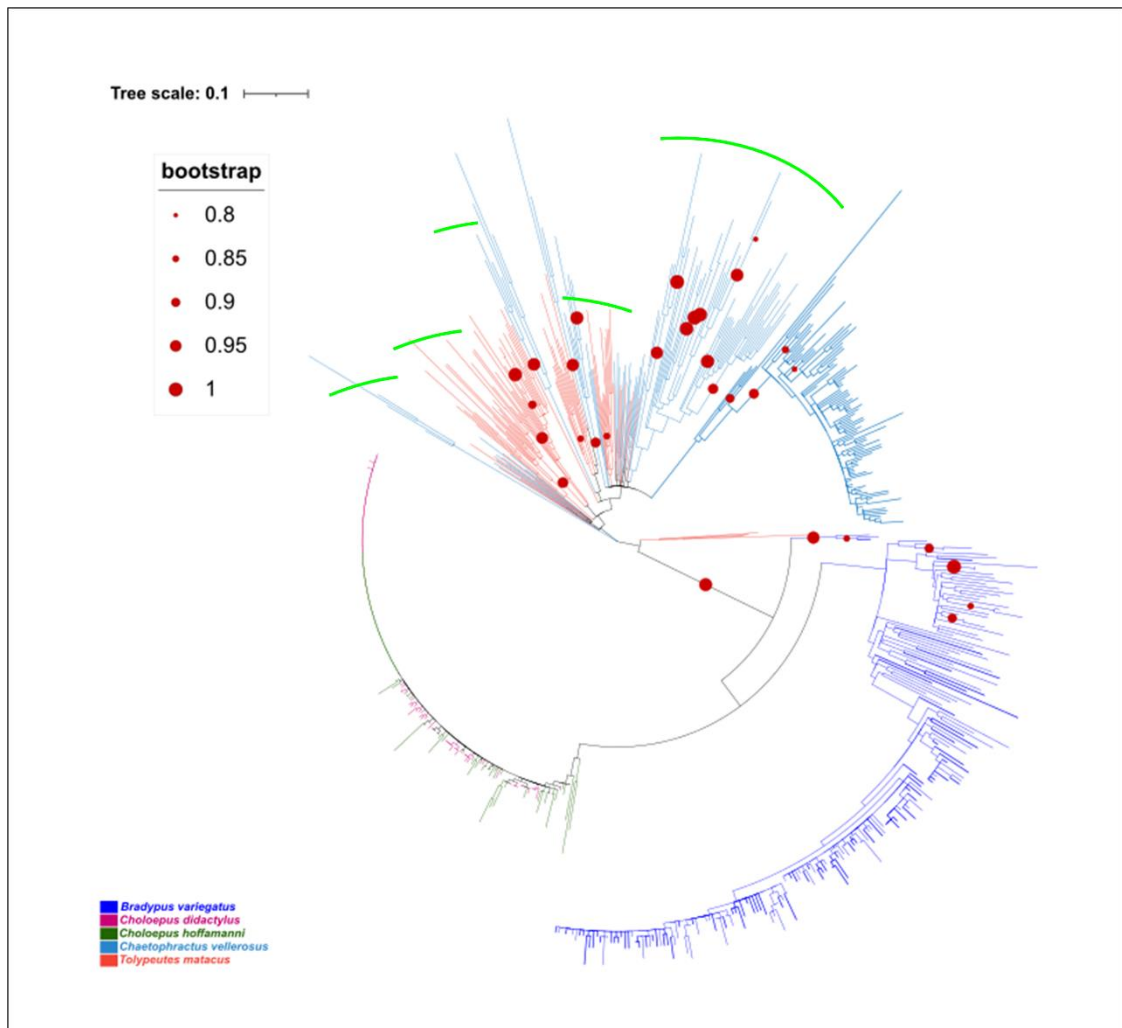


Figure 5: Maximum Likelihood tree of SATCHO1 copies from *Choloepus didactylus* and *C. hoffmanni* and copies from the satDNAs with 116 bp and 117 bp in length (BVA116, BVA117, CVE116, CVE117, and TMA117). The tree was inferred with 1000 replicates and minimum bootstrap support is 80%. Each color represents a species. BVA116 and CVE116 copies are in dark nodes. The bright green lines show the possible five/six groups of copies in *C. vellerosus* and *T. matacus*.

Searching for Centromeric Features in the SATCHO 1 and SATCHO 2 Families

Blastn searches in the nucleotide collection using the 117 bp satDNAs, CVE117 and TMA117, and their putative subfamilies as queries produced similarity hits among the satDNAs from the armadillos *C. vellerosus* and *T. matacus* with three repetitive sequences from the armadillo *Euphractus sexcinctus* (AMD-BglII811, AMD-EcoRI836, and AMD-EcoRI837). These sequences have been mapped in the centromeric regions or in heterochromatic short arms of some chromosomes of *E. sexcinctus* (Liu et al., 2011). The authors did not classify the identified repetitive sequences as any kind of repetitive DNA. This species belongs to the Chlamyphoridae family, the same as the other two armadillos studied herein. Unfortunately, we did not have chromosome spreads to map these putative satDNA sequences by FISH to check if they exhibit a similar hybridization pattern as that described in their relatives (Liu et al., 2011).

Table 2: Identification of CENP-B box motifs in the SATCHO1 family.

SAT_DNA	CENP-B_box motif	bp shared with HSA	bp shared with ECD
ECD	NTTCGNNNNANNCGGGN	9	-
HSA	CTTCGTTGGAAACGGGA	-	9
SATCHO1	CTTTTTGAAAGCAGTC	5	5
DNO_173	CTTTGCCGAGAACGGAG	4	6
*117_BVA	ACTTTTTTGAACCGAGT	4	5
*117_BVA	CTTTTTGAACCGAGTA	5	4
*117_CVE	ACTTCTTTAATTGATGG	2	3
*117_CVE	CTGTGAATGCAGCAGGA	4	5
*117_CVE	CTTCTTTAATTGATGGC	3	4
*117_TMA	GCTCTGTGAATGCAGTA	3	4
3029_ MTR	CTTCGGTGGAAGCCGGT	5	9
811_ CVE	CCTTGTTTAAAAGGAGC	5	5
812_ CVE	AGTTGTGGGAAGCTGGG	4	6
815_ CVE	CCTTGTTTGATTCGTGG	4	6
451_ TMA	ATGCTGTGAATGCAGGA	3	6

***satDNAs with more than one putative motif, the one with more similarity with the ECD was chosen.**
The green highlighted regions in the CENP-B box motif represent the region corresponding to the ECD. The bp shared with more than one motif are in different colors.

We also looked for centromeric features such as the CENP-B box motif in the new SATCHO1 satDNA families. We performed alignments among the satDNAs and the motifs of Evolutionary Conserved Domain (ECD – TTCGNNNNANNCGGG) and the Human CENP-B box motif (CTTCGTTGGAAACGGGA). We found CENP-B box-like motifs in all

eight satDNA sequences (Table 2). The best score was found in the sequence MTR3029, which shares 16 bp out of 17 bp with the Human CENP-B box-like motif (Table 2). The best matches in *C. vellerosus* and *T. matacus* were found in the sequences CVE811, CVE812, CVE815, and TMA451.

Using the software EMBOSS we identified many short-dyed palindrome sequences distributed in all the SATCHO1 family and in the BVA2291 satDNA from the SATCHO2 family (Figure 6). RNAfold identified the optimal secondary structures with minimum free energy for all SATCHO1 and SATCHO2 satDNA families.

Figure 6: Small dyad palindromic sequences from a) BVA116, b) BVA117, c) BVA2291, d) MTR3029, e) CVE116, f) CVE117, g) CVE811, h) CVE812, i) CVE815, j) TMA117, k) TMA451.

Discussion

In this work we analyzed the repetitive DNA content in the genomes of six Xenarthra species. Using RepeatExplorer2 and TAREAN we verified that the repetitive content varies from ~25% in *Tamandua tetradactyla* to ~45% in *Chaetophractus vellerosus*. LINEs are the most abundant sequences, followed by unclassified clusters of repetitive DNAs, LTRs, and satDNAs. The unclassified clusters are undescribed repetitive elements that may belong to the groups of MITEs, SINEs, or Helitrons, for example (Novák et al., 2020). Our results are in line with the recent curation of TEs in 248 placental mammal genomes, in which LINEs were the most abundant repetitive sequences in the same Xenarthra species studied herein (Osmanski et al., 2023). These authors also identified that *Dasypus novemcinctus* and *C. vellerosus* genomes have SINEs as their principal TE which are recent genome insertions.

The present work has focused on the satDNA content. We found that the satDNA content was variable among the species analyzed, and that armadillos have more satDNAs than the other species. Most satDNAs identified are described here for the first time. Five species have at least one satDNA with a repetition motif with >1.000 bp in length, an uncommon size compared with eukaryotic genomes (Schmidt, 1998; Plohl et al., 2012). Some satDNAs share short stretches of homology with previously identified TEs. Therefore, investigating SINEs and LINEs within Xenarthra genomes proves invaluable, given their association with the emergence and evolutionary dynamics of certain satellite DNA sequences.

We also searched for satDNAs that could belong to the SATCHO1 and SATCHO2 families and investigated their possible centromeric roles in Xenarthra genomes. These satDNAs may facilitate the assembly of centromeric structures and the correct chromosome segregation during cell divisions (Garrido-Ramos, 2017). These two families are the most abundant satDNAs in the genomes of the two-toed sloths *Choloepus didactylus* and *C. hoffmanni*, they are located at the centromeric regions of these two species, and showed features of satDNAs with functional centromeres roles (Sena et al., 2020). We identified 117 bp satDNAs in the genomes of *B. variegatus*, *C. vellerosus*, and *T. matacus*, and these repeats have the same size as the repeats from SATCHO1 previously identified by Sena et al. (2020) in *Choloepus hoffmanii*. *Bradypus variegatus* and *C. vellerosus* also have a 116 bp satDNA. These two satDNAs are members of SATCHO1. The sloth *B. variegatus* satDNAs BVA116 and BVA117 share more identical sites with SATCHO1 (>76% nucleotide identity) than all

the others (Table 1). These results agree with the phylogenetic relationship between *Bradypus* and *Choloepus*, the closest genera analyzed herein with an estimated mean divergence time of ~37 million years ago (Mya) (Delsuc et al., 2019; Casali et al., 2022). The only satDNA in the genome of *M. tridactyla* with some identity with SATCHO1 was the higher order (HOR) satDNA MTR3029 (~60.5% of identity) (Table 1.) *M. tridactyla* and *Choloepus* are part of the order Pilosa but are allocated in different suborders, whose estimated divergence time is ~58Mya (Gibb et al., 2016; Varela et al., 2019; Casali, 2021). The armadillo species that we analyzed also have satDNAs in HOR organization made up by multiples of the basic ~117 bp repeat unit. The armadillos *C. vellerosus* and *T. matacus* are phylogenetically more distantly related to *Choloepus* than the other three species studied herein. The estimated mean divergence time between armadillos and sloths is ~60 Mya (Gibb et al., 2016; Casali, 2021), which is in agreement with the low nucleotide identity among their 117 bp satDNAs and the SATCHO1 sequence (Table 1). All our results suggest that these satDNAs (BVA116, BVA117, MTR3029, CVE116, CVE117, CVE811, CVE812, CVE815, TMA117, and TMA451) are members of SATCHO1 (Figure 2; Table 1). Unexpectedly, we did not find satDNAs sharing homology with SATCHO1 in *T. tetradactyla*.

The Maximum Likelihood analysis with copies of satDNAs BVA116, BVA117, CVE116, CVE117, TMA117, and SATCHO1 from *C. didactylus* and *C. hoffmanni* resulted in a tree with a topology that coincides with the Xenarthra phylogenetic tree proposed by Gibb et al. (2016) (Fig. 4). The sloths and the armadillos were divided into two great branches. SATCHO1 copies from CDI and CHO are highly similar and did not separate into two different branches. The same happened with the copies from CVE and TMA, which are also very similar to each other and did not separate into different branches. Nevertheless, there are some arrangements among their copies, forming five or six groups. The copies from CVE116 did not mix with CVE117 and TMA117 but occupied a branch close to them. The BVA116 copies were not totally separate from BVA117. These sequences, especially CVE117 and TMA117, may be located in different chromosomes or chromosome regions as Liu et al. (2011) described. FISH mapping with long reads may confirm this hypothesis.

Considering that SATCHO1 is a centromeric satDNA with centromeric features, these new sequences from the SATCHO1 family may also be centromeric. We did not have chromosome preparations from the species analyzed in order to map these satDNAs. Then, to better characterize all satDNAs identified herein, we ran *in silico* analyses of their sequences,

searching for centromeric features. Melters et al. (2013) described that some eukaryotic genomes have centromeric regions enriched by tandem repeat sequences, which have a set of features related to the kinetochore assembly (Kasinathan & Henikoff, 2018). Such sequences have putative secondary structures with Minimum Free Energy (MFE), an enrichment of small dyad symmetries, HORs, and putative CENP-B boxes. Our results indicate the presence of these centromeric features in all sequences from the SATCHO1 family. They have an enrichment of small dyad symmetries, may form a putative secondary structure, and putative motif of CENP-B box on their sequences.

In functional centromeres from Old World monkeys, horse, chicken, fission yeast, and plants, there are small dyad symmetries (< 10 bp) which are hotspots for secondary structures and curvature non-B-DNA conformation, like hairpins, cruciforms, triplexes, and other structures (Kasinathan & Henikoff, 2018). Great apes and mouse have functional centromeres characterized by the presence of HORs and the CENP-B box (Muro et al., 1992; Masumoto et al., 2004). All these characteristics cause secondary structures where CENP-A is deposited allowing proper kinetochore assembly.

Interestingly, Liu et al. (2011), using restriction enzyme assays on the armadillo *Euphractus sexcinctus* genome, described some repetitive DNAs, including three different centromeric sequences (AMD-BglII811, AMD-EcoRI836, and AMD-EcoRI837). However, although the authors did not describe these centromeric repetitive sequences as satDNAs, they share ~75% of pairwise identity with our CVE117 and TMA117 satDNAs. The putative satDNAs identified by Liu et al. (2011) have >800 bp in length and contain internal subunits with ~113 bp and ~117 bp, like a HOR array. The sequences AMD-BglII811 and EcoRI837 had the same distribution pattern, and in FISH experiments mapped to the centromeric region of 18 chromosome pairs (3, 4, 7, 8, 10-13, 16-25) and the heterochromatic region of the short or long arm of pairs 3, 4, 7, and 19 (Liu et al., 2011). The AMD-EcoRI836 mapped onto the centromeric region of pair 12. Two other different repetitive sequences, with no nucleotide identity with SATCHO1, mapped in the centromeric region of pairs 12 (AMD- Eco RI 832) and 9 AMD- Eco RI 832.

We verified that CVE811, CVE812, CVE815, and TMA451 are in HOR organization, as the centromeric putative satDNAs present in *E. sexcinctus*. In mouse and human genomes, the HOR satellites are intimately related to the functional centromere (Talbert & Henikoff, 2022). Primates, for instance, have the alpha satDNA (~171 bp repeat) as the main

centromeric and most abundant satDNA (Garrido-Ramos, 2017; Ahmad et al., 2020). In the human chromosomes, the alpha satDNA is fundamental for kinetochore assembly and shows different organizations among different chromosomes and within the centromeric and pericentromeric regions (Alexandrov et al., 2001; Logsdon & Eichler, 2022). In the centromere core, it has a HOR organization with higher pairwise identity inter-copies than the peripheric centromere region. However, the latter region has simple copies of alpha, not in HOR, and other satDNA variants of alpha sequences (Logsdon & Eichler, 2022).

