UNIVERSIDADE FEDERAL DE MINAS GERAIS

Instituto de Ciências Biológicas

Programa de Pós-Graduação em Genética

Mirela Pelizaro Valeri

IDENTIFICAÇÃO *DE NOVO* E CARACTERIZAÇÃO DE DNAS SATÉLITES EM MACACOS-DE-CHEIRO (*Saimiri*, Cebidae, Platyrrhini) E PEIXES-BOI (*Trichechus*, Trichechidae, Sirenia)

Belo Horizonte

2022

Mirela Pelizaro Valeri

IDENTIFICAÇÃO *DE NOVO* E CARACTERIZAÇÃO DE DNAS SATÉLITES EM MACACOS-DE-CHEIRO (*Saimiri*, Cebidae, Platyrrhini) E PEIXES-BOI (*Trichechus*, Trichechidae, Sirenia)

Tese

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

Belo Horizonte

2022

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

ATA DE DEFESA DE DISSERTAÇÃO / TESE

ATA DA DEFESA DE TESE	154/2022
	Entrada 2º/2017
Mirela Pelizaro Valeri	**CPF: 412.114.798-70**

Às quatorze horas do dia **17 de março de 2022**, reuniu-se remotamente a Comissão Examinadora de Tese, indicada pelo Colegiado do Programa, para julgar, em exame final, o trabalho intitulado: **"Identificação de novo e caracterização de DNAs satélites em macacos-de-cheiro (Saimiri, Cebidae, Platyrrhini) e peixes- boi (Trichechus, Trichechidae, Sirenia)"**, requisito para obtenção do grau de Doutora em **Genética.** Abrindo a sessão, a Presidente da Comissão, **Marta Svartman**, após dar a conhecer aos presentes o teor das Normas Regulamentares do Trabalho Final, passou a palavra à candidata, para apresentação de seu trabalho. Seguiu-se a arguição pelos Examinadores, com a respectiva defesa da candidata. Logo após, a Comissão se reuniu, sem a presença da candidata e do público, para julgamento e expedição de resultado final. Foram atribuídas as seguintes indicações:

Prof./Pesq.	Instituição	CPF	Indicação
Marta Svartman	UFMG	101.787.258-97	Aprovada
Gustavo Campos e Silva Kuhn	UFMG	260.136.648-62	Aprovada
Francisco Pereira Lobo	UFMG	012.273.736-94	Aprovada
Fernando Araujo Perini	UFMG	034.512.456-12	Aprovada
Diogo Cavalcanti Cabral de Mello	UNESP	059.618.894-32	Aprovada

Edivaldo Herculano Correa de Oliveira	Instituto Evandro Chagas	318.888.012-04	Aprovada

Pelas indicações, a candidata foi considerada: **Aprovada**

O resultado final foi comunicado publicamente à candidata pela Presidente da Comissão. Nada mais havendo a tratar, a Presidente encerrou a reunião e lavrou a presente ATA, que será assinada por todos os membros participantes da Comissão Examinadora.

Belo Horizonte, 17 de março de 2022.

Marta Svartman

Gustavo Campos e Silva Kuhn

Francisco Pereira Lobo

Fernando Araujo Perini

Diogo Cavalcanti Cabral de Mello

Edivaldo Herculano Correa de Oliveira

Assinatura dos membros da banca examinadora:

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 **1313573** e o código CRC **DE320D8D**.

Referência: Processo nº 23072.214252/2022-90   SEI nº 1313573

AGRADECIMENTOS

Às secretárias do programa de Pós Graduação em Genética, Raíssa Campos e Daniela Pereira Rezende por todo o auxílio e risadas.

À querida Selminha, responsável por um café incrível e por tantas comidas deliciosas, mas principalmente por não nos deixar com fome no meio dos experimentos. Não creio que seja possível fazer pesquisa sem um bom café (acompanhado pão de queijo e/ou bolo).

Aos meus maravilhosos pais Sérgio e Marli e irmão Guilherme, pelo amor sempre presente e pelo suporte emocional. Sem vocês eu não teria chegado tão longe.

Ao meu amor Artur Jordão, por todo o incentivo e por fazer parte da minha vida. Você deixa tudo mais leve. Não posso deixar de fora sua mãe Ivanilde e irmã Manuela, pois quem tem amizades assim tem tudo.

Ao programa de Pós-Graduação em Genética da Universidade Federal de Minas Gerais.

RESUMO

As sequências de DNAs repetitivos estão presentes no genoma da maioria dos eucariotos. Parte da porção repetitiva dos genomas é composta por DNAs satélites (satDNAs), longas cadeias de sequências repetidas em tandem que, mesmo não codificando proteínas, apresentam papel biológico importante na manutenção e regulação do genoma. Apesar dessas características, são parte pouco explorada do genoma da maioria dos mamíferos. No capítulo 1, caracterizamos dois satDNAs presentes em quatro espécies de primatas neotropicais do gênero *Saimiri*. Identificamos os satDNAs alfa e CapA a partir da análise do genoma sequenciado de *S. boliviensis* usando o software RepeatExplorer. O satDNA centromérico alfa possui ~340 pb e alta homogeneidade interespecífica entre as sequências. Já o CapA apresenta monômeros com cerca de 1.500 pb e está presente principalmente nas regiões subteloméricas dos cromossomos submetacêntricos e em algumas regiões intersticiais, mas com localização variável entre as espécies de *Saimiri*. A localização do CapA, associada a outras informações, como o número fundamental de braços autossômicos, pode ajudar na classificação das espécies presentes no gênero, que sempre foi tema de debate tanto sobre o número de espécies quanto a suas relações filogenéticas. No capítulo 2, usando como ponto de partida o genoma sequenciado do peixe-boi *Trichechus manatus* e o software TAREAN, identificamos e caracterizamos o TMAsat, o primeiro satDNA descrito no grupo e presente nas duas famílias de Sirenia. TMAsat possui cerca de 687 pb e localização centromérica em *T. manatus* e *T. inunguis*. Também detectamos o satDNA em *D. dugon* e *H. gigas* (espécie já extinta). Uma análise comparativa dos monômeros de TMAsat nessas quatro espécies não indicou sequências espécie-específicas, apesar dos ~43 Ma de divergência entre Trichechidae e Dugongidae, contrariando a previsão da Evolução Combinada. Detectamos o que parece ser a sequência ancestral do TMAsat em elefantes e hiraxes, sem o padrão típico de satDNA nesses genomas, indicando que a sequência do TMAsat se expandiu como um satDNA em Sirenia há menos de 69 Ma.

Palavras-chave: Repetições em tandem. RepeatExplorer. TAREAN. Primatas neotropicais. Peixe-boi amazônico. Peixe-boi marinho.

ABSTRACT

Repetitive DNA sequences are present in most eukaryotic genomes. A fraction of the repetitive portion is composed of satellite DNA (satDNAs), long arrays of tandemly repeated sequences that even without encoding proteins display important biological roles in genome maintenance and regulation. In spite of this, they are still an underexplored fraction of most mammalian genome. In chapter 1, we characterized two satDNAs present in four neotropical primate species from the *Saimiri* genus. We identified the alfa and CapA satDNAs from the analysis of the *S. boliviensis* sequenced genome using the RepeatExplorer software. The alfa centromeric satDNA has ~340 bp and high interspecific sequence homogeneity. CapA, on the other hand, has monomers of around 1,500 bp and is mainly present in the subtelomeric regions of the submetacentric chromosomes and in some interstitial regions, but with variable location among *Saimiri* species. The CapA location, associated with other information, such as the fundamental autosomal number, may help in the classification of species present of the genus, whose number and phylogenetic relationships have always been a matter of debate. In chapter 2, using the manatee *Trichechus manatus* sequenced genome and the TAREAN software, we identified and characterized TMAsat, the first satDNA described in the group and present in the two Sirenia families. TMAsat is about 687 bp and has a centromeric location in *T. manatus and T. inunguis*. We also detected this satDNA in *D. dugon* and *H. gigas* (an extinct species). A comparative analysis of TMAsat monomers in these four species did not indicate species-specific sequences despite the ~43 Ma divergence between Trichechidae and Dugongidae, in disagreement with the prediction of Combined Evolution. We detected the ancestral TMAsatlike sequence in elephants and hiraxes, without the typical satDNA pattern in these genomes, indicating that the TMAsat sequence expanded as a satDNA in Sirenia less than 69 Ma.

Keywords: Tandem repeats. RepeatExplorer. TAREAN. New world monkeys. Amazonian manatee. West Indian manatee.

LISTA DE FIGURAS

Capítulo 1

Capítulo 2

LISTA DE TABELAS

LISTA DE ABREVIATURAS

2n - número diplóide

BLAST - *basic local alignment search tool*

Bp - *base pairs*

DAPI - 4′,6-Diamidine-2′-phenylindole dihydrochloride

DDU - *Dugong dugon*

ECD - *evolutionarily conserved domain*

ENCs - *Evolutionary New Centromeres*

ESU - Unidades Evolutivas Significativas

FISH - Hibridação *in situ* Fluorescente, *Fluorescent in situ Hibridization*

FN - *fundamental number*

HGI – *Hydrodamalis gigas*

HOR - *higher-order repeat*

IUCN - *International Union for Conservation of Nature*

LINEs - *long interspersed nuclear elements*

LTRs - *long terminal repeats*

Ma - Milhões de anos

MCS - *most common sequence*

Mya - *Million years ago*

NCBI - *National Center for Biotechnology Information*

NF - número fundamental

NMDS - *Non-metric multidimensional scaling*

NOS1AP - *nitric oxid synthase 1 adaptor protein*

NWMs - *New World monkeys*

Pb - pares de base

PCR - Reação em cadeia da polimerase, *polymerase chain reaction*

SatDNAs - DNAs satélites, *satellite DNAs*

SBO - *Saimiri boliviensis*

SINEs - *short interspersed nuclear elements*

SSC - *Saimiri sciureus*

SUS - *Saimiri ustus*

SVA - *Saimiri vanzolinii*

TIN - *Trichechus inunguis*

TMA - *Trichechus manatus*

SUMÀRIO

1. INTRODUÇÃO

1.1 DNAs satélites

As sequências de DNA repetitivo correspondem a grande parte dos genomas de eucariotos, podendo representar mais da metade do conteúdo de DNA nuclear. Os principais tipos de DNAs repetitivos são as repetições dispersas e as repetições em tandem. Os elementos transponíveis fazem parte das repetições dispersas e ocupam grande parte do genoma dos eucariotos. Eles são divididos em duas classe: a classe I ou retrotransposons, que são amplificados e inseridos em outro local do genoma através de um RNA intermediário, e que inclui LTRs (*long terminal repeats*), os LINEs (*long interspersed nuclear elements*) e SINEs (*short interspersed nuclear elements*); e a classe II, composta pelos transposons de DNA, que não usam um RNA intermediário para se moverem dentro do genoma (Wicker *et al*., 2007; Biscotti *et al*., 2015).