The HOR satDNAs from armadillos have their repetitive structure made up by multiple units of CVE117 in the *C. vellerosus* and TMA117 in *T. matacus*. Despite the three CVE HOR satDNAs having similar monomer lengths, they share ~75% of pairwise identity, suggesting that they are different sequences. However, we found that in the contigs from the CVE assembled genome with hits with these satellites, the size of the main tandemly repeated sequence varied among these three satDNAs, with most sequences showing 811 bp. Similarly, the human genome has different satDNAs in the centromeric region, and these sequences are variations of the alpha satDNA. The armadillo *E. sexcinctus* also has different satDNAs in HOR repeats in the centromeric area. Further analyses using long reads and FISH mapping of CVE11, CVE812, CVE815, and TMA451 may allow to verify if these satDNAs are related to the centromeric/pericentromeric regions.

A search for satDNAs from the SATCHO2 family resulted in the identification of only one homologous sequence (BVA2291). BVA2291 and SATCHO2 share ~69% of pairwise identity and have centromeric characteristics, such as a likely CENP-B box, enrichment of small palindromic sequences, and putative secondary structures with MFE. SATCHO2 was mapped by FISH to the centromeric region of all chromosomes of *C. hoffmanni* (Sena et al., 2020). BVA2291 and SATCHO2 are satDNAs with an uncommon repetition unit size (> 1 kbp). Plants and animals, in general, have satDNAs around 150-360 bp in length (Garrido-Ramos, 2017). Large satDNAs with >1 kbp are uncommon or underrepresented. Interestingly, we identified several potential satDNAs with large motif sizes in four species genomes (*B. variegatus*, *M. tridactyla*, *C. vellerosus*, and *T. matacus*) (Supplementary File 1). Fifteen out of the 39 new satDNAs identified by RepeatExplorer2, have repeats >1 to 2.9 kbp long (BVA2291, BVA2538, BVA2491, MTR2430, MTR1550, MTR1830, CVE2860, CVE2746, CVE1261, TMA2225, TMA2013, TMA1206, TMA1834, TMA2844, and TMA2844) and seven have repeats >3 kbp long (BVA6441, MTR8286, MTR3029, MTR4268, CVE4176,

CVE7849, and TMA3521) (Supplementary File 1). In mammals, there are some examples of large satDNAs with >1 kbp, as the walbrep and kervRep, recently described in the red-necked wallaby (*Notamacropus fogriseus*) (Hayashi et al., 2022b; Koga et al., 2023a, b, c), the 1.715 (SATI), 1.709 (SATIV) and the satDNA 1.711b, from bovids (Taparowsky and Gerbi, 1982; Skowronski et al., 1984; Adega et al., 2006), and CapA from *Cebus apella* (Fanning et al., 1993). Some large satDNAs have been reported as megasatellites, sequences with >90 bp up to 10 kb in length and basic motifs multiple of three, a class of large minisatellites (Thierry et al., 2008; Tekaia et al., 2013; Descorps-Declère & Richard, 2022). Recent research in which megasatellites of 58 vertebrate genomes were screened led to the suggestion that this class of repetitive DNA is young compared with fungi megasatellites and is still under an ongoing process of formation and evolution (Descorps-Declère & Richard, 2022). These authors focused on megasatellites present in coding regions and suggested that in vertebrates, megasatellite formation occurs with simultaneous events of duplication and expansion, followed by mutations, and propagates by 'jumping' between two unrelated genes or by gene conversion between paralogs. They also suggested that megasatellites may occur in heterochromatic regions (Descorps-Declère and Richard, 2022). The megasatellites walbrep and kervRep, for instance, originated from the endogenous retrovirus walb and kerv and are present in large heterochromatic portions (Hayashi et al., 2022b; Koga et al., 2023a, b, c). Thus, megasatellites are an intriguing type of repetitive sequence that deserves more attention, and the analysis of Xenarthra genomes and its repetitive content may help to understand this class of tandem repeats.

In conclusion, using *in silico* tools, we identified 39 new satDNA sequences in six Xenarthra species. From this total, 11 sequences exhibited some nucleotide identity with SATCHO1 and SATCHO2 and presented putative centromeric features. We also identified 22 satDNAs with an uncommonly large monomer size (> 1 kbp). Our results indicate that the study of Xenarthra genomes, mainly their repetitive content, has the potential to yield important data regarding mammalian genome evolution.

References

Ahmad, S. F., Singchat, W., Jehangir, M., Suntronpong, A., Panthum, T., Malaivijitnond, S., & Srikulnath, K. (2020). Dark Matter of Primate Genomes: Satellite DNA Repeats and Their Evolutionary Dynamics. *Cells*, *9*(12), 2714. https://doi.org/10.3390/cells9122714

Alexandrov, I., Kazakov, A., Tumeneva, I., Shepelev, V., & Yurov, Y. (2001). Alpha-satellite DNA of primates: Old and new families. *Chromosoma*, *110*(4), 253–266. https://doi.org/10.1007/s004120100146

Alkan, C., Cardone, M. F., Catacchio, C. R., Antonacci, F., O'Brien, S. J., Ryder, O. A., Purgato, S., Zoli, M., Della Valle, G., Eichler, E. E., & Ventura, M. (2011). Genome-wide characterization of centromeric satellites from multiple mammalian genomes. *Genome Research*, *21*(1), 137–145. https://doi.org/10.1101/gr.111278.110

Bao, W., & Jurka, J. (2010). Origin and evolution of LINE-1 derived "half-L1" retrotransposons (HAL1). *Gene*, *465*(1–2), 9–16. https://doi.org/10.1016/j.gene.2010.06.005

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., & Sayers, E. W. (2012). GenBank. *Nucleic Acids Research*, *41*(D1), D36–D42. https://doi.org/10.1093/nar/gks1195

Casali, D. de M. (2021). *Exploring morphological data partitioning and reassessing the phylogeny, divergence times and morphological evolution in Pilosa (Mammalia: Xenarthra)*. [Doctoral Dissertation, UFMG]. Institutional Repository of UFMG. https://repositorio.ufmg.br/handle/1843/38557

Churakov, G. (2005). A Novel Abundant Family of Retroposed Elements (DAS-SINEs) in the Nine-Banded Armadillo (*Dasypus novemcinctus*). *Molecular Biology and Evolution*, *22*(4), 886–893. https://doi.org/10.1093/molbev/msi071

Delsuc, F., Ctzeflis, F. M., Stanhope, M. J., & Douzery, E. J. P. (2001). 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*, *268*(1476), 1605–1615. https://doi.org/10.1098/rspb.2001.1702

Delsuc, F., Vizcaíno, S. F., & Douzery, E. J. (2004). Influence of Tertiary paleoenvironmental changes on the diversification of South American mammals: A relaxed molecular clock study within xenarthrans. *BMC Evolutionary Biology*.

Descorps-Declère, S., & Richard, G.-F. (2022). Megasatellite formation and evolution in vertebrate genes. *Cell Reports*, *40*(11), 111347. https://doi.org/10.1016/j.celrep.2022.111347

Edgar, R. C. (2004). MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, *32*(5), 1792–1797. https://doi.org/10.1093/nar/gkh340

Fanning, T. G., Seuànez, H. N., & Forman, L. (1993). Satellite DNA sequences in the New World primate *Cebus apella* (Platyrrhini, Primates). *Chromosoma*, *102*(5), 306–311. https://doi.org/10.1007/BF00661273

Garrido-Ramos, M. (2017). Satellite DNA: An Evolving Topic. *Genes*, *8*(9), 230. https://doi.org/10.3390/genes8090230

Gibb, G. C., Condamine, F. L., Kuch, M., Enk, J., Moraes-Barros, N., Superina, M., Poinar, H. N., & Delsuc, F. (2016). Shotgun Mitogenomics Provides a Reference Phylogenetic Framework and Timescale for Living Xenarthrans. *Molecular Biology and Evolution*, *33*(3), 621–642. https://doi.org/10.1093/molbev/msv250

Kasinathan, S., & Henikoff, S. (2018). Non-B-Form DNA Is Enriched at Centromeres. *Molecular Biology and Evolution*, *35*(4), 949–962. https://doi.org/10.1093/molbev/msy010

Kohany, O., Gentles, A. J., Hankus, L., & Jurka, J. (2006). Annotation, submission and screening of repetitive elements in Repbase: Repbase Submitter and Censor. *BMC Bioinformatics*, *7*(1), 474. https://doi.org/10.1186/1471-2105-7-474

Kuhn, G. C. S., Küttler, H., Moreira-Filho, O., & Heslop-Harrison, J. S. (2012). The 1.688 Repetitive DNA of *Drosophila*: Concerted Evolution at Different Genomic Scales and Association with Genes. *Molecular Biology and Evolution*, *29*(1), 7–11. https://doi.org/10.1093/molbev/msr173

Kumar, S., Stecher, G., Li, M., Knyaz, C., & Tamura, K. (2018). MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Molecular Biology and Evolution*, *35*(6), 1547–1549. https://doi.org/10.1093/molbev/msy096

Letunic, I., & Bork, P. (2021). Interactive Tree Of Life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research*, *49*(W1), W293–W296. https://doi.org/10.1093/nar/gkab301

Liao, D. (1999). Concerted Evolution: Molecular Mechanism and Biological Implications. *The American Journal of Human Genetics*, *64*(1), 24–30. https://doi.org/10.1086/302221

Liu, Y., Ye, J., Fu, B., Ng, B. L., Wang, J., Su, W., & Nie, W. (2011). Molecular Cytogenetic Characterization of the Genome Organization of the 6-Banded Armadillo *(Euphractus sexcinctus)*. *Cytogenetic and Genome Research*, *132*(1–2), 31–40. https://doi.org/10.1159/000318706

Logsdon, G. A., & Eichler, E. E. (2022). The Dynamic Structure and Rapid Evolution of Human Centromeric Satellite DNA. *Genes*, *14*(1), 92. https://doi.org/10.3390/genes14010092

Lower, S. E., Dion-Côté, A.-M., Clark, A. G., & Barbash, D. A. (2019). Special Issue: Repetitive DNA Sequences. *Genes*, *10*(11), 896. https://doi.org/10.3390/genes10110896

Masumoto, H., Nakano, M., & Ohzeki, J. (2004). The role of CENP-B and a-satellite DNA: de novo assembly and epigenetic maintenance of human centromeres. *Chromosome Research*, *12*, 543–556. https://doi.org/10.1023/B:CHRO.0000036593.72788.99

Melters, D. P., Bradnam, K. R., Young, H. A., Telis, N., May, M. R., Ruby, J., Sebra, R., Peluso, P., Eid, J., Rank, D., Garcia, J., DeRisi, J. L., Smith, T., Tobias, C., Ross-Ibarra, J., Korf, I., & Chan, S. W. (2013). Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biology*, *14*(1), R10. https://doi.org/10.1186/gb-2013-14-1-r10

Muro, Y., Masumoto, H., & Okazakit, T. (1992). Centromere Protein B Assembles Human Centromeric a-satellite DNA at the 17-bp Sequence, CENPB Box. *The Journal of Cell Biology*, *116*.

Nishihara, H., Kuno, S., Nikaido, M., & Okada, N. (2007). MyrSINEs: A novel SINE family in the anteater genomes. *Gene*, *400*(1–2), 98–103. https://doi.org/10.1016/j.gene.2007.06.003

Novák, P., Neumann, P., & Macas, J. (2020). Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nature Protocols*, *15*(11), 3745–3776. https://doi.org/10.1038/s41596-020-0400-y

Osmanski, A. B., Paulat, N. S., Korstian, J., Grimshaw, J. R., Halsey, M., Sullivan, K. A. M., Moreno-Santillán, D. D., Crookshanks, C., Roberts, J., Garcia, C., Johnson, M. G., Densmore, L. D., Stevens, R. D., Zoonomia Consortium†, Rosen, J., Storer, J. M., Hubley, R., Smit, A. F. A., Dávalos, L. M., … Zhang, X. (2023). Insights into mammalian TE diversity through the curation of 248 genome assemblies. *Science*, *380*(6643), eabn1430. https://doi.org/10.1126/science.abn1430

Plohl, M., Meštrović, N., & Mravinac, B. (2012). Satellite DNA Evolution. Em M. A. Garrido-Ramos (Org.), *Genome Dynamics* (Vol. 7, p. 126–152). S. Karger AG. https://doi.org/10.1159/000337122

Redi, C. A., & Capanna, E. (2012). Genome Size Evolution: Sizing Mammalian Genomes. *Cytogenetic and Genome Research*, *137*(2–4), 97–112. https://doi.org/10.1159/000338820

Schmidt, T. (1998). Genomes, genes and junk: The large-scale organization of plant chromosomes. *Trends in Plant Science*, *3*(5), 195–199. https://doi.org/10.1016/S1360-1385(98)01223-0

Sena, R. S., Heringer, P., Valeri, M. P., Pereira, V. S., Kuhn, G. C. S., & Svartman, M. (2020). Identification and characterization of satellite DNAs in two-toed sloths of the genus *Choloepus* (Megalonychidae, Xenarthra). *Scientific Reports*, *10*(1), 19202. https://doi.org/10.1038/s41598-020-76199-8

Stitou, S., Díaz De La Guardia, R., Jiménez, R., & Burgos, M. (1999). Isolation of a Species-Specific Satellite DNA with a Novel CENP-B-like Box from the North African Rodent *Lemniscomys barbarus*. *Experimental Cell Research*, *250*(2), 381–386. https://doi.org/10.1006/excr.1999.4516

Sullivan, L. L., Chew, K., & Sullivan, B. A. (2017). α satellite DNA variation and function of the human centromere. *Nucleus*, *8*(4), 331–339. https://doi.org/10.1080/19491034.2017.1308989

Superina, M., & Loughry, W. J. (2015). Why do Xenarthrans matter?: Table 1. *Journal of Mammalogy*, *96*(4), 617–621. https://doi.org/10.1093/jmammal/gyv099

Talbert, P. B., & Henikoff, S. (2022). The genetics and epigenetics of satellite centromeres. *Genome Research*, *32*(4), 608–615. https://doi.org/10.1101/gr.275351.121

Tekaia, F., Dujon, B., & Richard, G.-F. (2013). Detection and Characterization of Megasatellites in Orthologous and Nonorthologous Genes of 21 Fungal Genomes. *Eukaryotic Cell*, *12*(6), 794–803. https://doi.org/10.1128/EC.00001-13

Thakur, J., Packiaraj, J., & Henikoff, S. (2021). Sequence, Chromatin and Evolution of Satellite DNA. *International Journal of Molecular Sciences*, *22*(9), 4309. https://doi.org/10.3390/ijms22094309

Thierry, A., Bouchier, C., Dujon, B., & Richard, G. F. (2008). Megasatellites: A peculiar class of giant minisatellites in genes involved in cell adhesion and pathogenicity in *Candida glabrata*. *Nucleic Acids Research*, *36*(18), 5970–5982. https://doi.org/10.1093/nar/gkn594

Varela, L., Tambusso, P. S., McDonald, H. G., & Fariña, R. A. (2019). Phylogeny, Macroevolutionary Trends and Historical Biogeography of Sloths: Insights From a Bayesian Morphological Clock Analysis. *Systematic Biology*, *68*(2), 204–218. https://doi.org/10.1093/sysbio/syy058

Waters, Paul D., Dobigny, G., Pardini, Amanda T., & Robinson, Terence J. (2004). LINE-1 distribution in Afrotheria and Xenarthra: Implications for understanding the evolution of LINE-1 in eutherian genomes. *Chromosoma*, *113*(3). https://doi.org/10.1007/s00412-004-0301-9

Supplementary File 1: TAREAN basic results of satellite DNA identification in the genomes of six Xenarthra species.

Species	Cluster	Monomer length (bp)	Satellite Confidence	Satellite Probability	Genome Proportion %	C-Value	P-Value	AT Content %	RepBase	Nucleotide Collection	Blast againt WGS XEN
	1	116	HC	0.992	1.700	0.992	0.977	61.21	Run didn´t produce any hit	CDI: XR_005213860.1	Hits with 100% of query cover (QC) in CDI, CHO, CVE, and TMA
	9	172	HC	0.815	1.100	0.944	0.906	62.21	satélite ALR2 (151pb)	Hits with 100% f QC with macaca mulata (1e-61 of e-value; per id: 91.86%)	Hits with 97% of QC in CHO
	17	2291	HC	0.767	0.620	0.982	0.888	58.8	4 small fragments of TEs	MT505309.1 / MT505310.1 / XR_005215011.1	Hits with 100% of QC in CDI, CHO and 1% in CVE
***Bradypus variegatus* (BVA)**	76	2538	HC	0.955	0.026	0.973	0.909	29.94	6 fragments of TEs and 1 LTR	Hits with small fragments in different species ou of Xenarthra group	Hits with 8-100% of QC in CDI e CHO; 4-5% TMA; 3-5% in CVE;10% in DNO; 3-8% MTR and TTE
	32	2491	LC	0.0931	0.360	0.856	0.965	49	4 fragments of TEs	Hits with 6-98% of QC in different species outside theXenarthra group	Hits with 3% of QC in CDI, CHO, MTR, TTE, and DNO; 8% in CVE and 7-8% in TMA
	44	6441	LC	0.6240	0.180	0.972	0.827	51.62	8 fragments of different TEs	Hits with 3-62% of QC in different rodentia species	Hits with 9-52% of QC in MTR; 7-26% in TTE; 6-12% in CDI and DNO; 4-14% in CHO; 3-7% in CVE and TMA
	48	117	LC	0.0453	0.130	0.914	0.742	58.97	Run didn´t produce any hit	SATCHO1	Hits with the clones of SATCHO1
	26	696	HC	0.994	0.230	0.994	0.966	48.99	run didn´t produce any hit	no significant similarity found	Hits with 89% of QC in TTE
	29	8286	HC	0.975	0.220	0.984	0.951	48.26	18 small fragments of Tes	Hits with 1-23% of QC in different Species outside theXenarthra group	Hits with 76-92% of QC in TTE; 1-88% in CDI; 1-82% in CHO; 1-23% in BVA; 1-18% in CVE; 1-93% in TMA
	56	2430	HC	0.993	0.025	0.993	0.965	56.13	Small fragments of 1 DNA (Mariner/Tc1) and 1 LTR retrotransposon (Gypsy)	no significant similarity found	Hit with 99% of QC in TTE; 1-2% of QC in CDI e CHO
	57	1550	HC	0.981	0.025	0.997	0.963	53.03	4 small fragments of TEs	Hits with 6-7% of QC in different species outside theXenarthra group	Hits with 99% of QC in TTE (e-value=0.0 e per id=70.90%); 7% in BVA, CDI, CHO, DNO, and TMA (e-value=1e-16 e per id.=80%)
	58	3029	HC	0.981	0.025	0.983	0.960	52.53	2 small fragments of TEs	Hits with 6-7% of QC in different species outside theXenarthra group	Hits with 7-76% of QC in TTE (e-value: 8e-155; per id: 82%); 1-3% in BVA, CDI, CHO, CVE, DNO, and TMA;
***Myrmecophaga tridactyla* (MTR)**	73	1830	HC	0.989	0.013	0.994	0.946	23.44	4 small fragments of TEs	Hits with 6-7% of QC in different species outside theXenarthra group	Hits with 6-78% of QC in TTE (e-value: 8e-165; per id: 96%); 6-17% of QC in BVA, CDI, CHO, DNO, CVE, and TMA (e-value: 3e-68; per id: 94%);
	8	156	LC	0.0793	0.770	0.807	0.931	46.79	Fragment of TE (Penelope-like retroelement)	Hits with 100% of QC in different genomes (e-value:3e-09; Per.id=72.39%), these hits are related with telomeric regions	Hits with 100% of QC in TTE, CDI, CHO, and DNO (e-value=1e-15; per.id=75%); 39% in BVA (e-value=4e-06; per id=80.65%) em BVA; 99% in TMA (e-val=0.019; per id=69.75%); 99% in CVE (e-value=1e-06; per id=68.39%)
	63	4268	LC	0.0666	0.019	0.869	0.831	50.28	8 fragments of different TEs (including 859pb of L1-2 from CHO)	Hits with different contigs from CHO	Hits with 98% of QC in TTE (e-value: 0.0; per id: 88%); 8-50% of QC in CDI (e-value: 6e-177; per id: 73%); 8-35% of QC in CHO (e-value: 2e-178; per id: 73%); 8-36% in TMA (e-value: 4e-133; per id: 72%); 8-

											26% in CVE (e-value: 8e-128; per id: 71%); 8-16% of QC in BVA (e-value: 4e-55; per id: 74%),
***Tamandua tetradactyla* (TTE)**	46	745	LC	0.0727	0.20	0.856	0.868	59.06	1 small fragment of the satDNA BMSAT1 from blue whale (34bp)	Hits with 10% of QC in different species, including plants (e-value: 2e-05; per id: 79%)	Hits with 99% of QC in MTR (e-value: 1e-145; per id: 76%); 5-8% of QC in BVA, DNO, CDI, CHO, CVE, and TMA
	81	35	LC	0.0154	0.01	0.710	0.842	42.86	Run didn´t produce any hit	no significant similarity found	Hits with 100% of QC only in MTR (e-value: 1e-08; per id: 100%)
***Chaetophractus vellerosus* (CVE)**	8	812	HC	0.922	1.000	0.976	0.926	59.73	Run didn´t produce any hit	Hits only with *Euphractus sexcinctus* (GU550062.1 / GU550057.1 / GU550061.1); QC of 100%; e-value: 1e-74; per od: 71%	Hits with 100% of QC in TMA (e-value: 1e-43; per id: 70%); 5-13% in CDI and CHO; 6% in BVA (e-value: 0.012; per id: 81%); 5% in MTR (e-value: 0.002; per id: 87%)
	9	116	HC	0.975	0.940	0.976	0.951	59.48	Run didn´t produce any hit	no significant similarity found	Hits with 100% of QC in TMA (e-value: 2e-16; per id: 79%); 32% of QC in CHO (e-value: 0.037; per id: 84%)
	25	2860	HC	0.994	0.370	0.995	0.966	53.92	5 fragmentos de TEs	Hits with 1% QC in *Archocentrus centrarchus* (e-value: 0.024; per id: 83%)	Hits with 99% of QC in TMA (e-value: 0.0; ér id: 69%); 2-28% in DNO (e-value: 2e-16; per id: 65%); 1% of QC in CDI and MTR (e-value: 0.029; per id: 82-93%); 0% of QC in BVA (e-value: 0.045; per id: 100%; 25bp)
	36	117	HC	0.775	0.160	0.957	0.926	60.68	Run didn´t produce any hit	Hits only with *Euphractus sexcinctus* (GU550062.1 / GU550057.1 / GU550061.1)	Hits with 100% of QC in TMA (e-value: 1e-27; per id: 84%)95% of QC in DNO (e-value: 3e-10; per id: 76%); 30% of QC in TTE (e-value: 0.044; per id: 86%)
	58	811	HC	0.986	0.061	0.995	0.974	55.61	Run didn´t produce any hit	Hits only with *Euphractus sexcinctus* (GU550062.1 / GU550057.1 / GU550061.1)	Hits with 90% of QC in TMA (e-value: 2e-28; per id: 67%); 7% of QC in DNO (e-value: 0.025; per id: 76%); 6-8% of QC in CDI and CHO (e-value: 0.002; per id: 78%);
	70	2746	HC	0.976	0.031	0.990	0.937	46.5	5 fragments of TEs (1 of them is the L1-1C_Cho)	Hits with 27% in different spp (e-value: 2e-167/7e-155; per id: 78%) (Xenarthra and not xen)	Hits with 12-76% of QC in TMA (e-value: 1e-178; per id: 68%); 11-57% of QC in TTE (e-value: 7e-163; per id: 72%); 9-49% of QC in CDI DNO (e-value: 2e-169/5e-170; per id: 78%); 2-35% of QC in CHO, BVA, and MTR (e-value: 1e-145; per id: 72%); 1-2% of QC in CDI e CHO
	72	4176	HC	0.937	0.029	0.973	0.923	44.4	7 fragments of different TEs (2 of them are parts of HAL1-2A_Cho)	Hits with ~23-39%of QC in different species	Hits with 6-96% of QC in TMA (e-value: 0.0; per id: 87%); 1-64% of QC in DNO (e-value: 0.0; per id: 73%); 1-39% of QC in CDI, CHO, MTR, and TTE (e-value: 0.0/5e-14; per id: 76%); 1-2% of QC in BVA (e-value: 1e-11; per id: 74%)
	77	815	HC	0.993	0.024	0.990	0.961	57.55	Run didn´t produce any hit	Hits only with *Euphractus sexcinctus* (GU550062.1 / GU550057.1 / GU550061.1)	Hits with 58-100% of QC in TMA (e-value: 4e-27; per id: 68%); 6% of QC in CDI (e-value: 0.002/0.028; per id: 78-79%);
	99	45	HC	0.986	0.015	0.995	0.989	51.11	Run didn´t produce any hit	no significant similarity found -	Hits with 55% of QC in TMA (e-value:

										checar	0.032; per id: 96%)
	48	7849	LC	0.0952	0.098	0.903	0.801	47.52	4 fragments of ERV1	Hits of similarity with rodents and bats (QC: 89%; e-value: 0.0; per id: 78%)	Hits with 3-38% of QC in MTR (e-value: 0.0; per id: 72%); 4-23% of QC in TMA (e-value: 2e-48; per id: 70%); 3-22% of QC in CDI (e-value: 1e-62; per id: 74%); 1-20% of QC in DNO (e-value: 5e-42; per id: 81%); 3-15% of QC in CHO (e-value: 2e-61; per id: 75%);
	128	1261	LC	0.0194	0.010	0.722	0.648	65.98	3 fragments of TEs	Hits with 95-100% of QC in different species (Xenarthra and no Xen) (e-value: 1e-172/0.0; per id: 82%)	Hits with 86-100% of QC in BVA, CDI, CHO, MTR, TTE, TMA and DNO (e-value: 0.0; per id: 89%);
***Dasypus novemcinctus* (DNO)**	7	173	HC	0.917	0.63	0.956	0.ज़	63	Run didn´t produce any hit	No significant similarity found.	No significant similarity found.
	41	451	HC	0.975	0.410	0.982	0.953	64.52	Run didn´t produce any hit	Hit of 27% of QC in ESE (e-value: 0.012; per id: 71%)	Hits with 99% of QC in CVE (e-value: 2e-30/-1e-14, per id: 62-71%)
	46	2225	HC	0.741	0.320	0.946	0.888	63.1	4 fragments of TEs	Hit with 2% of QC in Gossyploidis kirkii (e-value:0.007; per id:83.64%)	Hits with 3-29% of QC in CVE and DNO (e-value: 5e-17; per id: 67%); 1-2% of QC in BVA, CDI, CHO, MTR, and TTE (e-value: 0.022-0.035; per id: 84-88%)
	73	2013	HC	0.889	0.034	0.964	0.904	41.73	7 fragments of TEs (1 of them is L1-1_DN (579pb))	Hits with 16-30% of QC in DNO and CHO (e-value: 8e-108; per id: 78%)	Hits with 69% of QC in CVE (e-value: 0.0; per id: 83%); 63% of QC in CDI (e-value: 1e-68 ; per id: 76%); 24-33% of QC in BVA, CHO, MTR, and TTE (e-value: 5e-55; per id: 75%)
	1	117	LC	0.10400	1.400	0.893	0.854	65.81	Run didn´t produce any hit	Hits only with *Euphractus sexcinctus* (GU550062.1 / GU550057.1 / GU550061.1)	Hits with 58-100% of QC in DNO and CVE (e-value: 2e-24; per id: 85%); 64% of QC in CDI, CHO, and TTE (e-value: 2e-05; per id: 91%); 56% in MTR (e-value: 6e-06; per id: 82%); 43% of QC in BV (e-value: 4e-04; per id: 80%)
***Tolypeutes matacus* (TMA)**	31	1206	LC	0.00828	0.570	0.725	0.449	47.93	1202bp correspondi to the LINE_1 (TolMat-1.65)	Hits with 12-99% of QC in DNO (e-value: 1e-40/0.0; per id: 85%)	Hits with 100% of QC in CVE and DNO (e-value: 0.0; per id: 88%); 5-21% of QC in BVA, CDI, CHO, MTR, and TTE (e-value: 2e-19; per id: 80%)
	34	1834	LC	0.01810	0.530	0.706	0.592	64.34	10 fragments of different repetitive DNA	Hits with 47-63% of QC in different species outside the Xenarthra group, the best values in DNO (e-value: 0.0/9e-177; per id: 75-91.96%)	Hits with 63% of QC in BVA, CDI, CHO, MTR, TTE, CVE and DNO (e-value: 0.0; per id: 83-93%)
	36	2844	LC	0.03400	0.520	0.775	0.712	65.08	7 fragments of different TEs (2 of them are TolMat-6.11 and TolMat-6.573 (both are from L1 family))	Hits with 3-14% of QC in different species outside the Xenarthra group, the best values are in DNO genome (e-value: 1e-18; per id: 74%)	Hits with 90% of QC in CVE (e-value: 1e-168: per id: 72%); 19-70% in DNO (e-value:2e-98; per id: 70%); 2-5% of QC in BVA, CDI, CHO, MTR, TTE (e-value:5e-12; per id: 80%);
	52	3521	LC	0.02780	0.180	0.747	0.635	49.76	6 fagments of TEs (4 are parts of TolMat-5.52_int and 1 of DasNov-5.218_LTR)	Hits with 4-34% of QC in different species outside the Xenarthra group, the values in DNO genome (e-value: 1e-108; per id: 70%)	Hits with 88% in DNO (e-value: 2e-171; per id: 78%); 60% of QC in CVE (e-value: 0.0; per id: 87%); 1-12% of QC in TTE (e-value: 1e-15; per id: 73%); 1-6% in BVA (e-value: 2e-07; per id: 81%)
	55	27	LC	0.01580	0.150	0.712	0.671	33.33	Run didn´t produce any hit	Hits with 70-100% of QC in	Hits with 100% of QC in CVE and

									different speceis outside the Xenarthra group (e-value: 4e-04; per id: 100%)	DNO (e-value: 3e-06/6; per id: 100%); 69-100% of QC in TTE (e-value: 0.11; per id: 100%); 55-74% in BVA, CDI, CHO, and MTR (e-value: 1.6-25; per id: 100%)
84	792	LC	0.03680	0.018	0.956	0.724	64.25	2 small fragments of TEs	Hit with 4% of QC in *Sarcophaga subvicina* (fly) (e-value: 0.023; per id: 96%)	Hits with 8-99% in CVE (e-value: 3e-17; per id: 70%); 5% of QC in BVA, CDI, CHO, TTE, and DNO (e-value: 0.008-0.027; per id: 83-85%)

Supplementary File 2: Matrix of pairwise identity percentage of each pair of SATCHO1 family.

SATDNA	SATCHO1	BVA116	BVA117	MTR3029	CVE116	CVE117	CVE811	CVE812	CVE815	TMA117	TMA451
SATCHO1	0	76.9	78.6	60.5	57.9	56.3	57.5	58.3	55.1	59.4	56.1
BVA116	76.9	0	83.76	57.5	55.93	51.7	54.62	57.14	52.76	57.27	53.33
BVA117	78.6	83.76	0	57.6	54.1	52.07	54.69	55.47	54.82	56.3	56.67
MTR3029	60.5	57.5	57.6	0	59.42	55.07	51.58	51.37	51.73	54.03	53.67
CVE116	57.9	55.93	54.1	59.42	0	67.64	62.39	65.83	66.39	77.78	61.54
CVE117	56.3	51.7	52.07	55.07	67.64	0	76.07	79.49	76.07	85.47	65.81
CVE811	57.5	54.62	54.69	51.58	62.39	76.07	0	75.52	76.01	68.38	55.84
CVE812	58.3	57.14	55.47	51.37	65.83	79.49	75.52	0	73.38	74.36	59.16
CVE815	55.1	52.76	54.82	51.73	66.39	76.07	76.01	73.38	0	73.5	55.99
TMA117	59.4	57.27	56.3	54.03	77.78	85.47	68.38	74.36	73.5	0	71.8
TMA451	56.1	53.33	56.67	53.67	61.54	65.81	55.84	59.16	55.99	71.8	0

Supplementary File 3: Sequence alignment among SATCHO1 and its subfamilies (left side) and the pairwise percentage of identity of each alignment (right side): BVA116 (a, b); BVA117 (c, d); MTR3029 (e, f); CVE116 (g, h); CVE117 (I, j); CVE811 (k, l); CVE812 (m, n); CVE815 (o, p); TMA117 (q, r); TMA451 (s, t). BVA – *Bradypus variegatus*; MTR – *Myrmecophaga tridactyla*; CVE – *Chaetophractus vellerosus*; TMA – *Tolypeutes matacus*.