Os DNAs satélites (satDNAs) são longas cadeias de repetições em tandem, presentes no genoma de quase todos os eucariotos estudados e são o principal componente da heterocromatina constitutiva (Plohl *et al*., 2012; Biscotti *et al*., 2015). A sequência do satDNA, sua abundância dentro do genoma, o tamanho médio do monômero e a localização cromossômica são extremamente variáveis entre as espécies. Tipicamente, um mesmo genoma pode conter vários satDNAs diferentes. Apesar de não codificarem proteínas, os satDNAs estão envolvidos em funções biológicas importantes, como a formação e manutenção da heterocromatina, inclusive na região dos centrômeros e telômeros, influenciando a integridade dos cromossomos e a estabilidade do genoma, a modulação da cromatina, a regulação gênica e o desenvolvimento (Shapiro e von Sternberg, 2005; Biscotti *et al*., 2015; Feliciello *et al*., 2021).

O termo DNA satélite se originou das bandas satélites formadas quando o DNA genômico era submetido a gradientes de densidade por centrifugação, como os gradientes de cloreto de césio. As repetições em tandem ficavam separadas por diferença de densidade em relação ao restante do DNA genômico (revisado em Tautz, 1993). Dentre as repetições em tandem, estão os micro e minissatélites, consideradas repetições moderadas no genoma. Os microssatélites são espalhados pela eucromatina e possuem monômeros com até 10 pares de bases (pb) e cadeias de até 100 pb. Já os minissatélites possuem unidades de repetição entre 10 e 100 pb, cadeias entre 0,5-30 kb e localização em várias regiões da eucromatina (Tautz, 1993; Charlesworth *et al*., 1994). Os satDNAs se diferenciam dos micro e minissatélites por suas longas cadeias, com centenas ou milhares de kilobases, e pela localização preferencial na

heterocromatina, como nos centrômeros, telômeros e blocos intersticiais (Figura 1a; Tautz, 1993; Biscotti *et al.*, 2015).
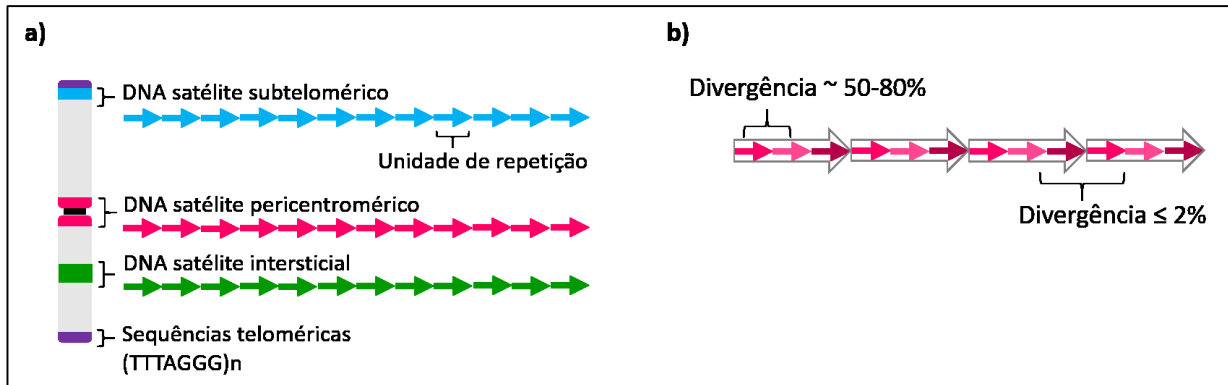
Figura 1. a) Localização preferencial dos satDNAs nas regiões heterocromáticas. b) Organização de satDNAs em *higher-order repeat* (HOR).

Além das longas cadeias com os monômeros repetidos em tandem, os satDNAs podem estar organizados em *higher-order repeat* (HOR), quando dois ou mais monômeros divergentes entre si formam juntos uma unidade de repetição, mas, quando comparadas, essas unidades apresentam alta identidade entre si (Figura 1b; Plohl *et al.*, 2012; Hartley e O'Neill, 2019; Thakur *et al.*, 2021). Essas unidades de repetição compostas são arranjadas em tandem formando longas cadeias, característica principal dos satDNAs. Esse tipo de organização pode variar dentro de um genoma, apresentando HOR cromossomo-específica, em que o número de monômeros, suas sequências e ordem podem variar dentro da unidade de repetição presente em determinado cromossomo, como é observado no satDNA centromérico alfa em humanos (Willard, 1985; Alkan *et al.*, 2007; Sullivan *et al.*, 2017).

DNAs satélites variam muito entre as espécies quanto ao tamanho da unidade de repetição, número de cópias e distribuição cromossômica (Plohl *et al.*, 2012). Os principais modelos evolutivos dos satDNAs são Evolução Combinada, Hipótese Biblioteca e Impulso Centromérico. A Evolução Combinada promove a homogeneização das cadeias de repetições dentro de um genoma, que são fixadas numa população, resultando em maior similaridade intra- do que interespecífica (Dover, 1982). A recombinação desigual, conversão gênica, transposição e amplificação por círculo rolante estão dentre os mecanismos moleculares responsáveis pela homogeneização das sequências dentro do genoma (Dover, 1982, 1986; Cohen e Segal, 2009). Assim, os satDNAs apresentam diferenças significativas mesmo entre espécies filogeneticamente próximas, podendo ser usados como marcadores taxonômicos e filogenéticos (Plohl *et al.*, 2012).

Já na Hipótese Biblioteca, espécies relacionadas compartilham um conjunto de satDNAs que estava presente no ancestral comum, a partir do qual cada satDNA pode se expandir ou retrair dentro do genoma de uma espécie, variando assim o perfil de abundância dos satDNAs em cada espécie (Fry e Salser, 1977; Meštrović *et al.*, 1998). Caso um determinado satDNA se expanda no genoma de uma espécie, será observado com baixo número de cópias em uma espécie relacionada. O outro modelo de evolução dos satDNA é o Impulso Centromérico, o qual prevê que sequências de satDNAs centroméricos, que facilitam a ligação dos centrômeros às fibras de fuso, segregam preferencialmente para o polo do óvulo durante a gametogênese, garantindo sua transmissão. Isso porque durante a meiose nas fêmeas apenas um dos quatro produtos da divisão celular se tornará o óvulo e os outros três serão glóbulos polares, sendo possível que sequências que ofereçam vantagem na ligação das fibras de fuso aumentem sua frequência na população (Thakur *et al.*, 2021).

Comparações interespecíficas entre as sequências de satDNAs, como sua localização cromossômica, abundância no genoma, relações com outras sequências e características da sua organização no genoma podem contribuir tanto para o entendimento de sua origem e evolução, quanto da evolução cromossômica e genômica das espécies analisadas (Kuhn *et al.*, 2008; Biscotti *et al.*, 2015). A disponibilidade de genomas sequenciados está aumentando expressivamente, mas análises de quantificação e caracterização das sequências repetitivas não têm sido feitas na mesma velocidade. Um dos motivos é a dificuldade na etapa de montagem dos genomas quando as *reads* resultantes do sequenciamento são curtas (<300 pb) e geralmente menores que a unidade de repetição. No caso dos DNAs satélites, a homogeneidade dos monômeros dentro do genoma é um desafio para a montagem dessas sequências em cadeias longas.

Um software muito útil para a identificação *de novo* e caracterização de sequências repetitivas é o RepeatExplorer2 (Novák *et al.*, 2013, 2017, 2020). Suas ferramentas podem ser executadas on-line na plataforma Galaxy, um gerenciador de *workflow* científico (Goecks *et al.*, 2010). O software funciona muito bem com sequências curtas, ideal para dados gerados por sequenciamento massivo, com *reads* de tamanho médio entre 100-300 pb. O RepeatExplorer2 funciona pelo agrupamento baseado em grafo, onde as *reads* amostradas aleatoriamente são todas comparadas entre si e agrupadas de acordo com sua similaridade. O grafo é formado por vértices e arestas, onde *reads* representam os vértices e a similaridade entre elas as arestas. O formato dos grafos ilustra a organização genômica da repetição e a variabilidade de suas sequências, podendo apresentar desde formas lineares típicas de elementos transponíveis como formatos circulares/globulares típicos de repetições em tandem. Devido a essa

comparação de todas as *reads* entre si, ocorre a identificação *de novo* de sequências repetitivas, sem a necessidade de uma base de dados com sequências conhecidas como referência. Após serem gerados os grupos, as *reads* que os compõem são montadas em *contigs* que serão usados para busca de similaridade com sequências repetitivas anotadas em bases de dados.

O TAREAN é uma ferramenta adicional do RepeatExplorer2 dedicada à identificação específica de satDNAs, que examina o formato dos grafos em busca de formas circulares ou globulares. A partir dos agrupamentos dos possíveis satDNAs, o TAREAN reconstrói o monômero da repetição em tandem baseado nos k-mers mais frequentes do *cluster* levando em consideração a orientação das *reads* que os compõe. O RepeatExplorer2 e o TAREAN têm sido usados para estudar DNAs repetitivos nos genomas de diversos organismos, como plantas (Kirov *et al*., 2017; Pamponét *et al*., 2019), insetos (Palacios-Gimenez *et al*., 2018, 2020; Silva *et al*., 2019), peixes e lagostins (Utsunomia *et al*., 2019; Boštjančić *et al*., 2020), e mamíferos (Pagán *et al*., 2012; Araujo *et al*., 2017; Sena *et al*., 2020).

1.2 *Saimiri*

Os primatas pertencentes à parvodem Platyrrhini são endêmicos do continente americano com distribuição desde o sul do México, passando pela América Central, até o norte da Argentina na América do Sul (Martin, 2012). Foram descritas aproximadamente 140 espécies de Platyrrhini, pertencentes a 20 gêneros, agrupados em três famílias: Cebidae, Atelidae e Pitheciidae. (Perelman *et al*., 2011; Schneider e Sampaio, 2015; Byrne *et al*., 2016; Alfaro, 2018). O número de espécies e gêneros de Platyrrhini reconhecidos e suas relações filogenéticas ainda são discutidos, com espécies novas sendo descritas e alguns grupos passando por constantes revisões taxonômicas. Isso é resultado do processo de intensa radiação adaptativa que originou um grupo muito diversificado em relação à morfologia, comportamento e cariótipos.

Dentro da família Cebidae, *Saimiri* é um gênero que possui taxonomia e filogenia confusas e está sob constante revisão, com suas espécies sendo conhecidas como macacos-de-cheiro. Elas estão distribuídas ao longo da Bacia Amazônica (Brasil, Colômbia, Venezuela, Guiana, Suriname e Guiana Francesa) e em um pequeno trecho na América Central, no Panamá e Costa Rica (Hershkovitz, 1984; Lavergne *et al*., 2010). A classificação das espécies dentro do gênero mudaram tanto que o número de taxa reconhecidos variou de um a 16 ao longo do tempo (Hershkovitz, 1984; Thorington, 1985; Costello *et al*., 1993; Boinski e Cropp, 1999; Cropp e Boinski, 2000; Groves C *et al*., 2005; Lavergne *et al*., 2010; Rylands e Mittermeier, 2013;

Alfaro *et al*., 2015). A classificação mais atual (Alfaro *et al*., 2015), e que consta na lista vermelha da International Union for Conservation of Nature (IUCN), considera 11 subespécies divididas em sete espécies (Alfaro *et al*., 2015; Alves *et al*., 2021; Heymann *et al*., 2021; Lynch *et al*., 2021; Paim *et al*., 2021; Silva Júnior *et al*., 2021; Solano-Rojas, 2021). Essas são: *S. boliviensis* (*S. b. boliviensis*, *S. b. peruviensis*), *S. sciureus*, *S. collinsi*, *S. vanzolinii*, *S. ustus*, *S. oerstedii* (*S. o. citrinellus*, *S. o. oerstedii*), *S. cassiquiarensis* (*S. c. albigena*, *S. c. macrodon* e *S. c. cassiquiarensis*). Segundo a IUCN, as populações de todas as espécies do gênero *Saimiri* estão diminuindo, exceto as de *S. cassiquiarensis*, para a qual não há dados suficientes. Já as categorias de risco são: menor preocupação para *S. boliviensis*, *S. sciureus*, *S. collinsi* e *S. cassiquiarensis*; quase ameaçada para *S. ustus*, e ameaçada/em perigo para *S. oerstedii* e *S. vanzolinii*.