Supplementary File 4: Satcho1 family consensus sequences identified by TAREAN.

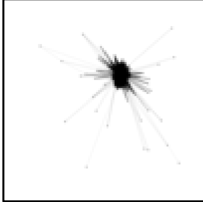
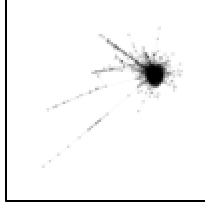
```
>BVA116
GAAAGCAATGTGTTATACAGTTCTGGAAACTAAGTTGAATGCAGCCTTCATTCACTTCTATGTGTTTCCCTTTGGCTGAAATAGCACTTTTTTA
ACAGAGACAGTTAGAGCCCTAC
>BVA117
GAAAGCAATATGTTTTACAGTTCTGGAAACTAACTTGAATGCAGCCTTGCCTGACTTGTATGTGTTTTCCCTGGGCTGAAAAATCACTTTTTTG
AACCGAGTAACTTGGAGCCCTAC
>BVA2291
GGACAGGGGCCCTTCACTAAACCTCCTCCTTCCCCACAGCAAACTGAGATTTGCCCTTGTGCGTCTCACCATAGCATGCTAAAAGGCGAGTCAA
AGGTCATCACCCTTCCGGCTTTCAGTTGTTCCTCGGGGGTGGAAATTCATCCTGAAATCGCTACTCCTTCCTCCAAGGGCTACAACACTTCCAA
AATCGCCTGCACACTGTGGAAGAAACATATGTTTCATACCACAGAGCCGCCAGTAGCCTAGGGTCCTGTTGCCTAGCAACAGGACGCTGTGTT
TGACAGTTGAGGGGGGCAGCGGAGCTCGCGCACGCGCACTGCGAGCACCCGCGGGAATTGAAAGCCGCTTGCGAAACTCTGCTGTCCCTCCCTG
TTCCCTCCACCGATCCGATCCCCAAAGAGATTTCTTCGCTGAGTTTTCTTAACTTGGCATGGACTGCTGCATATTGGGAACACCTGCGTAAATG
GATGCCTATTTAGAGTACTGGGAATAACTCTGTTAGCGATGGTGTGCTTTGTAGCTGATTCTGCAAAGAAGGCAGAGCCTCAAAGCCTTTATTT
GAAACTGGATATGAAGTTTTGTAATGTAGTGTAACTGTAAGTAGATGACAGGAGAGGAAAACAGGAAGGATAGGTGTGTGCAGCAATAAAAATT
ATCTCAATGAAAGTACACGGACATTAGGACGACTAGTACTTGTACTGTGAGAACAAATGTCTTGATGATGAACCATGAGTTCAAACGTGTGCAA
TGGAGAAGCAAAGCAATTAATTCAGATAATTTTTTCTTTCCTTCAACTTTTCTTTTCTTTATTAAATGTGGTGTTGTTTGCTGCTTTTGTGTTT
TTGTAGCTTATGGCTAATGTATAAAACGAATTCAATCCATTCAAACAATCAAAATTTAAACTTCTTATTTAAACAATAATAACCATAAGGGCAT
AATGATAATCTTCCAATGCAACGTAAATTTCATCACTTCAATTTCATCGGTTAGAACCTATTGAAAAAAAAAGAGATGGAGGAATATTCTTTTA
AAGGATGTTTCAGCATTTTTTATTTTATCTAGATGAAGAAATTGGTAGAGTTATATTCATATGGAAGGGAATAGATATACCTGAAAAGAAATGT
AATGTTTGCTATGAGACTTGATACTATATTGGTTTTGTTCTTTTAATGTTTCTAACATTAACAGTACTACTAACGTCCTGTTTTGTCTATACAT
CATGATTGTGTAATTGATGGCCCGTGGATCAGGCCCCTGGATGACACTGATGTACTTCAAGCAAGTTAATTTAATGATTATGGCAACATTCCTG
TGAAAGCATAGGAGTATGTGTCTGATGATAAGTGAAATATTATAAGAACAATTAAAGAAATTGATGTGCTTTATTGTGATGCATTTCCAATTAT
TGAAGAATCCCTTTTTTTAACTAGGAATATGAAAGTATATTAACGTGACAAAGAAGTGAACTAAATAAATGAGCTTCACGTGTGTAGGTTTCTC
TCCTCATTAAAGGCCCGGGACAATCACGGTGAGATTACACTGACCTGAATAAAAGGCTTATGAGAAGGGTTGTCACAGTCAATGCCACTTAGGC
ATACTCTGTGGGGTGTCACCAGATTTGGGAATGGAGAGATCATGGAAATCTTCATCCTTAGCAAAATCTTGCTTTCTGACCTGGATCAATATAT
GTGCCACTTCTGACATCAAGTTCTTCTATACACAGGTACACATATCAAGGGACTTGGGGTGATTTTTGGCAGTTATTTTTGAATTATTGTCACT
GACAGAGCCAGGTCATGTTGGGTACCACAACAGACAATTCCCTCAGGGCACTGAATATTCCTGGAGTAAGAAACAAAATGCACTCTGTGTGCAG
CCTCATGCCTGCCCTTGGGTGTTTCCAAACTGCTCCACAATTAGGCCTCCAAAACCTCAACTCAAAACATAGCAGCCTGTCCACATATCATCTG
GTTTCTGAACCACAGGAAAACCTTGATTGTCACTGACCATTTACACAGAGAATGACACCTGTAACAAACGCCAATAGAGCAAGAAATCCCACCT
TTAAGAAAGCCATAGGCCAAACTTTGCGTCAGCTCTCGACTATAGAGAGGACAGACACATCAAAGACGTCCAAATTTTTCTCACTGCCAGAGGG
TTCACTTGACTGAACATGGCCAAGCCTCATGGTGTGTTCGGCTAAGTGAGATTACTGGATGCTGTGCTGCAAGTCTCTGATCCAGCAGTCACTT
CTGTGATTTTACAGAAGAAAGGCAGGTCCTAGCAT
>MTR3029
GAAACAGAAACACCATTTGCGCCCCTGCTGTTGTTAGAGCGAGAAACAGCATGCTGCTGAACACACACTTAGGGCCTATAGAAAGAAGCCTGGA
AACCGGATATGAAAACGACGTCCGTTCTACATGTCAGTAGCTGCTCCTGGGCTTACCTGAGTGTGAAACAGCGTGTTGCTGAAACACTAAGGTC
GTGCGATTTCAGAGTAGCCAGAGAGTCTGCGTGGACAACGAAGGGACTCGGAAGCACCATTTACGTCCCTGCTGTTACTTCGGTGGAAGCCGGT
ATGTAGCTTTCAGACTCACTTGAGCGCTAGCTGAAACTAGCCAGAAATTCATTTTGGAAGACTAAAGTGAAACGAAACCACCGCTTACTCCCCT
GCTGTTATTAGAGGGGGGGAAAGGCCATGTTGCTTAAACGGTACCTTTAAAGCTCTTTGAAAGCCGTCTGAACAGCGGATTGGCAACCCATGTG
CTCTATAGCCCCAACGTGTTTCCCTGCTCTCGATAGAGTGAAAACCAGTGTGCCGCTAAGTCACTCAGTGCTAGAGGCAACAAACTGGAACTAC
AGTTAGGAAATGAAAGGGACAGGGAATAAATATAACCTTTTGCTTCCTTCTCTGCTGAGATTAGCGTGACCAAAAGCATGCTTGCTTAGACACA
CACACATGGTGCATAGCAAAGCAGCCTGAAAACCGCCTTTGAAAACGGTGCGCAATACACGGCCAGCAGCACCCCCTGCTGTTCTTTGAGCGTA
AAAGAGAATGCTGCTTCATCACGAATTTCGTGCTCTTTGAAAGCAGCCTGGAAATCGGATTCGACACGTAAGTGAAACAGACCCACGATTCGCT
CCCCTGCTCTCAGTAGAGCAGCAGGCTGCTCAACAAACAGTGCTTATAGAGAGAAGCCAGAAAATCGGGTATGGGAGTGACATGCATTCTACAG
GACGAATGATGGTTCTGTTTCACTTTCTTTGCGCAACCGAGCTAAAACTAAGCTGCATTAACGCGAGCCCTAACCCTAACTCTAACCATATAAC
GCCAAACACTAACCCTAACCCTAACGCTATCCTTGACCCAAAAGCTAGCCTTAACCCTAACCCTGACCCTGACCCTGACCCTAACCCTAACCCT
AAACCTTACCCTAACCCTAACCCTAACCCTCCCCCTGAGCCTGACCCTGACCCTAACCCCTACACCAACCCTTCAAATAGCAGTGAGGGTTTAA
GGAAGACACTGTTCTGCACTCCCATCCCTGGAGGGATGCCTGTGGTGCAGCAGTTTCACTTGCTTTGCCAAAACGATCTTTGGGCAACTTTCAA
ATAGCAGAAAGTGAGTGTTTCGTACAGTTGCTAGTTTATCACTCACGGAAGGGCAGGAAAGCTAAAGGGGGTCGCTATTTCACTGATTTTTCCA
AAACCGATTTTCTGGCTGCTTTCAACCAACACTGACTGCTTAATCAAGATTGGTTCCTTCAGTCACGTTACAGTAGGGAAGCGAAGGGGCGTTC
ACTATTTCTCTTACTTTCCAGAACCGAATTTCTGGCTGCTTTCAAATGGCATTAACTCCGTGTTTCACCGAGGTACAGGGGCTCACCAAACAAG
GGCAGGGAAAGAACGCCTCTTTCCGTGTTACCTACCGTGCAAAGCCGATTTCTGGGGCTACCTTCAAATAGCGCCAAGGGAGGGTTTAAATGAA
CAGCATCGTTCCCTCAGATAACAGAAGCGAAGCTACCGGTGGCTCTTTTTAACTCACTTTCCCAAACTGGTTTTCTGCCTAACGTGTTACAAAT
ACGACGACGTGAGTGATTTACCATGCTGCTCCTTTTATGCCAGCAACAGCAAGGAGGCATGTCATAGACATGCCAGCAGCATCCCCTGCCCTTA
TTTGAGTGGGAAACAGTACGCTGCCTAAACGCTCACTCAGTGCTCTCTGAAACTAGCCAGACATTCAGCTGGGGAAACTAAGTGAAACAGAGCC
ACCATTTGCGTCCCTGACGGTATTAGAGTGAGAAACAGACCGCAGCTTAAACACGCCCTTCGTGCCTTGCTTAAAGCAGCCCGAAAAATCGAAT
ATGAGGAGAAAGCGATACCGACCCCACAGGAAGCTTCCCTGCTCTCGGTGGCGTGAGTACCCGGGTGCTGCGAAAACACTCACGGTGCTATTTGA
AACTAGCCAGACATTCACTTTGGAAAAATAAGTGCGAGAGAAGCACCTTCTGATTCCCTACTGCTATTAGAGCGACGAAGAGTATGCCTGCTTA
AACACGCACACATCCTTCCTTTGGACAACAGCCTGAAGAACCGATTTGAAAGCTCTGTGCTATACACACGCCACCAACACCCCCTACTGCTGTGTG
ACTGAAAGAGAGAACGCTGCTTCAACACTTCTTTTTGTGCCTTTTGAAAGCAGCCTGAGAATCGGATATGAGAAGAAAGAGCTCTATACACACG
ATAGCTAGCCTGCTCTCATTAGACTGAAAAACCAGTGTGCTGCTGAAGCAGTCACTCAGCGCTATTTGAAACTAGCCTGATACTCAGTTGGGAA
GAATAAGTTGAAAGAACCACGATCTGCTGCCCGGCTGTTATCAGAGAGGAAACACAATACGCGGCTGCAACACTCTCTTAGTGTTCTTTGAACG
GAGCCTGACATTCGCTTACGAAACCTACGTGGGTGGAGTAGACACGTCAGCAGCATCCCTGCTCTTATTTGTGTGAAAAAAACAGTGCGCCGCT
TCAACACTCACTTAGTGGTCTTTGAAGCTAGTCAGAAATTCAGTTGGGAAAACGAAGTGAAGCGGAGCCACCATTTGCTTCCCGGATGTTGTTA
GAGTTCAACACAGTATGCTGTTTCAAGACTTACTTGGAGCTATTTGAAAGCAGTCTGGAAATCGGATTTGAATAATAAGTCATATGTACCCACG
TATAGCTTCACGTCTCTCATTGGAGTGAAAACCAGTGTGCTGCTTAAACATTCACTTAGTGCTATTTGAACCAAGCCGGAAATTCCGTTGGTTA
AAAAAAAAAAAAAAAATCAGT
>CVE116
GAATGCAGCAGTTTCTTCAGTTAAGGAAACTAAGGCAATTCCTCATTTAACTGGCTGCTAGGCTCTTTCTTAGCATTACATTCAATGGCAAAAT
CAGGTTTTACCTAATGCGCTGC
>CVE117
GAATGCAGCAGGATCATCAGTTTTCAAAGCCAAGCAAGTTACTTCTTTAATTGATGGCTATGCCCTTGTTTTCAGTTACAATCACTGTCAAATC
ACTGCTTTTACCTAACGCTCTGT
>CVE811
```