Todos os indivíduos de *Saimiri* analisados apresentaram número diploide 2n = 44, com as diferenças entre cariótipos relacionadas ao número de braços cromossômicos ou número fundamental (NF), que varia de 74 a 78. Alguns autores inclusive associaram o NF à taxonomia e origem geográfica dos indivíduos. (Bender e Mettler, 1958; Jones *et al*., 1973; Jones e Ma, 1975; Ma e Jones, 1975; Hershkovitz, 1984; Yonenaga Yassuda e Chu, 1985; Moore *et al*., 1990; Assis *et al*., 1998; Stanyon *et al*., 2000; Chiatante *et al*., 2017). Apenas os pares 5 e 15 foram detectados como responsáveis pela variação do NF de acordo com sua morfologia, podendo ser acrocêntricos ou submetacêntricos. São três tipos de cariótipos já registrados: pares 5 e 15 acrocêntricos (NF = 74), par 5 submetacêntrico e par 15 acrocêntrico (NF = 76), ou os dois pares submetacêntricos (NF = 78) (Ma e Jones, 1975; Hershkovitz, 1984; Moore *et al*., 1990; Chiatante *et al*., 2017). Foi demonstrado que a alteração da morfologia desses dois pares é devida ao reposicionamento centromérico (Chiatante *et al*., 2017) e não a inversões pericêntricas ou translocações, como era suposto (Ma e Jones, 1975; Lau e Arrighi, 1976). O reposicionamento centromérico é o surgimento de um centrômero em uma nova região cromossômica, seguido da inativação do centrômero antigo. Reposicionamento centromérico também foi detectado no cromossomo X de *Saimiri*, que possui morfologia diferente dos de outros primatas (Rocchi *et al*., 2012).

O bandeamento CBG nos cromossomos de *Saimiri* evidenciou heterocromatina constitutiva nas regiões pericentroméricas de todos os cromossomos e regiões subteloméricas da maioria dos cromossomos submetacêntricos, além de segmentos intersticiais em alguns cromossomos (Jones e Ma, 1975; Lau e Arrighi, 1976; Yonenaga Yassuda e Chu, 1985; Chiatante *et al*., 2017). Um bloco adicional de heterocromatina foi observado no braço curto do cromossomo 14 de alguns indivíduos, com consequente aumento do tamanho do cromossomo (Jones e Ma, 1975;

Ma e Jones, 1975; Moore *et al*., 1990; Chiatante *et al*., 2017). Esse polimorfismo foi observado em heterozigose na maioria dos casos e mais frequentemente em indivíduos oriundos da Bolívia (Moore *et al*., 1990). Essa variante com mais heterocromatina foi descrita como isoforma B, enquanto o cromossomo 14 mais comum é chamada de isoforma A (Chiatante *et al*., 2017). Esses autores também verificaram que houve uma inversão paracêntrica na isoforma B e hipotetizaram que sequências teloméricas podem ter sido translocadas para a região intersticial do braço curto deste cromossomo.

O satDNA mais estudado até o momento nos primatas é o satDNA centromérico alfa, presente nos cromossomos dos Catarrhini (primatas do Velho Mundo, os grandes macacos e humanos) e Platyrrhini (Fanning *et al*., 1993; Alkan *et al*., 2007). Esse satDNA apresenta divergência entre espécies e até entre cromossomos da mesma espécie, incluindo organização em HOR (Alves *et al*., 1998; Alkan *et al*., 2007; Cellamare *et al*., 2009; Sujiwattanarat *et al*., 2015). Os monômeros do satDNA alfa em Catarrhini possui aproximadamente 170 pb e na maioria dos Platyrrhini, cerca de 340 pb, uma derivação da unidade de 170 pb, provavelmente resultado de um evento que aconteceu no ancestral comum dos Platyrrhini, com duas unidades de 170 pb um pouco divergentes entre si, que se comportam como uma única unidade de repetição com o dobro do tamanho original (Fanning *et al*., 1993; Cellamare *et al*., 2009). Em duas espécies da família Pitheciidae, *Chiropotes satanas* e *Pithecia irrorata,* o monômero do alfa possui cerca de 540 pb, composto por quatro subunidades de 170 pb, sendo a terceira incompleta (Alves *et al*., 1994, 1998). Em Callithrichine foram descritos os satDNAs MarmoSAT e CarB, exclusivos do grupo (Alves *et al*., 1995; Araujo *et al*., 2017). MarmoSAT está presente nos gêneros *Callithrix*, *Mico*, *Callimico* e *Cebuella*, possui cerca de 171 pb e localização subtelomérica e, apesar de possuir um monômero com tamanho semelhante ao do alfa, não tem nenhuma relação com o satDNA centromérico (Araujo *et al*., 2017). Já CarB está presente em *Callithrix* e *Mico*, com monômeros de aproximadamente 1.528 pb e foi localizado em blocos subteloméricos de heterocromatina, chegando a ocupar quase metade de alguns cromossomos (Alves *et al*., 1995).

Um satDNA presente em todas as famílias de Platyrrhini é o CapA (Fanning *et al*., 1993; Valeri *et al*., 2018). Esse satDNA foi primeiro descrito por Fanning *et al*. (1993) em *Sapajus apella* (na época identificado como *Cebus apella*), com monômeros de aproximadamente 1.500 pb e ocupando cerca de 5% do genoma. CapA foi mapeado nos blocos de heterocromatina constitutiva dos pares 4, 6, 11, 12, 13, 17 e 21 e na região telomérica de alguns cromossomos de *S. apela*. Usando experimentos de *Southern blot*, os autores sugeriram que este satDNA poderia estar presente em outros gêneros de Platyrrhini. CapA está presente apenas em

Platyrrhini e se originou a partir de uma sequência intrônica do gene *nitric oxide synthase 1 adaptor protein*. Este satDNA está presente em representantes de Cebidae, Atelidae e Pitheciidae, excetuando os membros de Callithrichine e do gênero *Callicebus*. Sua abundância varia de menos de 1% até 5% dos genomas analisados, com localização cromossômica também variável, mas em todos os casos associada à heterocromatina (Valeri *et al*., 2018).

1.3 Sirenia

A ordem Sirenia está inserida em Paenungulata, juntamente com Proboscidea (elefantes) e Hyracoidea (híraxes, Upham *et al*., 2019). Sirenia é representada pelos únicos mamíferos herbívoros totalmente aquáticos, os peixes-boi e dugongos, que vivem entre os trópicos em águas oceânicas costeiras rasas, rios e estuários (Bertram e Bertram, 1973; Marsh *et al*., 1986; Domning, 2018). São quatro espécies viventes divididas nas famílias Dugongidae e Trichechidae. Em Dugongidae há apenas a espécie *Dugong dugon,* já que a vaca marinha de Steller, *Hydrodamalis gigas*, foi extinta ao redor de 1768 (Anderson e Domning, 2009). *D. dugon* ocorre na costa da Austrália e na costa leste do continente africano. Trichechidae agrupa três espécies de peixes-boi do gênero *Trichechus*, duas espécies marinhas, *T. senegalensis* e *T. manatus*, e uma espécie de água doce, *T. inunguis. T. senegalensis* habita a costa oeste da África, do Senegal a Angola. *T. manatus* está distribuída ao longo da costa do oceano Atlântico, da Flórida até o nordeste brasileiro (Barros *et al*., 2017), e *T. inunguis* ocorre ao longo da Bacia Amazônica. *D. dugon* e as três espécies de *Trichechus* são classificadas pela International Union for the Conservation of Nature and Natural Resources (IUCN) como vulneráveis, com suas populações em diminuição (Deutsch *et al*., 2008; Keith Diagne, 2015; Marmontel *et al*., 2016; Marsh e Sobtzick, 2019).

Apesar de semelhantes na morfologia externa, a principal diferença entre os membros de Dugongidae e Trichechidae é que a nadadeira caudal em Dugongidae é bifurcada e parecida com as de golfinhos e baleias, e não redonda como em Trichechidae (Marsh, 2009). A análise filogenética mais recente do grupo (de Souza *et al*., 2021) aponta que a divergência entre Dugongidae e Trichechidae ocorreu há cerca de 46,83 Ma e que dentro de Trichechidae a diversificação ocorreu há aproximadamente 6,56 Ma, separando *T. senegalensis* do clado formado por *T. manatus* e *T. inunguis*. Esta última divergência entre *T. manatus* e *T. inunguis* foi estimada em cerca de 1,34 Ma.

T. manatus é a espécie mais bem estudada e é dividida em duas subespécies: *T. manatus latirostris* (peixe-boi da Flórida), que ocorre na península da Flórida, e *T. manatus manatus* (peixe-boi marinho), que ocorre nas Américas Central e do Sul (Domning e Hayek, 1986).

Apesar dessa divisão, análises moleculares e morfológicas sugerem que a população da costa brasileira é mais divergente em relação às populações do Caribe e Flórida, e estas duas últimas mais semelhantes, apesar de serem subespécies diferentes, separando assim *T. manatus* em duas Unidades Evolutivas Significativas (ESUs), uma na costa brasileira e uma com populações do Caribe até a Flórida (Vianna *et al*., 2006; Barros *et al*., 2017; Lima *et al*., 2021). *T. manatus* conta com três haplótipos relacionados com a distribuição geográfica ao longo da costa atlântica (Figura 2; Vianna *et al*., 2006; Lima *et al*., 2019, 2021). As três populações de *T. manatus* são divididas em uma abaixo da foz do Rio Amazonas, uma distribuída na área de influência da foz do Rio Amazonas (desde a foz do Rio Amazonas até a Guiana), e a outra acima da Guiana. Também há registros da ocorrência de híbridos entre *T. inunguis* e *T. manatus* na região da foz do Rio Amazonas, onde essas espécies ocorrem em simpatria (Vianna *et al*., 2006; Lima *et al*., 2019). Aparentemente, a área sob influência da foz do Rio Amazonas atua como barreira na reprodução entre as populações (Barros *et al*., 2017; Lima *et al*., 2019).

   T. manatus e *T. inunguis* possuem cariótipos com 2n = 48 e 2n = 56, respectivamente, e tiveram seus padrões de bandeamento GTG e CBG descritos (Assis *et al*., 1998; Kellogg *et al*., 2007). O bandeamento GTG sugere que a diferença entre os cariótipos se deu por rearranjos Robertsonianos, em que há fusão/fissão dos cromossomos no centrômero, alterando o número diploide sem alterar o número fundamental (número total de braços autossômicos). A heterocromatina constitutiva evidenciada pelo bandeamento CBG em ambas as espécies está restrita às regiões pericentroméricas dos cromossomos. A pintura com sondas de cromossomos individuais humanos em *T. manatus* permitiu estabelecer uma assinatura cromossômica ligando esta espécie aos elefantes e reforçando sua posição filogenética dentre os Afrotheria (Kellogg *et al*., 2007). São poucos os estudos sobre a fração de DNA repetitivo em Sirenia, e nenhum deles aborda DNAs satélites. O pouco que se sabe é sobre os elementos transponíveis LINEs e SINEs presentes em *T. manatus* (Nikaido *et al*., 2003; Waters *et al*., 2004).
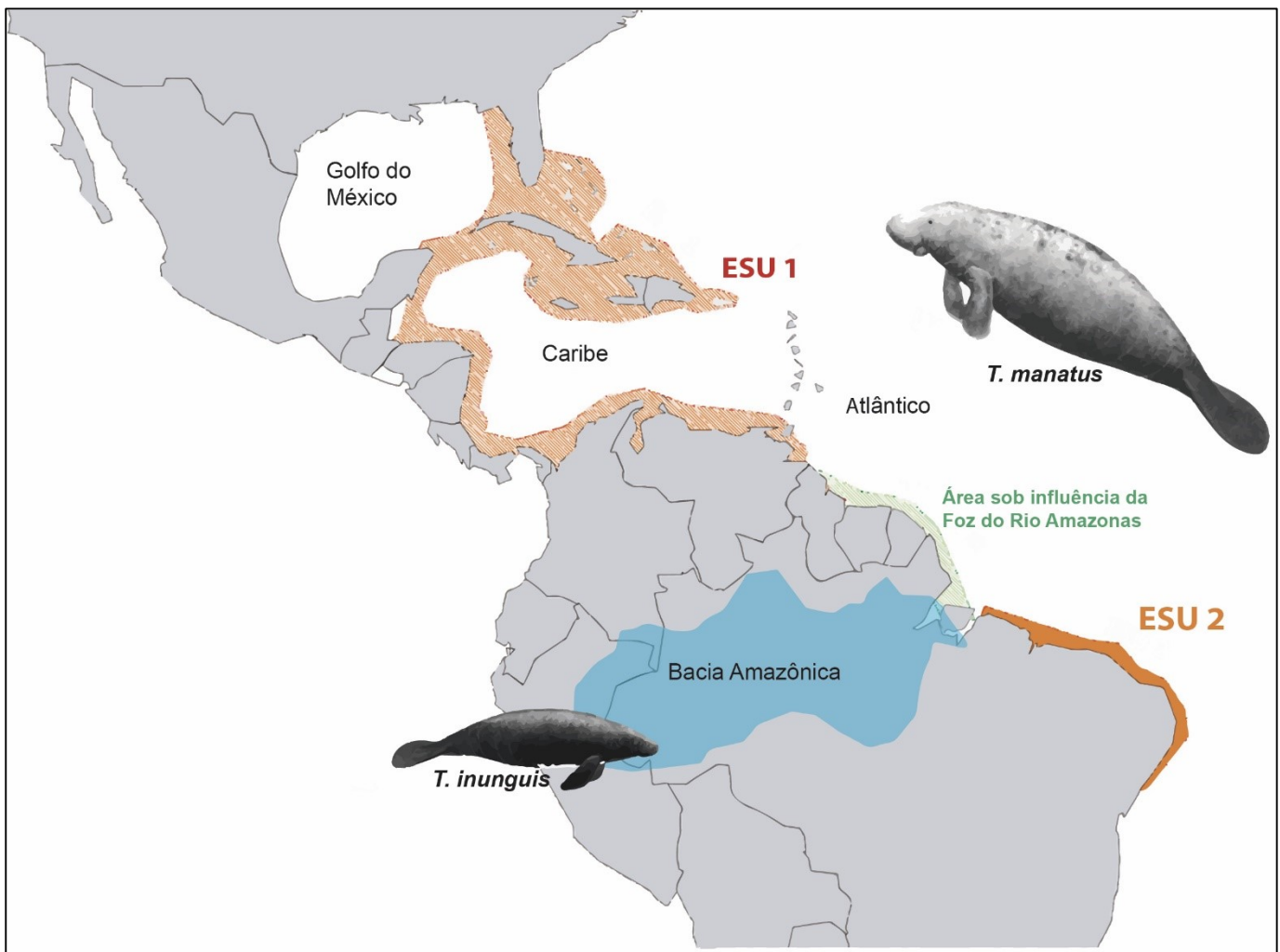
Figura 2. Mapa representando a distribuição do gênero *Trichechus* no continente americano. As ESUs de *T. manatus* baseadas em análises genéticas e morfológicas estão indicadas (Vianna *et al*., 2006; Barros *et al*., 2017; Lima *et al*., 2021). Adaptado de Lima *et al*., 2019.

2. OBJETIVOS

O objetivo deste trabalho é identificar e caracterizar sequências de DNAs satélites nos genomas de espécies do gênero *Saimiri* e da ordem Sirenia, especificamente:

Saimiri

1. Isolar e caracterizar sequências de DNAs satélites do genoma sequenciado de *Saimiri boliviensis*;

2. Mapear os DNAs satélites nos cromossomos de *S. boliviensis*, *S. sciureus*, *S. ustus* e *S. vanzolinii*;

3. Comparar as sequências dos DNAs satélites encontrados e suas características dentre as espécies do gênero *Saimiri*;

Sirenia

4. Isolar e caracterizar sequências de DNAs satélites de genomas sequenciados de *Trichechus manatus*, *Dugong dugon* e *Hydrodamalis gigas*;

5. Mapear as sequências de DNAs satélites nos cromossomos das espécies *Trichechus manatus* e *T. inunguis*;

6. Comparar as sequências encontradas e suas características dentre as espécies de Sirenia e também com as de outros mamíferos.

3. CAPÍTULO 1

Este capítulo é composto pelo artigo intitulado "Characterization of Satellite DNAs in Squirrel Monkeys genus *Saimiri* (Cebidae, Platyrrhini)", publicado no periódico *Scientific Reports* (DOI: https://doi.org/10.1038/s41598-020-64620-1). Nele usamos o genoma sequenciado de *S. boliviensis* e o software RepeatExplorer para identificar os satDNAs presentes na espécie e ampliamos a caracterização desses satDNAs com uma análise comparativa entre quatro espécies do gênero.

SCIENTIFIC REPORTS
natureresearch

Check for updates

OPEN | # Characterization of Satellite DNAs in Squirrel Monkeys genus *Saimiri* (Cebidae, Platyrrhini)

Mirela Pelizaro Valeri[1], Guilherme Borges Dias[2], Camila Nascimento Moreira[3], Yatiyo Yonenaga-Yassuda[3], Roscoe Stanyon[4], Gustavo Campos e Silva Kuhn[1] & Marta Svartman[1]✉

The genus *Saimiri* is a decades-long taxonomic and phylogenetic puzzle to which cytogenetics has contributed crucial data. All *Saimiri* species apparently have a diploid number of 2n = 44 but vary in the number of chromosome arms. Repetitive sequences such as satellite DNAs are potentially informative cytogenetic markers because they display high evolutionary rates. Our goal is to increase the pertinent karyological data by more fully characterizing satellite DNA sequences in the *Saimiri* genus. We were able to identify two abundant satellite DNAs, alpha (~340 bp) and CapA (~1,500 bp), from short-read clustering of sequencing datasets from *S. boliviensis*. The alpha sequences comprise about 1% and the CapA 2.2% of the *S. boliviensis* genome. We also mapped both satellite DNAs in *S. boliviensis*, *S. sciureus*, *S. vanzolinii*, and *S. ustus*. The alpha has high interspecific repeat homogeneity and was mapped to the centromeres of all analyzed species. CapA is associated with non-pericentromeric heterochromatin and its distribution varies among *Saimiri* species. We conclude that CapA genomic distribution and its pervasiveness across Platyrrhini makes it an attractive cytogenetic marker for *Saimiri* and other New World monkeys.

Squirrel monkeys of the genus *Saimiri* (Cebidae, Platyrrhini) are medium sized neotropical primates inhabiting forest environments of South America. They range from about 10°N to 17°S including the Amazon basin, the Guianas, and coastal zones of Central America[1,2]. As for many other New World monkey (NWM) taxa the phylogenetic relationships within the genus *Saimiri* are still debated[3]. Even the number of species is uncertain, historically ranging from one to 16 distinguished species[4–11]. In a recent molecular report on mitochondrial D-Loop and cyt *b* sequences, Alfaro *et al.*[3] presented a provisional taxonomy of seven *Saimiri* species and various subspecies: (1) *S. boliviensis*, (2) *S. cassiquiarensis* (*S. c. cassiquiarensis*, *S. c. albigena*, *S. c. macrodon* A, *S. c. macrodon* B, and *S. c. macrodon* C), (3) *S. collinsi*, (4) *S. oerstedii* (*S. o. oerstedii* and *S. o. citronellus*), (5) *S. sciureus*, (6) *S. ustus* (A, B, and C lineages), and (7) *S. vanzolinii*.

Cytogenetic studies consistently showed that all *Saimiri* have a diploid number of 2n = 44, but can differ in fundamental numbers (FN, the number of chromosome arms) which range from 74 to 78[12]. Differences in FN in *Saimiri* have traditionally been expressed as the number of acrocentric chromosomes, with five (FN = 78) to seven pairs (FN = 74). FN variation was thought to correlate with geographic distribution and taxonomy. According to Jones *et al.*[13] individuals from Costa Rica, Panama and Iquitos – Peru had five acrocentric pairs (FN = 78), those originating from Leticia – Colombia had six pairs (FN = 76) and specimens from Georgetown – Guiana had seven pairs of acrocentric chromosomes (FN = 74). Differences in FN in *Saimiri* were previously thought to be the result of pericentric inversions and reciprocal translocations[14–16] but recently Chiatante *et al.*[12], using high-resolution BAC-FISH analysis, showed that centromere repositioning explains differences in FN in *Saimiri*. These authors, however, did not address the taxonomic issue within the genus.