GCTATTACTTAATGCGCTCTGGATGCACCGCGAGCATCCGTTTCAGAAACCAAGCCATTTCCTCGTTTAATTCCTGGCAATGCCCTTCTGACAG
TTTCACTTAAGGCCAAATCTCTGCTCTTACCTAATATTGTGGAAAGGTCCCAAGATCATCAGTTTCACATAGCAAGCAAAGTATTAGTTTCACT
GGCGGCTAGGCCCTTGTTTTCAGTTACACTGAATGCCACCTCAATCAGCTGACGCTGTGTAAATGCACCAGGATGATCTGTTTTTCAAACCGAA
TAATATACTTGTCAAACTCATGGCTCTGCCCTTGTTTTCAGTTACCGTGAATGCCAAATCACTGCTCATGCTTAAGGCTCTGTGAATGCACCAA
GATGATCAGTTTTCCAAACTGAAGAATATGCTGGTTTCACTCATCTCTATGTCCTTGTTTTCAGTTAGAATGAATGACAAATCACTGGTTCTGC
CTCCCACTCTGGGAAATGACTGGGAACCTCAGGTTTCAAGACCAAGGCAGTTCCTTGTTTAAAAGGAGCTTCTGCCCTTCTTTTCCCTTACAAT
GCTTTTCAAATCCCTGCCTTTACCTAACACTGTAGAGCGCCCCAAGACCTTCAGCTTGGAAATCACAGCCCCTCACTTCTTTCATGGAAGGAAA
CACCTGTTTTCACTTATGATCACTGTTAAAGATTGGTTTCACCTAGTTCTCTGTAAGTGAACCCAGAACGCCTGTTTGGCAAGCTAGCAAACAC
CGTCTCTATCTGGCCGCTCTTCCCTGCTTTTCACTAGCAGCCAATGCCAAATCT
>CVE812
GCTATTAGATAACACTGTGCGAATGCACCAGGAGAATCAGTTTTACAAACTAAGCAATATACTTGTTAAACTCGTGGATATGCCCTTGTTTTCA
GTTATAATGAATGCCAAATCAATGCTAATAACTAACGGTCTGTGAATGCACCAAGATCATCCGTTTTGTAAACTAAGCAGTATACTGCTTTCAC
TCATCTCTATGGCCTTGTTTTCAGTGTCAATGAATGCGAAATCACTGTTTCTGCCTGCCATTGTGGGAATTCACTTGGAACCTCATTATTGAAA
GCAAGGCAAATACTGTGTGAAATGGTGGCTATGCCCTTGTTTTCCCTTAAAAGTCTTTTCAAATCCCTGCATTTCCCTACCACTCTGTCAACGC
ACCAAGATCTTCAGCTTTGAAATCCAAGCCGTTTACTGCTTCAGGGAAGAATTCCACCAGTTTTCACTTATAATCATTGCCTAAACAATGTTTT
CACTTAATGATCTGTAAGTGAACCAAGAACATCAGTTGTGGGAAGCTGGGGAATGACTTCTTTTTGGCTGCTGTGCTATGGTTTTCACTAACAT
CCAAAGCCACATCTCTGCTGCTACCTAATGCCCTGTGAATGCACCCTGATTATCAGTTACCCAAAACAAGTAATTTCCTAGTTTAACTGGTGCT
ATGCCCTTTTGTCAGTTGCAGTGAATGCCAAACCTGTGCTATTAACTATTGCTGGGTGAATGTACCAAGATCATCAGTTTTACATACGAAACAA
CTTATTTGTTTGCCTCGCGGCTATGAACAGGTTTACAGTTACAATGACTGCCACATGACT
>CVE815
GCTATTGCAAACGCTCTGTGAATGCACCAGGAACACCAGTTTTCCAAACTAAGGGAAATCCTTGTTTGGTTCATGGCTATGCCCTTGTTTTCAG
TTACCAGTAATGAAAATCAGTGCTCTTACATACCGGCCTGCGAATGCACCAAGAACTTCAGTTTTCCAAGCTAAGGAACATACTCCTTTAACTC
ATTCCTAGCTACTTGTTTTCAGTTAGAATGAATGCCAAATCATAGGTTCTGCTTACCGCTCTGGGAATTCAAAGGAAACCTCAGTTTTCAAATC
CAAGCCTGTTACTGGTTTAAACGGTGCCTCTGCCCTTCTTTTCCCTTACAGTTCTTTGCAAATTCCGGCCTTAAGCTCACACTGTCAGAACGCC
CCAGGACCTTCAGCTTGGAAATGTAAGCGCCTCACCTGCTTCATGCAGCCAATGAACCTGTTTTCACTTATAATCAATCTCTTAAGATTAGTTT
CACCTAATCTTCTGTCACTGAACCCAGAATATGTGTTTTTGACAGCAAGGAAATGATTTCCTTTTTCGGCCTCTCTGCCCTGGAATTCACTAAC
AAACAATGCCAAAACTCTGCAATTATTTAATGCGCTGTGAATGCACTCTAATCATCAGTTTCTCAAACCAAGCAATCTCCTTGTTTGATTCGTG
GCGATGCCCTTCTCTTTGTTAGAATGAATGCCAAATCTCTGCTCTTACCTAATGCTGTGTCAATGTGCCAAGATCATCAGATTTACAAGGCAAG
CAAGTTACTTGCTTGTCTCGCGCCTATGCCCTTGTTTTCAGTTACAGTGAATGAAAGCTCACC
>TMA117
GAATGCAGTAGTTTCTTCAGTTTTGAAAACTAAGCAAGTTACTTGTTTAATTGGCTGCTATGCACTTGTTTTCACTTACAATCAATGTCAAAAC
ACTGGTTTTACCTAATGCTCTGT
>TMA451
GAATGCAGGAGGTAGTTCAATTTCCAAAGTTTAATAATATACTTGTTTGTTTCCCTGCTATGCTCTTGCTTTCACTCACAATGTAAAAACCCGG
ATTTTACGAAATTCTATGCCAATGCAGTAACTTCCTCACGTTTCTAATGTAAGACTAGTGTTTTTTGTTGTTTTAAATCTGTGTGTTAACTTAA
AATCATCTTCTATACACTGGTTTAACATAACACGCTGTGAATTCAGGAGTAAATTCACCTTTCAAAGTTAAGTAACATACTTGTTTGTTTGGCT
GCTATGCAATCACTTTCACTGAGTATGCCAACATACTGCTTTTTCCTGATGTTTCCCCATAGCACTAGTTGTTTTACTCCACAAATGTGAATGT
GTGTTTAATTTGGTGCTTTGCTTCTGTTTTCATTTACAGTGACCTTCAAACCACTGCTTTCACGAAACATGCTGT

Supplementary File 5: Cluster graph layout of satDNAs from the SATCHO1 family identified by TAREAN.

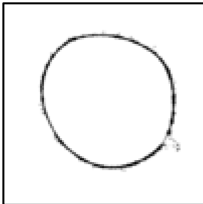
a) *Bradypus variegatus*

BVA116 BVA117

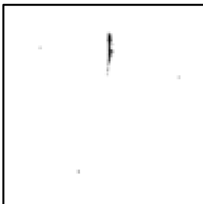
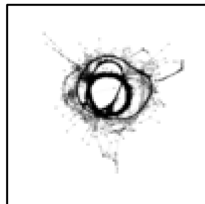
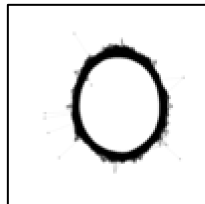
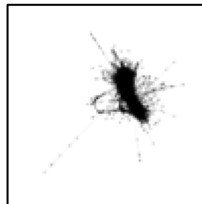
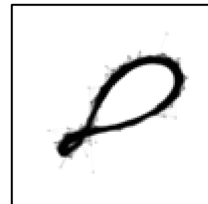
b)*Myrmecophaga tridactyla*

MTR3029



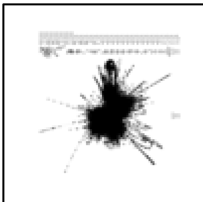
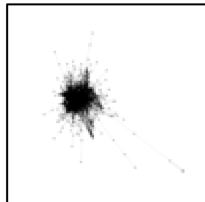
C)*Chaetophractus vellerosus*

CVE116 CVE117 CVE811 CVE812 CVE815

d)*Tolypeutes matacus*

TMA117 TMA451

5. CAPÍTULO 3

Identification and Characterization of the Most Abundant satDNAs in the Genomes of the Anteaters *Myrmecophaga tridactyla* and *Tamandua tetradactyla* (Myrmecophagidae, Xenarthra)

Radarane Santo Sena[1]; Gustavo Campos Silva Kuhn[1]; Marta Svartman[1]*
[1]Laboratório de Citogenômica Evolutiva, Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil

Running Title: Satellite DNAs in Anteater Species

*Corresponding author
E-mail: svartmanm@ufmg.br
Telephone number: 5531-34092612
Fax number: 5531-34092568

Abstract

Satellite DNAs (satDNA) represent a repetitive DNA class that maps to fundamental chromosome regions, as centromeres and telomeres. In this work, we analyzed the satDNA content in the genomes of the anteater species *Myrmecophaga tridactyla* and *Tamandua tetradactyla*. Using the TAREAN tool, we identified ten satDNA sequences, eight on *M. tridactyla* and two on *T. tetradactyla* genomes. In Blastn searches, we also identified the *M. tridactyla* satDNA MTR156 in *the T. tetradactyla* genome. The satDNA content contributes differently to the genome size of these two anteaters, comprising about ~1.05% and ~0.21 of the *M. tridactyla* and *T. tetradactyla* genomes, respectively. Based on copy number estimation, we chose the four most abundant satDNAs identified by TAREAN to map these sequences by fluorescent *in situ* hybridization (FISH) on the chromosomes of the two anteater species. The MTR156 sequence mapped in the telomeric region of all chromosomes of *M. tridactyla* and a telomeric probe mapped in the pericentromeric region of a few chromosomes from both species. The other satDNAs mapped in the constitutive heterochromatin located in the pericentromeric regions and in some interstitial areas of a few chromosomes.

Keywords: Satellite DNAs; Centromeres; TAREAN; Anteaters; FISH; Constitutive Heterochromatin.

Introduction

The ten recognized species of anteaters belong to the families Cyclopedidae, with seven species of pygmy anteaters, and Myrmecophagidae, with three species (https://xenarthrans.org/species/). Both families are united under the suborder Vermilingua, order Pilosa, Xenarthra (Superina and Loughry, 2015).

The Myrmecophagidae have a toothless mouth and a long tongue used to eat ants and termites, which compose their primary diet (https://xenarthrans.org/species/; Hayssen, 2014; Gaudin et al., 2018). They have a wide geographical distribution, spanning from Central to South America, but are under extinction risk, mainly due to anthropic activities that cause the loss of their habitats (https://xenarthrans.org/species/).

Despite its population decline, Myrmecophagidae species, especially the giant anteater *Myrmecophaga tridactyla*, are still poorly studied. Clozato et al. (2017) analyzed seven different Brazilian populations of *M. tridactyla*, from nine different states, and four different biomes (Pantanal, Cerrado, Amazon Forest and Atlantic Forest), and concluded that the Cerrado is an important area to preserve this species. Coimbra et al. (2022) investigated the Brazilian populations of the giant anteater regarding their distribution, ecology, phylogeography, and historical demography. They analyzed 66 samples used by Clozato et al. (2017) and added 28 new samples from museum specimens and road-kills, from the same four biomes studied by Clozato et al. (2017). They found two divergent clusters of mitochondrial DNAs, one related to the Amazon Forest and the other to three different biomes, Cerrado, Pantanal, and Atlantic Forest. These authors concluded that the Cerrado is a priority area for the species conservation because it presents better habitats and conditions to favor genetic variability by providing the possibility of gene flow between the two more distant populations in the Amazon and Atlantic forests (Coimbra et al., 2022). A populational study of the lesser anteater *Tamandua tetradactyla* mitogenome revealed high levels of mitochondrial variability (Ruiz-García et al., 2021).

Several Xenarthra species, including the Myrmecophagidae, present relatively large genomes (>4Gb) (https://www.genomesize.com/search.php; Redi and Capanna, 2012), which may be due to repetitive sequences. Repetitive DNA makes up a substantial portion of the eukaryotic genome (>50%) (Garrido-Ramos, 2017) and is composed of different families of dispersed or tandemly arrayed sequences. Dispersed repetitive sequences are divided into transposable elements (TEs), pseudogenes, segmental duplications, and some gene families.

The tandemly arrayed sequences are subdivided into ribosomal DNA, telomeric DNA, microsatellites, minisatellites, and satellite DNAs (satDNA). The TEs are the most abundant class of repetitive DNA in some eukaryotic genomes and are split into retrotransposons (Class I elements) and DNA transposons (Class II elements), a classification based on their origin and transposition mechanism (López-Flores and Garrido-Ramos, 2012).

Transposable elements are the most studied repetitive sequence of the Myrmecophagidae family. Nishihara et al. (2007) discovered a new SINE (short interspersed repetitive element) in the genomes of *M. tridactyla* and *T. tetradactyla*, which is a DNA transposon composed of a tRNA-derived region with a poly-A tail, named MyrSINE. Recently, Osmanski et al. (2023) made a curation of TEs from 248 mammalian genomes, including *M. tridactyla* and *T. tetradactyla,* and found that these two anteaters have newly inserted TEs, which mean they still present some transposition activity. *M. tridactyla* has more recent insertions of SINEs, whereas recent LINEs are more common in *T. tetradactyla*.

The recent advances in Whole Genome Sequencing (WGS) technologies allowed surveys like that of Osmanski et al. (2023) and provided the means to identify and characterize other significant classes of repetitive DNAs, the satDNAs. This class of tandemly arrayed sequences is located at fundamental chromosome regions, such as centromeres and telomeres, and at interstitial chromosome regions. SatDNAs typically present a high evolution rate due to the mechanism of Concerted Evolution, which allows the satDNAs to be used as species-specific markers in some species groups (Dover 1982). Another characteristic of advances in WGS is the creation of softwares and pipelines developed to identify and characterize repetitive DNA content. For instance, the pipeline RepeatExplorer2 and its tool TAREAN (Novák et al., 2017) have been extensively used to identify and characterize the satDNAs sequences on assembled genomes.

In this work we identified and characterized the most abundant satDNA sequences in the genomes of *M. tridactyla* and *T. tetradactyla*, including chromosome mapping of the satDNAs which had an estimate of more than 300 tandem copies.

Material and Methods

Identification and *in Silico* Characterization of SatDNAs

We performed a graph-based clustering analysis with raw reads on TAREAN to identify and characterize the most abundant satellite DNAs (satDNA) on the giant anteater *Myrmecophaga tridactyla* (MRT), and on the lesser tamandua *Tamandua tetradactyla* (TTE) genomes. We used the TAREAN to analyze paired-end Illumina reads from *M. tridactyla* (accession: SRX8010125) and *T. tetradactyla* (accession: SRX4562107) and performed the analysis with default options without a prior sample size for the runs. The software randomly chose and analyzed 5,693.729 reads of *M. tridactyla* (~13% of GP) and 4,938.038 reads of *T. tetradactyla* (~12% of GP).

After initial identification by TAREAN, we used the satDNA sequences as queries on searches on Repbase and Genbank to detect similarities among our sequences and previously described sequences present on these databases. We also searched the new satDNAs directly on *M. tridactyla* (accession: GCA_004026745.1) and *T. tetradactyla* (GCA_004025105.1) genomes available on GenBank using Blastn with default parameters. On Blast searches, we identified that MTR156, the most abundant satDNA of *M. tridactyla*, is also present in the *T. tetradactyla* genome. We then retrieved copies of the MTR156 satDNA from each genome to perform pairwise evolutionary distances on MEGAX. We build a phylogenetic tree with the MTR156 copies from both species using the Maximum Likelihood method with 1000 replicates.

Biological Samples

Tissue samples from a male *M. tridactyla* and a female *T. tetradactyla* were obtained from Fundação de Parques Municipais e Zoobotânica de Belo Horizonte/MG, Brazil (with the permit from Instituto Brasileiro do Meio Ambiente e dos Recursos Naturais Renováveis, Processo Sisbio 28422-5, conceded to Marta Svartman). These samples were used to establish fibroblast cultures from which we obtained chromosome preparations according to Stanyon and Galleni (1991). Total genomic DNAs were extracted with the Wizard Genomic Purification kit (Promega).

Molecular Analysis

To perform FISH experiments using the satDNA sequences as probes, we extracted total genomic DNAs from the *M. tridactyla* and *T. tetradactyla* sampled tissues and amplified the satDNA sequences by PCR on their DNAg. The PCR products were purified and sequenced. We did not manage to amplify the MTR156 sequence from *T. tetradactyla*, so we used the probe from *M. tridactyla* for FISH in *T. tetradactyla* chromosomes.

We used the following oligo sequences designed from the satDNAs consensus identified by TAREAN: MTR156F' (TGCACCCAGAAGAGTCACTC), MTR156R' (CAATTTGATGCAGGGTCGGT), MTR696F' (TCCAAGGGAGCGTTTCTCTC), MTR696R' (GAAACACTGACTTCGTGCTC), MTR1550F' (CGTACCCTAACCCGAACCCT), MTR1550R' (GGGATACGGTCGGTGCTAAG), MTR3029F' (TCACTTTCTTTGCGCAACCG), and MTR3029R' (AACAGAACCATCATTCGTCC) from *M. tridactyla*; TTE35F' (TATTCCATCTGGGGCAGTGG), TTE35R' (TACAGATGGAACCCTCCCCA), TTE745F' (CTTAAAGCACACCTACGAAC), TTE745R' (AGGTTAGGGTTATGGATAGG) from *T. tetradactyla*. The thermocycling conditions followed were: 95ºC-5min, 30 cycles of 95ºC-1min, the temperature of melting (TM) varied for each oligo pair 46ºC-1min (TTE745), 52ºC-1min (TTE35), 56ºC-1min (MTR156), and 58ºC (MTR696 and MTR3029), 72ºC-1min, and final elongation was 72ºC-10min. The PCR products were purified by Wizard SV Gel and PCR Clean-up System Kit (Promega) and sequenced with an ABI 3730 Platform (Applied Biosystems). The amplification for the TTE35 did not work, so we used the MTR35 product produced by PCR, ran an agarose gel (1%), cut the oligos dimers from the gel, purified, sequenced, and used the sample as a probe for FISH experiments.