A class of markers that can be used to study karyotype evolution and address taxonomic issues is satellite DNAs (satDNAs). These sequences consist of tandem repeats organized in large arrays (up to Mb size) typically associated with chromosome landmarks such as centromeres, telomeres, and heterochromatic regions [e.g.[17–19],

[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]Department of Genetics and Institute of Bioinformatics, University of Georgia, Athens, GA, United States of America. [3]Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brazil. [4]Department of Biology, University of Florence, Florence, Italy. ✉e-mail: svartmanm@ufmg.br

Species	Specimen	Sex	Origin	FN	Pair 5	Pair 15
S. boliviensis	SBO1	Male	Unknown	76	SM	A
S. sciureus	SSC 782	Male	Presidente Figueiredo – Amazonas	74	A	A
S. sciureus	SSC 770	Female	Santarém – Pará	74	A	A
S. sciureus	SSC2	Male	Unknown	74	A	A
S. vanzolinii	SVA 321	Female	Lake Mamirauá, Tefé - Amazonas	76	SM	A
S. vanzolinii	SVA 322	Male	Lake Mamirauá, Tefé - Amazonas	76	SM	A
S. ustus	SUS 739	Male	Hydroelectric plant, Samuel - Rondônia	78	SM	SM
S. ustus	SUS 740	Female	Hydroelectric plant, Samuel - Rondônia	78	SM	SM
S. ustus	SUS 742	Female	Hydroelectric plant, Samuel - Rondônia	78	SM	SM
S. ustus	SUS 746	Female	Hydroelectric plant, Samuel - Rondônia	78	SM	SM
S. ustus	SUS 747	Female	Hydroelectric plant, Samuel - Rondônia	78	SM	SM
S. ustus	SUS786	Female	Hydroelectric plant, Samuel - Rondônia	78	SM	SM

Table 1. Identification, sex, collection site, fundamental number and morphology of chromosome pairs 5 and 15 of the analyzed specimens. SBO - *Saimiri boliviensis*; SSC - *Saimiri sciureus*; SVA - *Saimiri vanzolinii*; SUS - *Saimiri ustus*; FN – fundamental number; SM - submetacentric; A - acrocentric.

reviewed in[20]. Several satDNAs show a concerted mode of evolution, in which new mutations are homogenized within satDNA arrays in a genome and differentially fixed in reproductively isolated populations [reviewed in[20–22]. Mechanisms such as gene conversion and unequal crossing-over are involved in the evolutionary process known as molecular drive, responsible for the concerted evolution of satDNAs[21]. The rapid concerted evolution of satDNAs can result in high intraspecific sequence homogeneity and interspecific differences, making satDNAs potential taxonomic markers and, in some cases, allowing their use for phylogenetic inferences[20,23]. SatDNAs have been used as cytogenetic markers facilitating species identification in many taxa, including primates[24], frogs[25], fish[26] and plants[27].

The alpha is the most studied satDNA in primates. It has a centromeric location and its monomer length in Old World primates is ~170 bp. Most NWMs (Platyrrhini) have a derived alpha with ~340 bp but species of the Pitheciidae family have a monomer of ~550 bp composed by four ~170 bp subunits with the third one incomplete[17,28,29]. Alpha satDNA is often highly divergent among species and also among chromosomes of the same species[18,28]. CapA is a satDNA present only in NWMs, with ~1,500 bp monomer length and was found in the three Platyrrhini families, with different chromosome localization and abundance varying from less than 1% up to 5%[17,30,31].

The repetitive DNA fraction of the *Saimiri* genomes, including their satDNAs, are largely unexplored. In this work, we employed bioinformatic and cytogenetic tools to characterize the satDNAs of *Saimiri*. We characterized the two most abundant satDNAs of the genus and used these sequences to analyze the karyotypes of several individuals. Alpha and CapA comprise ~1% and 2.2% of the *S. boliviensis* genome, respectively. The alpha satDNA has ~340 bp, a centromeric location and high interspecific monomer homogeneity, while CapA has ~1500 bp and is associated with constitutive heterochromatin. This satDNA was mainly located in distal regions of the short arms and in the interstitial heterochromatin of some chromosomes, showing different chromosome localization among *Saimiri* species. Novel markers may help to clarify the taxonomic and phylogenetic relationships among *Saimiri* taxa.

Results

Chromosome banding. The 12 *Saimiri* individuals analyzed presented the expected diploid number of 2n = 44, but their fundamental numbers (FNs) varied due to the presence of different numbers of acrocentric chromosomes, which ranged from 10 to 14 (five to seven pairs) (Table 1). The karyotypes were arranged according to Stanyon *et al.*[32]. GTG-banding allowed the identification of all chromosomes.

The specimens identified as *S. sciureus* (SSC782, SSC770 and SSC2) had a FN = 74 and both pairs 5 and 15 were acrocentric. The *S. boliviensis* (SBO1) and the two *S. vanzolinii* specimens (SVA321 and SVA 322) had a FN = 76 with a submetacentric pair 5 and an acrocentric pair 15. All the *S. ustus* samples (SUS739, SUS740, SUS742, SUS746, SUS747 and SUS786) had a FN = 78 and both pairs 5 and 15 were submetacentric. These results supported previous conclusions[12] that centromere shifts in pairs 5 and 15 explained the morphological variation of these chromosomes and the consequent differences in FNs (Fig. 1a; Supplementary Fig. S1).

CBG-banding revealed, in addition to the pericentromeric constitutive heterochromatin, the presence of distal and interstitial heterochromatic blocks in several chromosome pairs (Fig. 1b; Supplementary Fig. S2). The distribution and abundance of distal and interstitial heterochromatin was slightly different between species. Generally, distal CBG bands were located in the short arms of submetacentric chromosomes. In all analyzed specimens, heterochromatic blocks were detected in the distal regions of the short arms of pairs 4, 7, 8, 10, 11, and 13 and in the proximal regions of both arms of chromosome 2. Interstitial CBG bands in both arms of chromosome 2 were more evident in *S. sciureus* than in *S. vanzolinii* and were very light in *S. boliviensis* and all *S. ustus*. Pairs 5 and 15 had distal CBG bands in their short arms only in their submetacentric form. The *S. sciureus* specimens SSC770, SSC782 and SSC2 had the acrocentric form of chromosomes 5 and 15 and thus did not present the distal CBG bands in these pairs. The *S. boliviensis* and *S. vanzolinii* specimens showed distal CBG bands in pair 5. Distal CBG
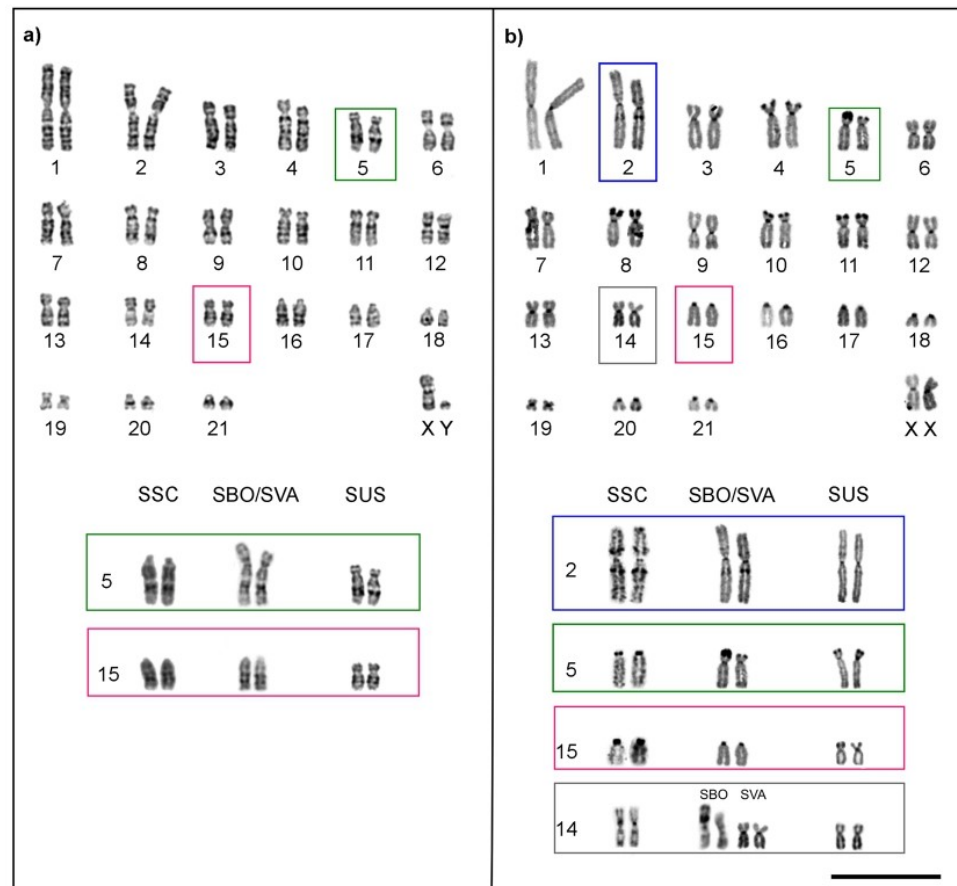
Figure 1. (**a**) Representative *Saimiri* karyotype with FN = 78 and five acrocentric pairs, pairs 5 and 15 are submetacentric (SUS 739); the lower panels show the acrocentric pairs 5 and 15 (SSC 770, SSC 782, SSC2); submetacentric pair 5 and acrocentric 15 (SBO1, SVA 321 and SVA 322), submetacentric pairs 5 and 15 (SUS). (**b**) CBG-banded *Saimiri* karyotype with FN = 76 (SVA 321), and in the boxes below pairs 2, 5, 15 and 14 from *S. sciureus* (SSC), *S. boliviensis* (SBO), *S. vanzolinii* (SVA) and *S. ustus* (SUS). Chromosome pair 2 is highlighted in blue, pair 5 in green, pair 14 in gray and pair 15 in pink. Bar = 10 μm.

bands were detected in pairs 5 and 15 of all *S. ustus* analyzed. In SSC782 and SBO1, a CBG band was detected only in one homologue of pair 14, in the proximal and distal regions of the short arm.

Satellite DNAs identification and chromosome mapping.   After careful analysis of RepeatExplorer's results, we identified two clusters that corresponded to potential satDNAs. Cluster 5 (CL5) comprises 22,193 reads (out of 2,230,692), representing ~1% of the *S. boliviensis* genome. RepeatExplorer includes partial assembly of reads into contigs and RepeatMasking of contigs using the RepBase metazoan library. This analysis indicated that CL5 corresponded to the centromeric repeat. Extending the search for CL5 sequences to the nr/nt GenBank database revealed that this cluster represents the well-known alpha satDNA, known to have a centromeric location in simian primates. Similarity searches of CL5 sequences against the *S. boliviensis* reference genome (accession GCA_000235385.1) revealed that these sequences are organized in tandem and that monomers are ~340 bp in length, which was confirmed by PCR in all *Saimiri* analyzed species (Supplementary Fig. S3a).

Fluorescent *in situ* hybridization (FISH) with the alpha satDNA in squirrel monkey chromosomes revealed its presence in the centromeric region of all chromosomes in *S. boliviensis*, *S. sciureus*, and *S. vanzolinii* (Fig. 2). In *S. ustus* it was absent from pair 12 and from one homologue of pair 7. Furthermore, the hybridization signal in chromosome 6 of SSC770, SSC 782 and SSC2 was much weaker when compared to the other species. In all *Saimiri* species, the alpha satDNA signal was more intense in acrocentric chromosomes and in the metacentric pair 19.