Banding Patterns and Fluorescent in Situ Hybridization (FISH)

We performed GTG- and CBG-banding in the chromosome spreads of *M. tridactyla* and *T. tetradactyla*, according to Seabright (1971) and Sumner (1972), respectively.

FISH experiments were performed using as probes ~150ng of each sequenced PCR product, labeled with the Biotin-Nick Translation mix or Digoxigenin-Nick Translation mix (Roche Applied Science). We also used a commercial telomeric probe with a Biotin tag (5'-[Biotin-5] $TTAGGG_8$). The probes were denatured in 50% formamide/2xSSC at 98ºC for 10min. Chromosome spreads were dehydrated in an ethanol series (70%, 90%, and 100%, 3 min each), denatured at 70% formamide/2xSSC at 75ºC for 2min (MTR) and 1:50min (TTE).

The chromosome preparations were incubated with the probes overnight at 37ºC in a wet chamber. Post-hybridization washes consisted of two baths of 2xSSC at 42ºC for 5min each and the immunodetection was performed with anti-digoxigenin/rhodamine (Roche) and avidin/FITC (Invitrogen), followed by two baths of PBT for 5min each. The chromosomes were counterstained with DAPI/Slow Fade (Invitrogen). Analyses and image acquisition were performed under a Zeiss Axioimager2 epifluorescence microscope using the AxioVision software (Zeiss).

Preliminary Results

TAREAN Results and Phylogenetic Analysis

Table 1: Must abundant satellite DNA identified by TAREAN and their main features.

Species	SatDNAs Identification	Cluster	Consensus Length (bp)	Satellite Probability	Genome Prop. (%)	AT Content (%)	Estimated Copy Number
***Myrmecophaga tridactyla* (MTR)**	MTR696	26	696	0.994	0.230	48.99	14.511,22
	MTR1550	57	1550	0.981	0.025	53.03	708,26
	MTR3029	58	3029	0.981	0.025	52.53	362,43
	MTR156	8	156	0.0793	0.770	46.79	216.746,12
***Tamandua tetradactyla* (TTE)**	TTE745	46	745	0.0727	0.20	59.06	10.790,82
	TTE35	81	35	0.0154	0.01	42.86	11.484,51

TAREAN identified eight new satDNAs in the *M. tridactyla* genome, six with high confidence and two with low confidence and two new satDNAs in the *T. tetradactyla* genome, both with low confidence (Table 1). These satDNAs comprise about ~1.05% and ~0.21 of the *M. tridactyla* and *T. tetradactyla* genomes, respectively. Considering that these species have an estimated genome size of ~4Gb, the satDNA content does not appear to contribute much to their genome size. TAREAN identified eight new satDNAs in *M. tridactyla*, but we analyzed only four sequences because they have >300 estimated copies and presented hits on the assembled genome with more than three tandemly organized copies: the satDNAs MTR156, MTR696, MTR1550, and MTR3029. TAREAN did not identify the

MTR156 in the *T. tetradactyla* genome, but we found this sequence in Blastn searches on the assembled genome of this species forming long arrays (>32Kb).

Searches on Repbase and GenBank databases using the six selected identified satDNAs as queries revealed that MTR696 and TTE35 did not have significant sequence identity with any previously described DNA sequence. MTR156 showed sequence identity with a Penelope-like retroelement, an ancient eukaryotic class of repetitive DNA that shares a common ancestor with the telomerase reverse transcriptase. MTR156 is composed of ~156 bp long repeats and has a higher-order organization (HOR) comprised of twenty-six fragments of six bp each (~68% of pairwise identity). MTR156 also shares some pairwise identity with other MTR satDNAs (~51% with MTR696, ~57% with MTR1550, and ~64% with MTR3029) and with the vertebrate conserved telomeric sequence (TTAGGG)n (~71%). MTR1550 has similarity with small fragments of four different transposable elements, including the same Penelope-like retroelement present in the MTR156. MTR3029 also showed three small fragments of transposable elements, including the Penelope-like retroelement. MTR3029 also shares ~60% of pairwise identity with SATCHO1, a centromeric satDNA from the two-toed sloths *Choloepus didactylus* and *C. hoffmanni* (Sena et al., 2020). TTE745 showed similarity with a small fragment with the BMSAT1 satDNA from the blue whale *Balaenoptera musculus*, and a small fragment with a transposable element. TTE35 did not have similarities with any sequence present on RepBase or GenBank.

The MTR156 copies retrieved from the assembled genomes of *M. tridactyla* and *T. tetradactyla* presented ~0.4% nucleotide variability. The maximum likelihood tree with retrieved copies from these two species did not show clustering into species-specific branches (Fig. 1). Retrieved copies from the other satDNAs showed the following nucleotide variability values: ~0,12% in MTR696, ~0,21% in MTR1550, ~0,28% in TTE35, and ~0,22% in TTE745. We did not get copies of MTR3029 from the *M. tridactyla* assembled genome.

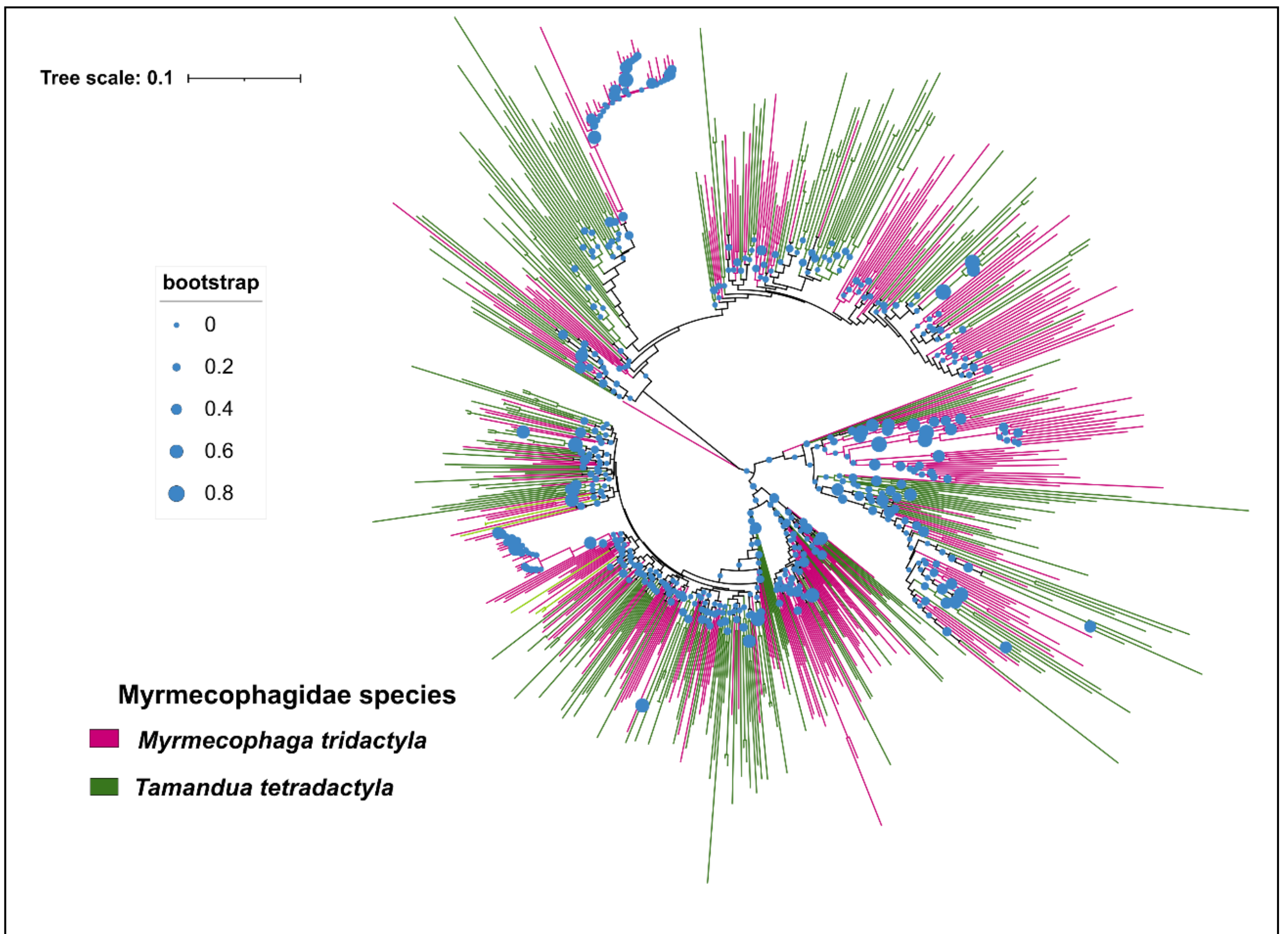


Figure 1: Maximum-Likelihood tree of MTR156 copies retrieved from assembled genomes of *Myrmecophaga tridactyla* and *Tamandua tetradactyla.*

SatDNAs Chromosome Mapping

The GTG-banding allowed the identification of all chromosome pairs of *M. tridactyla* and *T. tetradactyla*. Our sample of *M. tridactyla* is a male with 2n = 60 and FN = 104 (fundamental number), and the *T. tetradactyla* sample is a female with 2n = 54 and FN = 104 (Fig. 2 a, b). These karyotypes were earlier described for *M. tridactyla* (O'Brien et al., 2006) and *T. tetradactyla* (Svartman, 2006). The CBG-banding revealed the presence of constitutive heterochromatin in the pericentromeric region of all chromosomes in *M. tridactyla*, except for pairs 25, 26, and the sex chromosomes (Fig. 2c). Some chromosome pairs had interstitial bands in their short arms (pairs 6-11, 13, 19, 20, and 24), in the long arms (4 and 13), or in the subtelomeric region (2, 7, 8, 13, 14, 21, 24, 25, 27, 28, 29) (Fig. 2c). CBG banding in the chromosomes of *T. tetradactyla* revealed the presence of constitutive heterochromatin in the pericentromeric regions of 11 autosome pairs (1, 2, 4, 7-9, 11, 13, 17, 18, and 21), in the long

arms of four pairs (1, 2, 16, and 19), and on the subtelomeric region of three pairs (6, 7, and 14) (Fig. 2d).

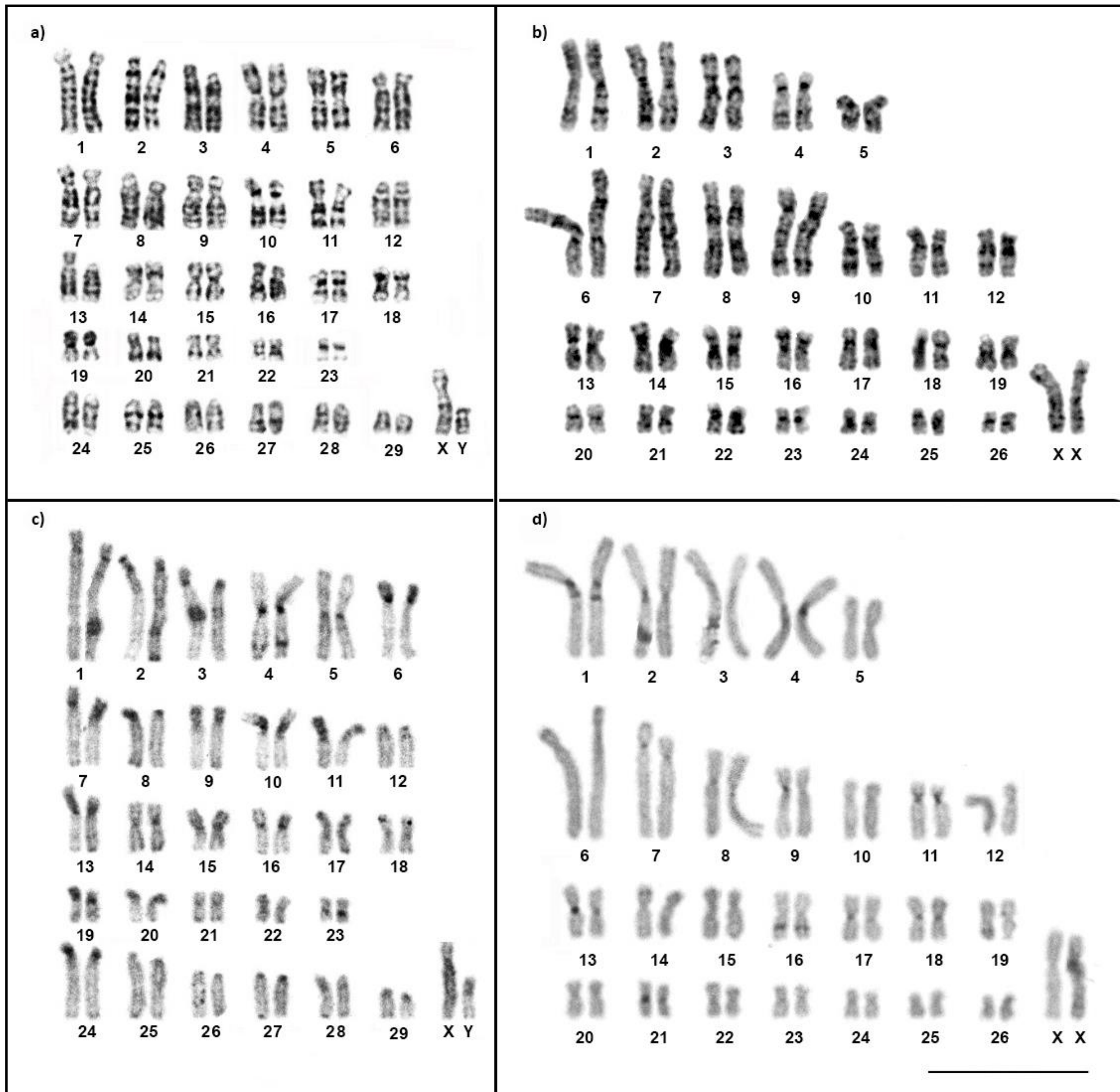


Figure 2*:* Karyotype of *M. tridactyla* (a-c) and *T. tetredactyla* (b-d) after: (a-b) GTG-banding and (c-d) CBG-banding. Bar: 10µm.

Chromosome mapping by FISH was performed for the most abundant satDNAs identified by TAREAN in the *M. tridactyla* (Fig. 3 a-b) and *T. tetradactyla* (Fig. 3 e-g) genomes. The most abundant satDNA MTR156 mapped to the subtelomeric regions of all *M. tridactyla* autosomes, in the pericentromeric regions of the meta/submetacentric pairs 3, 6, 7, 12, 13, 18, 21, and the acrocentric pairs 24-29. The MTR696 probe revealed signals in the

pericentromeric region of pairs 3, 6-8, 12, and 25-27. MTR1550 hybridized in the pericentromeric region of pair 27, whereas MTR3029 mapped to the pericentromeric region of pair 3. The telomeric probe (Tel) revealed signals in the telomeric regions of all chromosomes and additional pericentromeric signals in pairs 4, 7, 10, 13, 17, and 21.The satDNA TTE35 and TTE745 mapped in the pericentromeric regions of different *T. tetradactyla* chromosome pairs: TTE35 in pairs 4, 8, 12, and 16; and TTE745 in pairs 9, 13, 17, and 21. The telomeric probe hybridized to all the telomeric regions, to the pericentromeric regions of pairs 9, 13, 17, 18, and 21, and in an interstitial region of the long arm of pair 6. The MTR156 probe revealed signals in the pericentromeric regions of pairs 4, 13, 15-18, 21, in the subtelomeric regions of pairs 6 and 25, and in the same site where the telomeric probe hybridized in the long arm of pair 6. The pericentromeric and interstitial signals mapped in the chromosomes from the two anteater species coincided with regions of constitutive heterochromatin revealed after CBG-banding (Fig. 2 c-d).

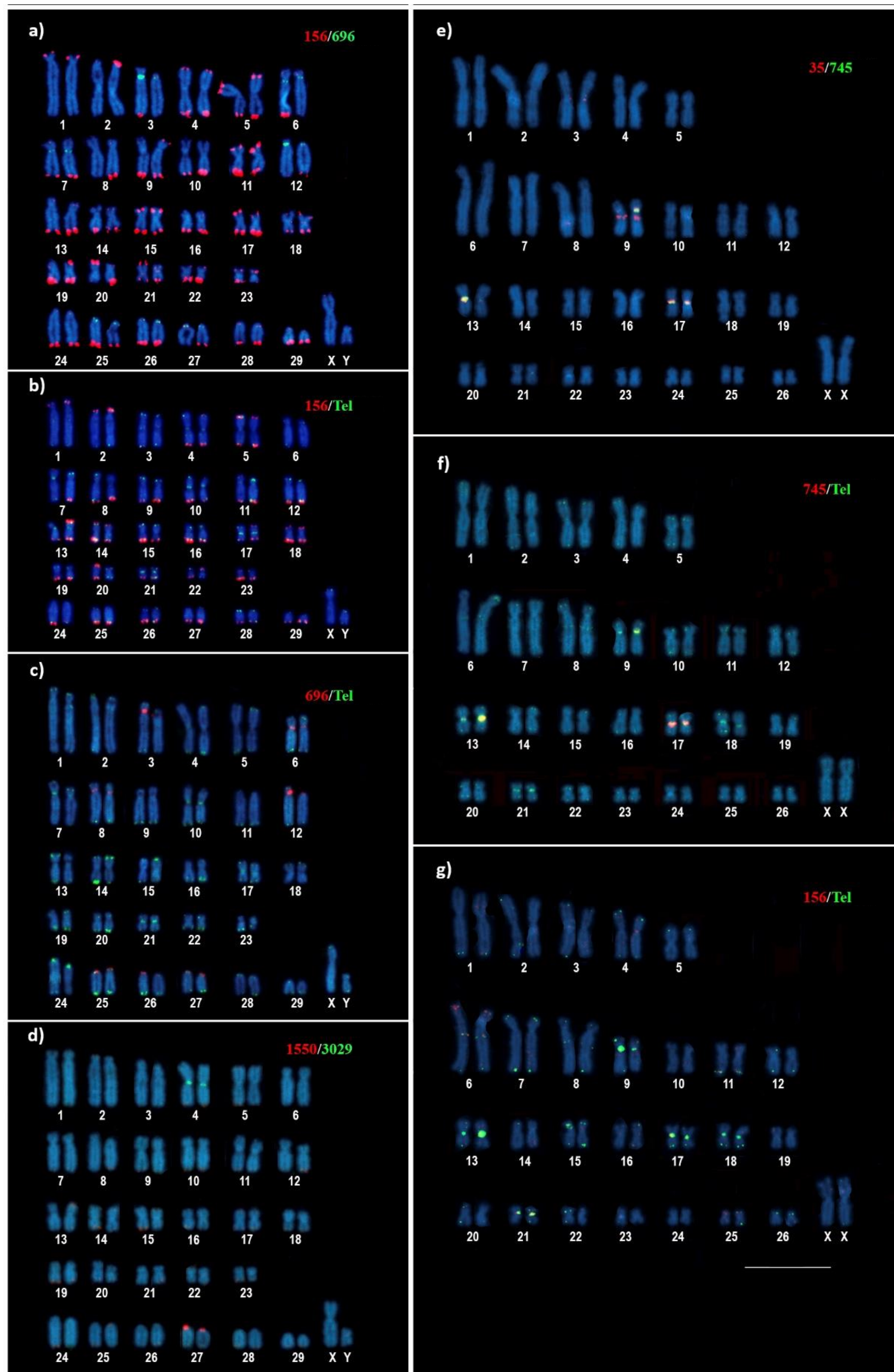


Figure 3: Karyotypes of *M. tridactyla* (a-d) and *T. tetradactyla* (e-g) after FISH experiments. Probes from satDNAs MTR156/MTR696 (a), MTR156/Tel (b), MTR696/Tel (c), MTR1550/MTR3029 (d), TTE35/745 (e), TTE745/Tel (f), and MTR156/Tel (g). The probes used are indicated in each karyotype. The probes in red were immunodetected with anti-digoxigenin/rhodamine and the probes in green with anti-biotin/FITC. Chromosomes were counterstained with DAPI. Bar = 10μm.

Perspectives

In order to conclude the study of satDNAs of *M. tridactyla* and *T. tetradactyla* we aim to perform more *in silico* and FISH analyses. We will analyze the available *T. tetradactyla* genome long reads and chromosomes assembled. This analysis may help to map the satDNAs on the chromosomes of *T. tetradactyla*, including the MTR156 and may help clarify if the MTR156 has a different sequence on the *T. tetradactyla* genome that we did not identify yet. We also intend to perform new FISH experiments using probes produced with the entire TTE35 sequence and a probe with a partial sequence of MTR156, a segment with a lower pairwise identity, together with the telomeric probe. Both probes will have a biotin or digoxigenin tag and we expect to get more precise FISH results with them. We will also perform a fiber-FISH assay with the new MTR156 and telomeric probes to verify how these repetitive sequences are organized in the telomeric regions. These future experiments shall contribute to better characterize the satDNAs of these two anteater species.

References

Clozato, C. L., Miranda, F. R., Lara-Ruiz, P., Collevatti, R. G., & Santos, F. R. (2017). Population structure and genetic diversity of the giant anteater (*Myrmecophaga tridactyla*: Myrmecophagidae, Pilosa) in Brazil. *Genetics and Molecular Biology*, *40*(1), 50–60. https://doi.org/10.1590/1678-4685-gmb-2016-0104

Coimbra, R. T. F., Magalhães, R. F., Lemes, P., Miranda, F. R., & Santos, F. R. (2022). Integrative Phylogeography Reveals Conservation Priorities for the Giant Anteater *Myrmecophaga tridactyla* in Brazil. *Diversity*, *14*(7), 542. https://doi.org/10.3390/d14070542

Dover, G. Molecular drive: a cohesive mode of species evolution. (1982). *Nature*, *299(5879)*, 111-117.

Garrido-Ramos, M. (2017). Satellite DNA: An Evolving Topic. *Genes*, *8*(9), 230. https://doi.org/10.3390/genes8090230

Gaudin, T. J., Hicks, P., & Di Blanco, Y. (2018). *Myrmecophaga tridactyla* (Pilosa: Myrmecophagidae). *Mammalian Species*, *50*(956), 1–13. https://doi.org/10.1093/mspecies/sey001

Hayssen, V. (2011). *Tamandua tetradactyla* (Pilosa: Myrmecophagidae). *Mammalian Species*, *43*, 64–74. https://doi.org/10.1644/875.1

López-Flores, I., & Garrido-Ramos, M. A. (2012). The Repetitive DNA Content of Eukaryotic Genomes. Em M. A. Garrido-Ramos (Org.), *Genome Dynamics* (Vol. 7, p. 1–28). S. Karger AG. https://doi.org/10.1159/000337118

Nishihara, H., Kuno, S., Nikaido, M., & Okada, N. (2007). MyrSINEs: A novel SINE family in the anteater genomes. *Gene*, *400*(1–2), 98–103. https://doi.org/10.1016/j.gene.2007.06.003

Novák, P., Ávila Robledillo, L., Koblížková, A., Vrbová, I., Neumann, P., & Macas, J. (2017). TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic acids research, 45(12)*, e111-e111.

Osmanski, A. B., Paulat, N. S., Korstian, J., Grimshaw, J. R., Halsey, M., Sullivan, K. A. M., Moreno-Santillán, D. D., Crookshanks, C., Roberts, J., Garcia, C., Johnson, M. G., Densmore, L. D., Stevens, R. D., Zoonomia Consortium†, Rosen, J., Storer, J. M., Hubley, R., Smit, A. F. A., Dávalos, L. M., … Zhang, X. (2023). Insights into mammalian TE diversity through the curation of 248 genome assemblies. *Science*, *380*(6643), eabn1430. https://doi.org/10.1126/science.abn1430

Redi, C. A., & Capanna, E. (2012). Genome Size Evolution: Sizing Mammalian Genomes. *Cytogenetic and Genome Research*, *137*(2–4), 97–112. https://doi.org/10.1159/000338820

Rossi, L. F., Chirino, M. G., Luaces, J. P., & Merani, M. S. (2014). Determination of the karyotype of the giant anteater *Myrmecophaga tridactyla* (Myrmecophagidae: Xenarthra) using classical and molecular cytogenetic techniques. *Italian Journal of Zoology*, *81*(3), 322–327. https://doi.org/10.1080/11250003.2014.943308

Ruiz-García, M., et al., (2021). Comparative mitogenome phylogeography of two anteater genera (*Tamandua* and *Myrmecophaga*; Myrmecophagidae, Xenarthra): Evidence of

discrepant evolutionary traits. *Zoological Research*, *42*(5), 525–547. https://doi.org/10.24272/j.issn.2095-8137.2020.365

Seabright, M. (1971). A rapid banding technique for human chromosomes. *The Lancet, 298(7731)*, 971-972.

Stanyon, R., & Galleni, L. (1991). A rapid fibroblast culture technique for high resolution karyotypes. *Bolletino Di Zoologia*, *58*(1), 81–83. https://doi.org/10.1080/11250009109355732

Sumner, A. T. (1972). A simple technique for demonstrating centromeric heterochromatin. *Experimental Cell Research*, *75*(1), 304–306. https://doi.org/10.1016/0014-4827(72)90558-7

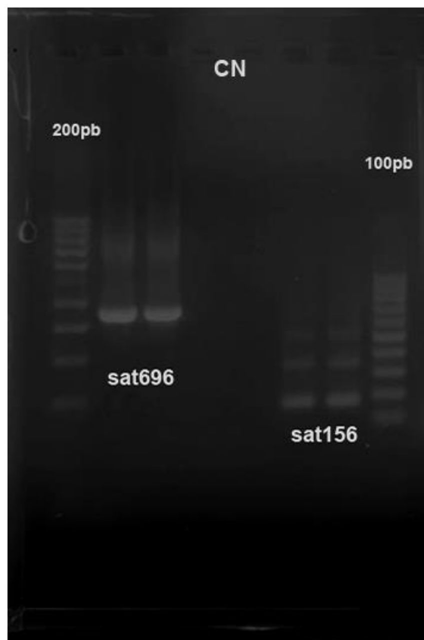
Superina, M., & Loughry, W. J. (2015). Why do Xenarthrans matter?: Table 1. *Journal of Mammalogy*, *96*(4), 617–621. https://doi.org/10.1093/jmammal/gyv099

Svartman, M., Stone, G., & Stanyon, R. (2006). The Ancestral Eutherian Karyotype Is Present in Xenarthra. *PLoS Genetics*, *2*(7), e109. https://doi.org/10.1371/journal.pgen.0020109

Supplementary Files

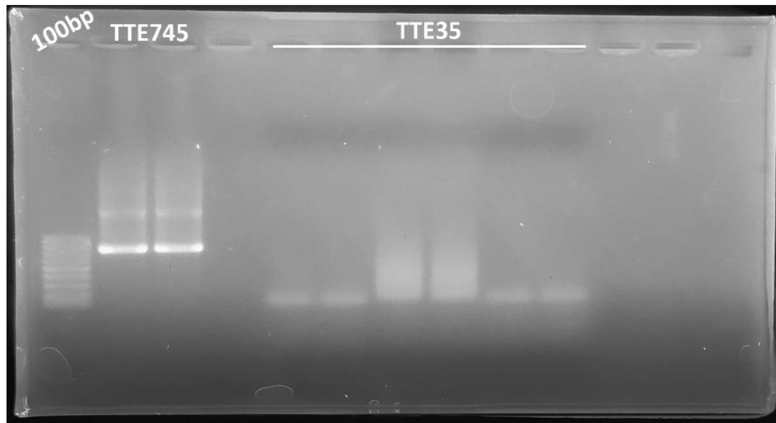
Supplementary File 1: TAREAN results with all satDNAs identified.

Species	Cluster	Monomer length (bp)	Satellite Confidence	Satellite Probability	Genome Proportion %	C-Value	P-Value	AT Content%	Genome Size	Estimated Copy Number
***Myrmecophaga tridactyla* (MTR)**	26	696	HC	0.994	0.230	0.994	0.966	48.99	4391220000	14.511,22
	29	8286	HC	0.975	0.220	0.984	0.951	48.26		1.165,90
	56	2430	HC	0.993	0.025	0.993	0.965	56.13		451,77
	57	1550	HC	0.981	0.025	0.997	0.963	53.03		708,26
	58	3029	HC	0.981	0.025	0.983	0.960	52.53		362,43
	73	1830	HC	0.989	0.013	0.994	0.946	23.44		311,94
	8	156	LC	0.0793	0.770	0.807	0.931	46.79		216.746,12
	63	4268	LC	0.0666	0.019	0.869	0.831	50.28		195,49
***Tamandua tetradactyla* (TTE)**	49	16278	HC	0.994	0.14	0.991	0.971	60.93	4019580000	345,71
	46	745	LC	0.0727	0.20	0.856	0.868	59.06		10.790,82
	81	35	LC	0.0154	0.01	0.710	0.842	42.86		11.484,51



Supplementary File 2: Agarose gel of purified PCR products from satDNAs MR156 and MTR696 amplified from *Myrmecophaga tridactyla* genomic DNA.

Supplementary file 3: Agarose gel of purified PCR products from satDNAs MTR1550 and MTR3029 amplified from *Myrmecophaga tridactyla* genomic DNA.



Supplementary File 4: Agarose gel of purified PCR products from satDNAs TTE745 and dimers of the TTE35 primers after PCR with *Tamandua tetradactyla* genomic DNA.

6. CONCLUSÕES

Com o intuito de investigar a relação dos DNAs satélites (satDNA) com a evolução genômica e cariotípica da superordem Xenarthra, realizamos a identificação e caracterização de 41 sequências de satDNA presentes nos genomas de oito espécies de Xenarthra (*Bradypus variegatus*, *Choloepus didactylus*, *C. hoffmanni*, *Chaetophractus vellerosus*, *Dasypus novemcinctus*, *Myrmecophaga tridactyla*, *Tamandua tetradactyla* e *Tolypeutes matacus*). Essa identificação, bem como a caracterização dessas sequências, permitiu constatarmos que essa classe de DNA repetitivo não tem uma contribuição expressiva para o aumento do tamanho dos genomas das espécies analisadas. Estudamos comparativamente as sequências identificadas, realizando, quando possível, o seu mapeamento cromossômico por hibridação *in situ* fluorescente (FISH), e comparando com os padrões de bandeamento GTG e CBG. Assim, identificamos uma família de satDNA (SATCHO1) com ampla distribuição, presente em pelo menos cinco das oito espécies analisadas e outra família (SATCHO2) presente em três espécies estudadas. Os representantes dessas duas famílias apresentam características que sugerem alguma função centromérica como: presença de sítios de ligação da proteína centromérica B (CENP-B box), pares de pequenos dímeros e a formação de possíveis estruturas secundárias estáveis. Além disso, em experimentos de FISH usando os satDNAs SATCHO1 e SATCHO2 como sondas, constatamos que as duas sequências ocupam regiões pericentroméricas nos cromossomos de *C. hoffmanni*. Este achado reforça o possível papel dessas sequências no correto funcionamento do cinetócoro durante as divisões celulares. Estudos citogenômicos para mapeamento cromossômico desses satDNAs são necessários para melhor elucidarmos as relações entre essas sequências e os genomas dessas e de outras espécies de Xenarthra.

O satDNA MTR156, presente nos genomas de *M. tridactyla* e *T. tetradactyla*, mostrou ter relação com sequências teloméricas, tanto em análises *in silico*, como em experimentos de FISH. Os resultados de FISH coincidiram com os resultados de bandeamento CBG que evidenciaram heterocromatina constitutiva nas regiões centroméricas e em algumas regiões cromossômicas intersticiais. O aprofundamento dos estudos desta sequência faz-se necessário para elucidar sua relação com sequências teloméricas.