The second potential satDNA cluster, CL3, was initially found split into four clusters with a strongly connected component to each other (CL6, CL7, CL8 and CL12). These were merged using the RepeatExplorer cluster merger tool, thus recovering a larger CL3. This cluster comprises 49,193 reads, or ~2.2% of the genome, representing the second most abundant repeat family in the *S. boliviensis* genome (Supplementary Table S1). Analysis of this sequence in the *S. boliviensis* reference genome revealed tandem repeats with a ~1,500 bp monomer length confirmed by PCR (Supplementary Fig. S3b). Similarity searches on the nr/nt database from GenBank using partially
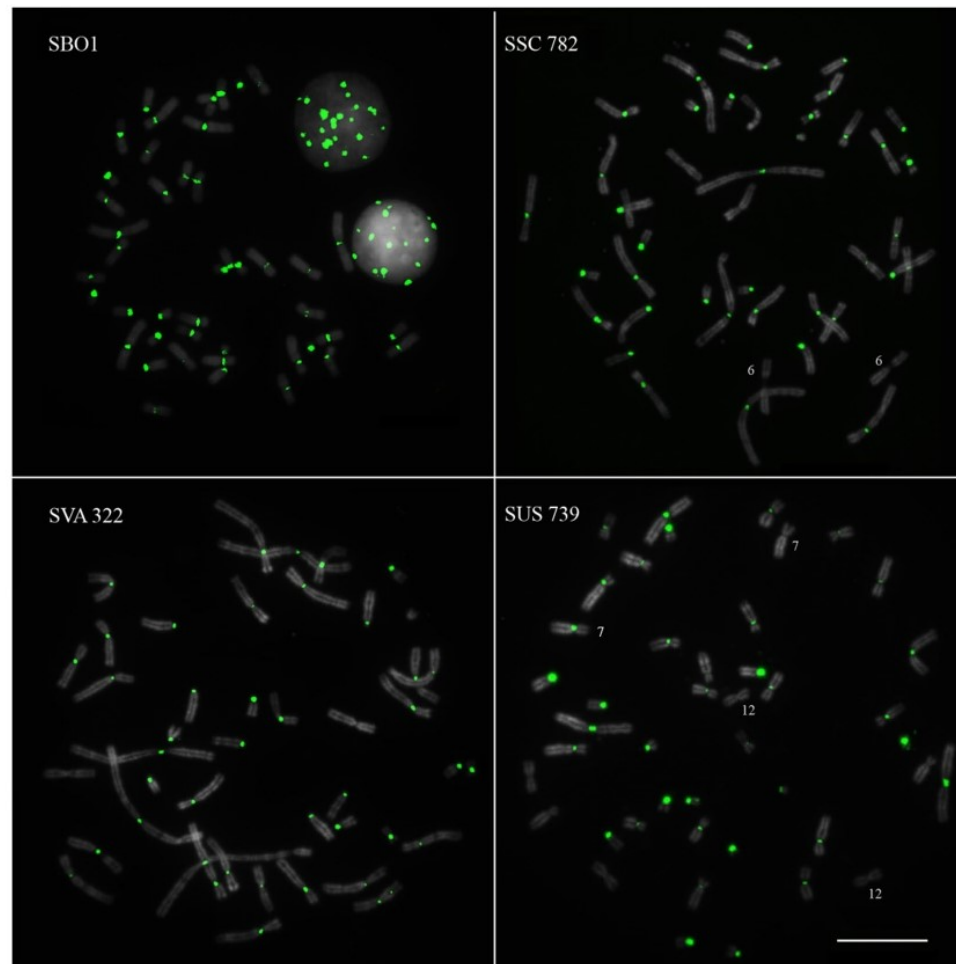
Figure 2. Metaphases of *S. boliviensis* (SBO1), *S. sciureus* (SSC 782), *S. vanzolinii* (SVA 322) and *S. ustus* (SUS 739) after FISH with the alpha satDNA probe. The signal in pair 6 of SSC 782 was weaker than in the other species; pair 12 and one homologue of pair 7 of SUS 739 did not show any signal. Bar = 10 μm.

assembled contigs from the reads of CL3 as queries revealed that this sequence is homologous to a satDNA named CapA described in *Sapajus apella*[17].

We have previously obtained a probe for CapA when studying the origin and distribution of this satDNA in mammals[31] and we used the same probe to analyze CapA distribution in the *Saimiri* genus. FISH with the CapA probe showed a distribution largely coincident with the heterochromatic regions revealed after CBG-banding in the four *Saimiri* species. Signals were mainly located in the distal regions of the short arms of submetacentric chromosomes and in the interstitial heterochromatin of some chromosome pairs (Fig. 3). There was a slight variation in CapA localization among *Saimiri* species. The difference of CapA localization between *S. boliviensis* and *S. vanzolinii*, that share the same FN = 76, was its presence in chromosome 14 of *S. boliviensis* and its absence in the same chromosome of *S. vanzolinii*. Besides *S. boliviensis*, only *S. sciureus* had CapA mapped to pair 14. Only in *S. sciureus* (FN = 74) CapA was not detected in pair 11 and its signal was much more intense in pair 2. CapA was not detected in pair 13 of *S. sciureus* and *S. ustus* (FN = 78). As observed with CBG bands, pairs 5 and 15 had distal CapA signals in their short arms only in their submetacentric form. Thus, *S. sciureus* did not show CapA in pairs 5 and 15; *S. boliviensis* and *S. vanzolinii* had the satDNA mapped in pair 5 and *S. ustus* in both pairs.

Alpha satDNA sequence analysis. Sequences of the alpha satDNA were obtained for all four *Saimiri* species studied herein from either the reference genome (*S. boliviensis*), from clones obtained by Kugou *et al.*[33] for *S. sciureus*, and from cloning and Sanger sequencing performed in this study (*S. sciureus*, *S. vanzolinii* and *S. ustus*). These sequences were aligned and compared for intra and interspecies diversity, using neighbor-joining clustering and non-metric multidimensional scaling (NMDS) ordinations. Alpha sequences showed the same value (0.16 substitutions per site) for intra and interspecific diversity. Both the neighbor-joining tree and NMDS ordinations revealed that alpha sequences do not display species-specific clustering (Fig. 4). Neighbor-joining clustering suggested a divergent group of alpha sequences, but in a branch without support. Interestingly, NMDS
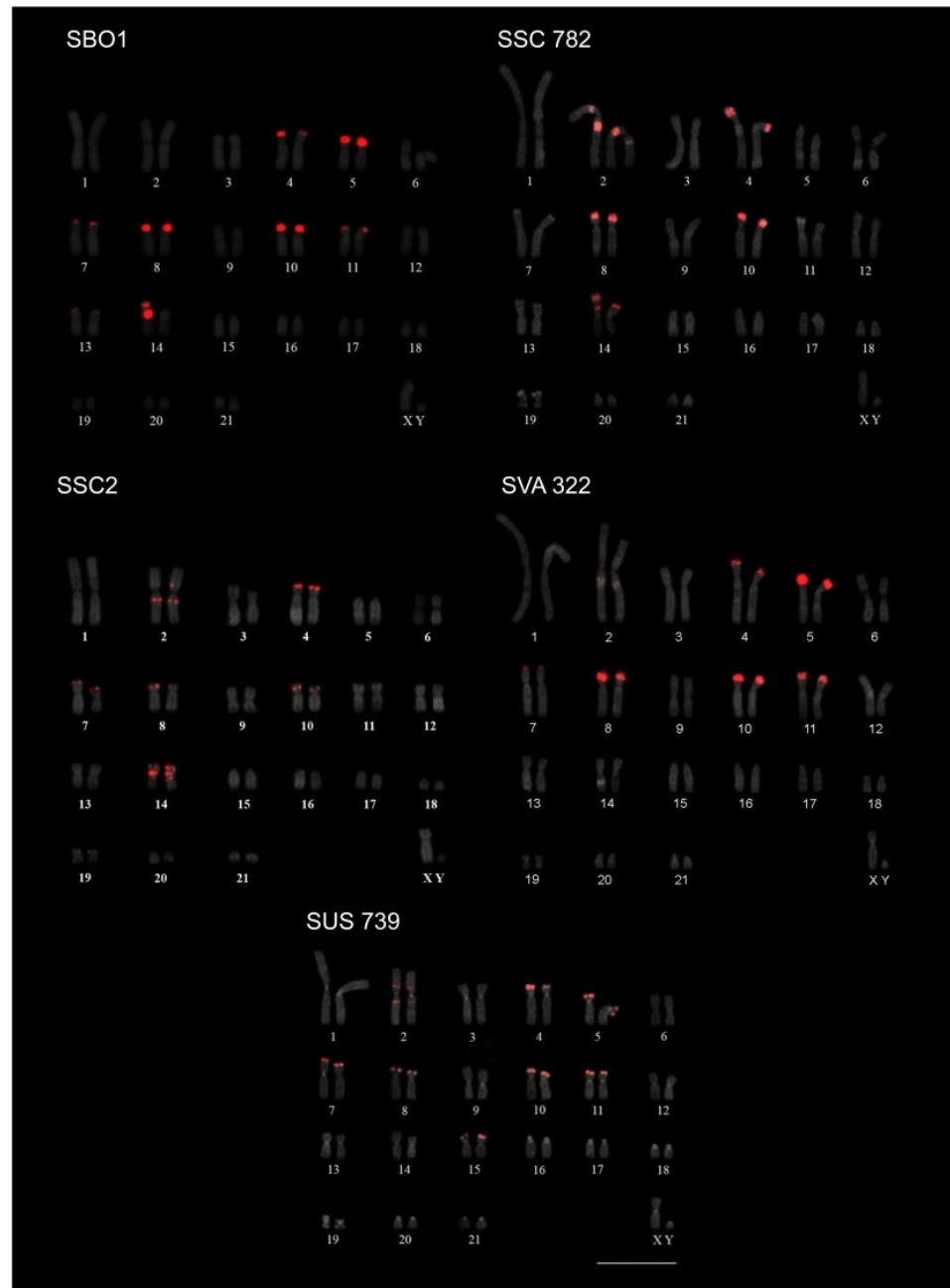
Figure 3. FISH with CapA in the karyotypes of *S. boliviensis* (SBO1), *S. sciureus* (SSC 782 and SSC2), *S. vanzolinii* (SVA 322) and *S. ustus* (SUS 739). Bar = 10 μm.

showed the same group of alpha sequences clearly distinct from the others, indicating a potential chromosome variant. The neighbor-joining and NMDS results were supported by the Maximum Likelihood method (Tamura 3-parameter model) analysis (Supplementary Fig. S4).