7. REFERÊNCIAS

Abba, A. M., Cassini, G. H., Túnez, J. I., & Vizcaíno, S. F. (2018). The enigma of the Yepes' armadillo: *Dasypus mazzai*, *D. novemcinctus* or *D. yepesi*? *Revista Del Museo Argentino de Ciencias Naturales, Nueva Serie*, *20*(1), 83–90. https://doi.org/10.22179/REVMACN.20.578

Alkan, C., Cardone, M. F., Catacchio, C. R., Antonacci, F., O'Brien, S. J., Ryder, O. A., Purgato, S., Zoli, M., Della Valle, G., Eichler, E. E., & Ventura, M. (2011). Genome-wide characterization of centromeric satellites from multiple mammalian genomes. *Genome Research*, *21*(1), 137–145. https://doi.org/10.1101/gr.111278.110

Anderson, R. P., & Handley, C. J. (2001). A new species of three-toed sloth (Mammalia: Xenarthra) from Panamá, with a review of the genus *Bradypus*. *Proceedings of the Biological Society of Washington*, *114*(1), 1–33.

Anderson, R. P., & Handley, C. O. (2002). Dwarfism in insular sloths: Biogeography, selection, and evolutionary rate. *Evolution*, *56*(5), 1045–1058. https://doi.org/10.1111/j.0014-3820.2002.tb01415.x

Azevedo, N. F., Svartman, M., Manchester, A., De Moraes-Barros, N., Stanyon, R., & Vianna-Morgante, A. M. (2012). Chromosome painting in three-toed sloths: A cytogenetic signature and ancestral karyotype for Xenarthra. *BMC Evolutionary Biology*, *12*(1), 36. https://doi.org/10.1186/1471-2148-12-36

Bao, W. & Jurka, J. (2010). Origin and evolution of LINE-1 derived " half-L1 " retrotransposons (HAL1). *Gene*, *465*(1–2), 9–16. https://doi.org/10.1016/j.gene.2010.06.005

Benirschke, K. (2006). Xenarthra. In S. O'Brien, J. Menninger, & W. Nash (Eds.), *Atlas of Mammalian Chromosomes, 93*, 81–93.

Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Ostell, J., Pruitt, K. D., & Sayers, E. W. (2018). GenBank. *Nucleic Acids Research*, *46*(D1), D41–D47. https://doi.org/10.1093/nar/gkx1094

Biscotti, M. A., Canapa, A., Forconi, M., Olmo, E., & Barucca, M. (2015a). Transcription of tandemly repetitive DNA: functional roles. *Chromosome Research*, *23*(3), 463–477. https://doi.org/10.1007/s10577-015-9494-4

Biscotti, M. A., Olmo, E., & Heslop-Harrison, J. S. (Pat.) (2015b). Repetitive DNA in eukaryotic genomes. *Chromosome Research*, *23*(3), 415–420. https://doi.org/10.1007/s10577-015-9499-z

Charlesworth, B., Sniegowski, P., & Stephan, W. (1994b). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature*, *371*(6494), 215–220. https://doi.org/10.1038/371215a0

Churakov, G., Smit, A. F. A., Brosius, J., & Schmitz, J. (2005). A novel abundant family of retroposed elements (DAS-SINEs) in the nine-banded armadillo (*Dasypus novemcinctus*). *Molecular Biology and Evolution*, *22*(4), 886–893. https://doi.org/10.1093/molbev/msi071

Corin-Frederick, J. (1969). Les formules gonosomiques dites aberrantes chez les Mammifères

Euthériens. *Chromosoma*, *27*, 268–287.

Da Silva, M. J., Fogarin Destro, R., Gazoni, T., Narimatsu, H., Pereira Dos Santos, P. S., Haddad, C. F. B., & Parise-Maltempi, P. P. (2020). Great Abundance of Satellite DNA in *Proceratophrys* (Anura, Odontophrynidae) Revealed by Genome Sequencing. *Cytogenetic and Genome Research*, *160*(3), 141–147. https://doi.org/10.1159/000506531

Defler, T. (2019). The Xenarthrans: Armadillos, Glyptodonts, Anteaters, and Sloths. In N. H. Landman & P. J. Harries (Eds). *History of terrestrial mammals in South America*. (pp. 117-138). Cham, Switzerland: Springer. http://link.springer.com/10.1007/978-3-319-98449-0

Delsuc, F., Scally, M., Madsen, O., Stanhope, M. J., De Jong, W. W., Catzeflis, F. M., Springer, M. S., & Douzery, E. J. P. (2002). Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. *Molecular Biology and Evolution*, *19*(10), 1656–1671. https://doi.org/10.1093/oxfordjournals.molbev.a003989

Delsuc, F., Vizcaíno, S. F., & Douzery, E. J. P. (2004). Influence of Tertiary paleoenvironmental changes on the diversification of South American mammals : a relaxed molecular clock study within xenarthrans. *BMC Evolutionary Biology*, *4*(1), 1–13. http://www.biomedcentral.com/1471-2148/4/11

Delsuc, F., Superina, M., Tilak, M. K., Douzery, E. J. P. & Hassanin, A. (2012). Molecular phylogenetics unveils the ancient evolutionary origins of the enigmatic fairy armadillos. *Molecular Phylogenetics and Evolution, 62*(2),673-680.

Dobigny, G., Yang, F., O'Brien, P. C. M., Volobouev, V., Kovács, A., Pieczarka, J. C., Ferguson-Smith, M. A., & Robinson, T. J. (2005). Low rate of genomic repatterning in Xenarthra inferred from chromosome painting data. *Chromosome Research*, *13*(7), 651–663. https://doi.org/10.1007/s10577-005-1002-9

Dover, G. (1982). Molecular drive: A cohesive mode of species evolution. *Nature*, *299*(5879), 111–117. https://doi.org/10.1038/299111a0

Dover, G. A. (1986). Molecular drive in multigene families: How biological novelties arise, spread and are assimilated. *Trends in Genetics*, *2*(C), 159–165. https://doi.org/10.1016/0168-9525(86)90211-8

Dumbovic, G., Forcales, S. V., & Perucho, M. (2017). Emerging roles of macrosatellite repeats in genome organization and disease development. *Epigenetics*, *12*(7), 515–526. https://doi.org/10.1080/15592294.2017.1318235

Feijó, A., Vilela, J. F., Cheng, J., Schetino, M. A. A., Coimbra, R. T., Bonvicino, C. R., ... & Cordeiro-Estrela, P. (2019). Phylogeny and molecular species delimitation of long-nosed armadillos (*Dasypus*: Cingulata) supports morphology-based taxonomy. *Zoological Journal of the Linnean Society, 186*(3), 813-825.

Finnegan, D. J. (1989). Eukaryotic transposable elements and genome evolution. *Trends in genetics, 5*, 103-107.

Gardner, A. L. (2008). *Mammals of South America, Volume 1- Marsupials, Xenarthrans, Shrews, and Bats* (A. L. Gardner (ed.)). University of Chicago Press.

Garrido-Ramos, M. A. (2017). Satellite DNA: An evolving topic. *Genes*, *8*(9), 230.

https://doi.org/10.3390/genes8090230

Gaudin, T. J., Hicks, P., & Di Blanco, Y. (2018). *Myrmecophaga tridactyla* (Pilosa: Myrmecophagidae). *Mammalian Species*, *50*(956), 1–13. https://doi.org/10.1093/mspecies/sey001

Goldschmidt, B., & De Almeida, J. (1993). Cytogenetic studies in sloths (*Bradypus variegatus*). *Revista Brasileira de Genética*, *16*(4), 939–948.

Jorge, W. (1981). *Estudo cromossômico de algumas espécies da ordem Edentata*. UNESP.

Jorge, W. (1982). Chromosome study in the "Pichiciego Argentino," *Chlamyphorus truncatus* Harlan (Dasypodidae, Edentata). 23: 101 – 7. *Mammalian Chromosomes Newsletter*, 101–107.

Jorge, W., Dennis, A. M. J., & Benirschke, K. (1978). Chromosome Estudies in Edentata. *Cytobios*, 157–172.

Jorge, W., Orsi-Souza, A. T., & Best, R. (1985). The somatic chromosomes of Xenarthra. In G.Gene Montgomery (Ed.), *The evolution and ecology of armadillos, sloths, and vermilinguas* (pp. 121–129). Smithsonian Institution.

Jorge, W., & Pinder, L. (1990). Chromosome study on the maned sloth *Bradypus torquatus* (Bradypodidae, Xenarthra). *Cytobios*, *62*(248), 21–25.

Liu, Y., Ye, J., Fu, B., Ng, B. L., Wang, J., Su, W., Yang, F., & Nie, W. (2011). Molecular cytogenetic characterization of the genome organization of the 6-banded armadillo (*Euphractus sexcinctus*). *Cytogenetic and Genome Research*, *132*(1–2), 31–40. https://doi.org/10.1159/000318706

McKenna, M. C., & Bell, S. K. (1997). *Classification of mammals: above the species level.* Columbia University Press.

Miranda, F. R., Casali, D. M., Perini, F. A., Machado, F. A., & Santos, F. R. (2018). Taxonomic review of the genus *Cyclopes Gray*, 1821 (Xenarthra: Pilosa), with the revalidation and description of new species. *Zoological Journal of the Linnean Society*, *183*(3), 687–721. https://doi.org/10.1093/zoolinnean/zlx079

Miranda, F. R., Chiarello, A., Rohe, F., & Mourão, G. (2015). *Instituto Chico Mendes de Conservação da Biodiversidade - Avaliação do Risco de Extinção dos Xenartros Brasileiros* (ICMbio - Instituto Chico Mendes de Conservação da Biodiversidade (ed.)).

Mora, P., Vela, J., Ruiz-Ruano, F. J., Ruiz-Mena, A., Montiel, E. E., Palomeque, T., & Lorite, P. (2020). Satellitome analysis in the ladybird beetle *Hippodamia variegata* (Coleoptera, coccinellidae). *Genes*, *11*(7), 1–13. https://doi.org/10.3390/genes11070783

Moraes-barros, N., & Arteaga, M. C. (2015). Genetic diversity in Xenarthra and its relevance to patterns of neotropical biodiversity. *Journal of Mammalogy*, *96*(4), 690–702. https://doi.org/10.1093/jmammal/gyv077

Nishihara, H., Kuno, S., Nikaido, M., & Okada, N. (2007). MyrSINEs: A novel SINE family in the anteater genomes. *Gene*, *400*(1–2), 98–103. https://doi.org/10.1016/j.gene.2007.06.003

Novák, P., Neumann, P., Pech, J., Steinhaisl, J., & MacAs, J. (2013). RepeatExplorer: A Galaxy-based web server for genome-wide characterization of eukaryotic repetitive

elements from next-generation sequence reads. *Bioinformatics*, *29*(6), 792–793. https://doi.org/10.1093/bioinformatics/btt054

Novák, P., Robledillo, L. Á., Koblížková, A., Vrbová, I., Neumann, P., & Macas, J. (2017). TAREAN: A computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Research*, *45*(12). https://doi.org/10.1093/nar/gkx257

O'Leary, M. A, Bet al. (2013). The Placental Mammal Ancestor and the Post–K-Pg Radiation of Placentals. *Science*, *339*(6120), 662–667. https://doi.org/10.1126/science.1229237

Ohno, S. (1972). So much'junk'DNA in our genome." In *Evolution of Genetic Systems, Brookhaven Symp. Biol*. (pp. 366–370).

Palazzo, A. F., & Gregory, T. R. (2014). The Case for Junk DNA. *PLoS Genetics*, *10*(5). https://doi.org/10.1371/journal.pgen.1004351

Pereira, H. R. J., Jorge, W., & da Costa, M. E. L. T. (2004). Chromosome study of Anteaters (Myrmecophagideae, Xenarthra) - A preliminary report. *Genetics and Molecular Biology*, *27*(3), 391–394. https://doi.org/10.1590/s1415-47572004000300014

Pita, S., Panzera, F., Mora, P., Vela, J., Cuadrado, Á., Sánchez, A., ... & Lorite, P. (2017). Comparative repeatome analysis on *Triatoma infestans* Andean and Non-Andean lineages, main vector of Chagas disease. *PLoS One*, *12(7)*, e0181635.

Rossi, L. F., Chirino, M. G., Luaces, J. P., & Merani, M. S. (2014). Determination of the karyotype of the giant anteater *Myrmecophaga tridactyla* (Myrmecophagidae: Xenarthra) using classical and molecular cytogenetic techniques. *Italian Journal of Zoology*, *81*(3), 322–327. https://doi.org/10.1080/11250003.2014.943308

Ruiz-Ruano, F. J., López-León, M. D., Cabrero, J., & Camacho, J. P. M. (2016). High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Scientific Reports*, *6*(1), 1–14. https://doi.org/10.1038/srep28333

Silva, D. M. Z. A., et al. (2017). High-throughput analysis unveils a highly shared satellite DNA library among three species of fish genus *Astyanax*. *Scientific Reports*, *7*(1), 1–12. https://doi.org/10.1038/s41598-017-12939-7

Sonta, S. I., Hayata, I., Sasaki, M., & Kondo, N. (1980). Karyotype and sex determining in the two-toed sloth, *Choloepus didactylus*. *Chromosome Information Service*, *28*, 15–17.

Steiner, C. C., Houck, M. L., & Ryder, O. A. (2011). Species identification and chromosome variation of captive two-toed sloths. *Zoo Biology*, *30*(6), 623–635. https://doi.org/10.1002/zoo.20360

Superina, M., & Loughry, W. J. (2015). Why do xenarthrans matter? *Journal of Mammalogy*, *96*(4), 617–621. https://doi.org/10.1093/jmammal/gyv099

Svartman, M. (2013). Reconstruction of the ancestral Xenarthran karyotype. *eLS*. https://doi.org/10.1002/9780470015902.a0024972

Svartman, M., Stone, G., & Stanyon, R. (2006). The ancestral Eutherian karyotype is present in Xenarthra. *PLoS Genetics*, *2*(7), 1006–1011. https://doi.org/10.1371/journal.pgen.0020109

Tautz, D. (1993). Notes on the definition and nomenclature of tandemly repetitive DNA

sequences. *DNA Fingerprinting: State of the Science*, 21–28. https://doi.org/10.1007/978-3-0348-8583-6_2

Thakur, J., Packiaraj, J., & Henikoff, S. (2021). Sequence, chromatin and evolution of satellite DNA. *International Journal of Molecular Sciences*, *22*(9). https://doi.org/10.3390/ijms22094309

Valeri, M. P., Dias, G. B., do Espírito Santo, A. A., Moreira, C. N., Yonenaga-Yassuda, Y., Sommer, I. B., Kuhn, G. C. S., & Svartman, M. (2021). First Description of a Satellite DNA in Manatees' Centromeric Regions. *Frontiers in Genetics*, *12*. https://doi.org/10.3389/fgene.2021.694866

Vizcaíno, S. F., & Loughry, W. J. (2008). The Biology of the Xenarthra. In S. F. V. and W. J. Loughry (Ed.), *University Press of Florida. 7(1)*. University Press of Florida. https://doi.org/10.1896/1413-4411.7.1.59

Waters, P. D., Dobigny, G., Pardini, A. T., & Robinson, T. J. (2004). LINE-1 distribution in Afrotheria and Xenarthra: Implications for understanding the evolution of LINE-1 in eutherian genomes. *Chromosoma*, *113*(3), 137–144. https://doi.org/10.1007/s00412-004-0301-9

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., Flavell, A., Leroy, P., Morgante, M., Panaud, O., Paux, E., SanMiguel, P., & Schulman, A. H. (2007). Aunified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, *8*, 973–982. https://doi.org/10.1038/nrg2165

Yang, F., et al. (2006). Comparative genome maps of the pangolin, hedgehog, sloth, anteater and human revealed by cross-species chromosome painting: Further insight into the ancestral karyotype and genome evolution of eutherian mammals. *Chromosome Research*, *14*(3), 283–296. https://doi.org/10.1007/s10577-006-1045-6

Yu, H. T., et al. (2012). Use of a cytogenetic whole-genome comparison to resolve phylogenetic relationships among three species: Implications for mammalian systematics and conservation biology. *Theriogenology*, *77*(8), 1615–1623. https://doi.org/10.1016/j.theriogenology.2011.12.006