Discussion

We identified two abundant satDNAs in the genome of *Saimiri boliviensis*, alpha and CapA. We characterized both for repeat size, abundance, and chromosome localization in *S. boliviensis*, *S. sciureus* *S. vanzolinii*, and *S. ustus*. The alpha satDNA comprises ~1% of the *S. boliviensis* genome and its ~340 bp monomer length was confirmed in all four *Saimiri* species. This repeat structure with ~340 bp is found in most NWMs and is probably the ancestral form in the group, evolved through a duplication of the ~170 bp monomer found in Catarrhini[18,29].
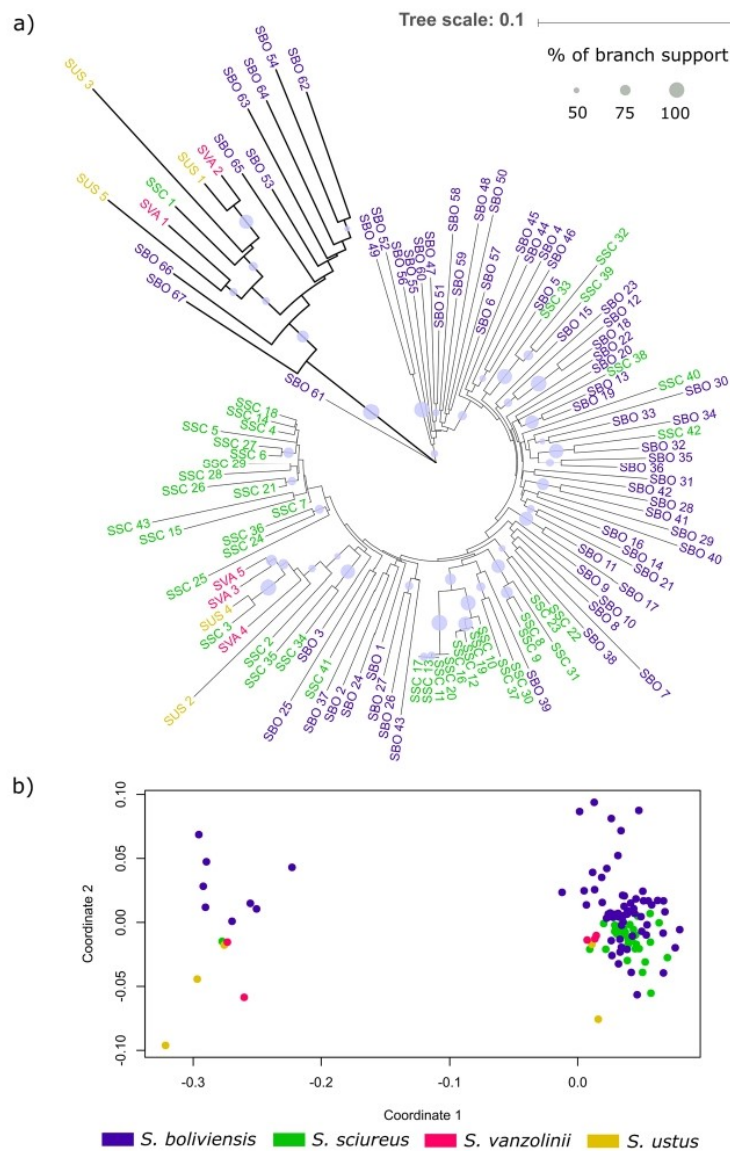
Figure 4. (**a**) Neighbor-joining clustering of alpha sequences of *S. boliviensis* (SBO), *S. sciureus* (SSC), *S. vanzolinii* (SVA) and *S. ustus* (SUS) inferred by the neighbor-joining method with 1000 bootstrap replicates. Tree visualized in iTOL v4.3.3 (https://itol.embl.de/)[52]. (**b**) Non-metric multidimensional scaling (NMDS) of evolutionary divergence among the same alpha sequences used in the phylogenetic analyses. The scaling represents euclidean distances for two dimensions (stress: 0.1127148). The sequences present in the left group of ordinations are the same of the branch in bold of the phylogenetic tree. Colors indicate the sequence taxa (bottom).

The alpha satDNA centromeric localization was observed in all chromosomes of *S. boliviensis*, *S. sciureus* and *S. vanzolinii*. Moreover, it appears to be absent or undetectable by FISH (due to low repeat number or sequence divergence) from chromosomes 7 and 12 in *S. ustus* (Fig. 5). Variable FNs among squirrel monkey species is due to centromere repositioning in pairs 5 and 15[12]. Our sample includes specimens with both morphologies of chromosomes 5 and 15 and the alpha satDNA was always detected at their centromeres, indicating that these evolutionary new centromeres (ENCs) are mature.

The alphoid satDNA family is known to be part of the centromere in simian primates (Catarrhini and Platyrrhini). In a phylogenetic analysis of primate alpha satDNA, Alkan *et al.*[18] reported that alpha satDNA repeats are evolutionarily distinct and heterogenous between human, chimpanzee and macaques. Alpha DNA was shown to have a genus-specific chromosome localization in small apes from the *Nomascus* and *Hylobates* genera, which could be used as a cytogenetic marker[24].
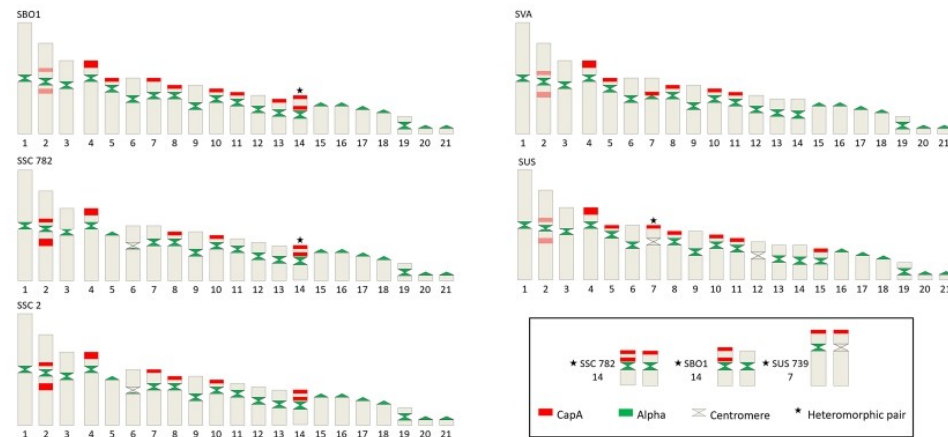
Figure 5. Schemes of *Saimiri* chromosomes showing the localization of the alpha and CapA satDNAs in *S. boliviensis* (SBO), *S. sciureus* (SSC 782 and SSC2), *S. vanzolinii* (SVA) and *S. ustus* (SUS). The heteromorphic pairs for satDNAs distribution are depicted in the inset.

Our sequence analysis of alpha repeats between *Saimiri* species revealed a high interspecific repeat homogeneity. The homogeneity found among alpha repeats of all analyzed species may be due to the recent *Saimiri* species diversification (1.4–0.8 million years ago-Mya)[3] and/or to hybridization events. In fact, hybridization has been reported between *S. sciureus* and *S. ustus*[34], between *S. boliviensis* and *S. sciureus*, and between *S. boliviensis* and *S. ustus*[35]. Another possibility is that the alpha repeats have been conserved due to their important role in centromere function and maintenance. For example, Smalec *et al.*[36] reported a conserved centromeric repeat (PM sat) among rodents of the genus *Peromyscus* and suggested that molecular drive was not the only agent in the evolution of this satDNA and that the homologous arrays may play a role in the chromosome evolution of the genus.

NMDS ordinations and neighbor-joining clustering clearly split *Saimiri* alpha repeats into two groups, one containing most sequences and the other containing only a few sequences but representing all four species. This could be an indication that monomers are evolving divergently among the chromosomes. In humans, alpha is known to have chromosome-specific sequences[37], and in *Callitrix jacchus* alpha repeats are divided into seven clusters that showed different hybridization patterns among chromosomes[29].

The second satDNA analyzed, CapA, has ~1,500 bp monomers, comprises ~2.2% of the *S. boliviensis* genome and is associated with constitutive heterochromatin. It is the second most abundant repetition in *S. boliviensis*, only surpassed by Alu transposable elements (Short Interspersed Nuclear Elements – SINEs). CapA was first identified after digestion of *Sapajus apella* (previously classified as *Cebus apella*) genomic DNA with restriction enzymes, and was shown to be ~1,500 bp long and to comprise about 5% of the genome[17,30]. In a recent study, we provided evidence that the CapA satDNA is homologous to an intronic sequence of the NOS1AP gene (*Homo sapiens nitric oxide synthase 1 adaptor protein*), and likely originated from this single-copy sequence through duplication and unequal crossing-over[31]. We also showed that CapA is present in representatives of the three Platyrrhini families (Cebidae, Atelidae and Pitheciidae; except in Callitrichines and in the *Callicebus* genus) with different genome abundance and chromosome localization, always associated with constitutive heterochromatin.

CapA repeats in *Saimiri* chromosomes are mostly enriched in the distal heterochromatin of submetacentric chromosomes, but the overall distribution and abundance differed among species (Fig. 5).

Interestingly, CapA was mapped in some chromosomes involved in rearrangements. For instance, CapA distribution in *Saimiri* chromosomes 5 and 15 was coincident with the CBG bands: they were absent in the acrocentric forms and present in the distal heterochromatin of the short arm in the submetacentric variants of these chromosomes. This variation may be related to the morphology of these chromosomes and perhaps with the process of ENCs formation in *Saimiri*. According to the putative ancestral Platyrrhini karyotype, the ancestral morphology of chromosome 5 and 15 is submetacentric and acrocentric, respectively[32,38]. Taking this into account, chromosome 5 would have lost CapA sequences when the chromosome became acrocentric and the new centromere occupied this region. Conversely, in chromosome 15, the CapA satDNA would have colonized the old centromere region after the chromosome became submetacentric.

SBO1 was previously analyzed by Chiatante *et al.*[12], who reported pair 14 as heteromorphic, with isoforms A and B differing by a paracentric inversion and two heterochromatic blocks, one distal and one in the proximal region of the short arm. This same polymorphism was also described in several specimens of squirrel monkeys, especially in those from Bolivia, in which heterozygotes for pair 14 were more common than homozygotes[14,39]. Chiatante *et al.*[12] suggested that an inversion in the isoform B may have carried some telomeric DNA sequences to the interstitial region of the short arm, explaining the heterochromatic blocks. Three of our specimens (SBO1, SSC 782 and SSC2) presented the isoform B of chromosome 14, and in all of them CapA was located in the interstitial heterochromatin and distal region. In SSC 782, CapA was also detected in the distal region of the short arm of isoform A. This indicates that CapA could have been involved in the generation of this paracentric inversion, instead of the suggested telomeric sequences[12].

Chromosome painting with human probes revealed that *Saimiri* 2q is formed by homologues of human chromosomes 14 and 15 (HSA 14 and 15) and their signals are intercalated due to inversions[32]. CapA was mapped on 2q in a region that may correspond to the breakpoint of one of this inversions. CapA was also observed in *Saimiri* 2p, which is homologous to HSA 9[32]. In *Sapajus apella* HSA 9 corresponds to pair 19, which has an interstitial heterochromatic block where CapA was observed and corresponds to an inversion[17,40]. Pair 7 of *Saimiri* is formed by the association of HSA 15/HSA 2[32] and presents CapA in the short arm of most specimens analyzed (Fig. 5), although seemingly not at the association region.

Although our data do no allow to pinpoint CapA to chromosome rearrangements, they indicate that a detailed analysis of these regions could provide some new interesting data.

CapA distribution differed more among *Saimiri* species than CBG bands. Even individuals of different species with the same karyotype and FN differed in their CapA localization, as for instance, *S. boliviensis* and *S. vanzolinii*, both with FN = 76 (Fig. 3). The specimens identified as *S. sciureus* (FN = 74) did not show CapA in pair 11 and the satDNA was more abundant in pair 2 when compared with the other species (Fig. 3). The specimen SSC 770 identified as *S. sciureus* has a similar karyotype and CapA distribution to SSC 782, however its collection site suggests that it may actually be a *S. collinsi* according to Alfaro et al.[3]. No karyotype has been described for *S. collinsi* or for samples collected within its geographic distribution, preventing us from further conclusions.

The variable chromosomal localization of CapA among the *Saimiri* species analyzed herein suggests that, combined with FN, this sequence may be used as a valuable tool in taxonomic identification. Hybridization in squirrel monkeys has been reported in captivity and in nature[4,6,35] and CapA mapping may also reveal the origin of chromosome sets in hybrids more precisely than chromosome morphology or banding patterns.

The rapid expansion and diversification of squirrel monkeys occurred in the Amazon basin with all the speciation events estimated in the range between 1.4–0.8 Mya[3] starting with the divergence between *S. boliviensis* and the ancestor to all other squirrel monkeys at 1.4–1.6 Mya[2]. Despite the very short time since diversification, we were able to detect different CapA chromosomal distributions among the *Saimiri* specimens analyzed. Conversely, the alpha satDNA displayed high interspecific repeat homogeneity which could be related to its role in centromere function and maintenance.

The cytogenetic information about *Saimiri* is poor when compared to morphological and molecular data. The number of specimens karyotypically analyzed is low, and most samples are not geotagged. Until now, the karyotypes of only five out of the seven recently recognized species[3] have been reported and they are very similar, including their banding patterns. In order to further validate CapA as a marker, more geotagged specimens need to be analyzed. The study of species not yet karyotyped will also help clarify *Saimiri*'s taxonomic puzzle and inform its systematics.

Our results indicate that CapA is a promising cytogenetic marker and that it could be useful for taxonomic, phylogenetic and conservation studies of *Saimiri*. In addition, the fact that CapA is present across Platyrrhini further extends its utility as a marker for chromosome and genome evolution studies in NWMs[31]. The availability of new markers is especially important in the face of threats of extinction to an alarming large number of NWM species due to rapid habitat loss[41].

Materials and Methods

Biological samples and chromosome banding. Chromosome spreads and genomic DNAs were obtained from fibroblast cultures of 12 squirrel monkeys: one *S. boliviensis*, three *S. sciureus*, two *S. vanzolinii*, and six *S. ustus*. We used previously established cell lines and the available details for the samples are provided in Table 1. The work did not involve the direct use of animals, so ethical permission was not required. Genomic DNAs were purified with the Wizard Genomic DNA Purification Kit (Promega) and chromosome spreads were obtained from cultured cells according to Stanyon and Galleni[42]. GTG- and CBG-banding were performed according to Seabright[43] and Sumner[44], respectively. The samples from *S. boliviensis* (SBO1) were previously analyzed in Chiatante et al.[12] and Capozzi et al.[45]. The *S. sciureus* (SSC2) sample was previously reported in Chiatante et al.[12]. The two *S. vanzolinii* (SVA 321 and SVA 322) samples were previously reported in Yonenaga-Yassuda and Chu[46].

Satellite DNAs identification. The RepeatExplorer pipeline[47,48] was used to identify satDNAs based on all to all similarity comparison of the Illumina reads of *S. boliviensis* (NCBI SRA access: SRR317821). A total of 2,230,692 Illumina reads (~100 bp long) comprising 7% of the estimated *S. boliviensis* genome were randomly sampled and used in this analysis. The results are represented as graph-based clusters of similar reads and the shape of the clusters is indicative of the nature of the different repeat families (e.g. globula and ring-like structures suggest tandemly organized repeats). The number of reads within the clusters (out of the total used in the analysis) indicates the abundance of that cluster in the genome. The clustering process tends to split large repeats into several clusters, but the pipeline has a separate re-clustering tool for a user-aided merging of the clusters. The reads that make up each cluster are partially assembled into contigs that can be used for repeat annotation. The minimum overlapping lengths used for clustering and assembly were 55 and 40 bp, respectively. Sequences in clusters with globular/ring-like structure were analyzed in detail through similarity searches against the *S. boliviensis* reference genome (accession GCA_000235385.1) using the BLASTn tool with default parameters[49]. Additionally, satDNA clusters were annotated through similarity searches against the whole non-redundant nucleotide collection in GenBank.

PCR amplification, cloning and sequencing of satellite DNAs. Isolation of alpha satDNA was performed through polymerase chain reaction (PCR) of *Saimiri* genomic DNAs using the following specific primer set: alpha-F (ACAGGGAAATATCTGCTTCTAAATC) and alpha-R (GCTTACTGCTGTTTCTTCCATATG). The thermocycling conditions were as follows: 95 °C—3 min, 35 cycles: 95 °C—30 sec; 60 °C—30 sec; 72 °C—1 min;

final elongation: 72 °C—3 min. The repeat monomers obtained by PCR were cloned into pGEM-T Easy vector plasmids (Promega) and used to transform *E. coli* strain XL1-BLUE (Phoneutria) through electroporation. Recombinant colonies were capillary sequenced with the ABI3130 platform (Applied Biosystems) and are available in GenBank under accession numbers MK879580-MK879592. We obtained three sequenced clones from *S. sciureus*, five from *S. vanzolinii* and five from *S. ustus*. These sequenced clones were used in molecular analyses and as FISH probes. CapA amplification was performed using genomic DNAs from *S. boliviensis* and *S. vanzolinii* and the primer set CapA-F (ACTTCCTCACTGACCTGTCTT) and CapA-R (GGGCTGATGCTTAATGTAGCA). CapA isolation and cloning were previously performed in Valeri *et al.*[31] using human DNA and the same primer set described. The sequenced product is deposited in GenBank under the accession number MG264524 and was used as a FISH probe in this study.

Fluorescent *in situ* hybridization (FISH).   FISH was performed using alpha and CapA sequences as probes on metaphase spreads of the four *Saimiri* species. SatDNA probes were prepared from pGEM-T Easy cloned sequences and labeled by nick translation with biotin-16-dUTP or digoxigenin-11-dUTP (Nick Translation mix, Roche Applied Science). Chromosomes were denatured in 70% formamide/2xSSC (saline-sodium citrate buffer) at 75% for 105 sec. The hybridization mix consisted of 100 ng of labeled probe in 50% formamide/2x-SSC and was denatured for 10 min at 98 °C and added to the chromosome spreads. Hybridization was carried at 37 °C for 16–20 hours. Post-hybridization washes consisted of three baths of 2xSSC at 45 °C for 5 min each. Immunodetection was performed with neutravidin+rhodamine or avidin+FITC conjugates (Roche Applied Science) and the slides were mounted with DAPI 1:500 in Slowfade (Life Technologies). Chromosome identification was based on the Q-banding pattern produced after DAPI staining. The analyses were performed under a Zeiss Axioimager 2 epifluorescence microscope equipped with a CCD camera and image acquisition was performed with the AxioVision software (Carl Zeiss MicroImaging, Jena, Germany).

Alpha satDNA sequence analyses.   In order to test if satDNA sequences of a species showed signs of concerted evolution we aligned the satDNA sequence monomers with MUSCLE[50] and performed clustering using the neighbor-joining method with 1000 bootstrap replicates in MEGA X[51]. Alpha sequences intra and interspecific diversities were calculated using the Maximum Composite Likelihood model in MEGA X. In addition, we computed pairwise distances between all sequences using a Maximum Composite Likelihood model in MEGA X and used these data to perform a NMDS analysis. The dataset used for these analyses was mainly composed of alpha monomeric sequences retrieved from GenBank (https://www.ncbi.nlm.nih.gov/genbank/). Thus, we performed BLASTn using the alpha consensus sequence as query against the *S. boliviensis* reference genome (accession GCA_000235385.1) and *S. sciureus* alpha clones (accession LC075928-LC075953) obtained by Kugou *et al.*[33]. Because GenBank lacks data from *S. vanzolinii* and *S. ustus*, we added sequences of these species obtained herein by cloning and Sanger sequencing. The total of alpha monomeric sequences used was 120, composed by 67 from *S. boliviensis*, 43 from *S. sciureus*, five from *S. vanzolinii* and five from *S. ustus*. The resulting neighbor-joining tree was visualized in iTOL v4.3.3 (https://itol.embl.de/)[52]. NMDS ordinations were generated with the R package vegan[53], representing the divergence between sequences values in Euclidian distances for two dimensions. RStudio v1.1.463[54] was used to conduct the NMDS analyses and to plot the ordinations.

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/) and in NCBI trace and short-read archive (https://trace.ncbi.nlm.nih.gov/Traces/sra/sra.cgi?). All accession numbers are provided in the Methods section.

References

1. Lavergne, A. *et al.* Phylogeny and phylogeography of squirrel monkeys (genus Saimiri) based on cytochrome b genetic analysis. *Am. J. Primatol.* **72**, 242–253 (2010).
2. Chiou, K. L., Pozzi, L., Lynch Alfaro, J. W. & Di Fiore, A. Pleistocene diversification of living squirrel monkeys (Saimiri spp.) inferred from complete mitochondrial genome sequences. *Mol. Phylogenet. Evol.* **59**, 736–745 (2011).
3. Lynch Alfaro, J. W. *et al.* Biogeography of squirrel monkeys (genus Saimiri): South-central Amazon origin and rapid pan-Amazonian diversification of a lowland primate. *Mol. Phylogenet. Evol.* **82**, 436–454 (2015).
4. Hershkovitz, P. Taxonomy of squirrel monkeys genus Saimiri (Cebidae, Platyrrhini): A preliminary report with description of a hitherto unnamed form. *Am. J. Primatol.* **7**, 155–210 (1984).
5. Hershkovitz, P. Uacaries, New World monkeys of the genus Cacajao (Cebidae, Platyrrhini): A preliminary taxonomic review with the description of a new subspecies. *Am. J. Primatol.* **12**, 1–53 (1987).
6. Thorington, R. W. The taxonomy and distribution of squirrel monkeys (Saimiri) in Handbook of squirrel monkey research (ed. Rosenblum, L. A. & Coe, C. L.) 1–33 (Springer, 1985).
7. Costello, R. K., Dickinson, C., Rosenberger, A. L., Boinski, S. & Szalay, F. S. Squirrel monkey (genus Saimiri) taxonomy in Species, species concepts and primate evolution (ed. Kimbel, W. H. & Martin, L. B.) 177–210 (Springer, 1993).
8. Cropp, S. & Boinski, S. The Central American squirrel monkey (Saimiri oerstedii): Introduced hybrid or endemic species? *Mol. Phylogenet. Evol.* **16**, 350–365 (2000).
9. Boinski, S. & Cropp, S. J. Disparate data sets resolve squirrel monkey (Saimiri) taxonomy: Implications for behavioral ecology and biomedical usage. *Int. J. Primatol.* **20**, 237–256 (1999).
10. Groves C., Wilson D.E. & Reeder D.A.M. Order Primates in Mammal species of the world: a taxonomic and geographic reference (ed. Wilson, D. & Reeder, D.) 111–184 (The Johns Hopkins University Press, 2005).
11. Rylands, A. B., Mittermeier, R.A., Bezerra, B. M., Paim, F. P. & Queiroz, H. L. Family Cebidae (squirrel monkeys and capuchins) in Handbook of the mammals of the world, volume 3. Primates (ed. Mittermeier, R. A., Rylands, A. B. & Wilson, D. E.) 348–413 (Lynx Edicions, 2013).

12. Chiatante, G. *et al.* Centromere repositioning explains fundamental number variability in the New World monkey genus Saimiri. *Chromosoma* **126**, 519–529 (2017).
13. Jones, T. C., Thorington, R. W., Hu, M. M., Adams, E. & Cooper, R. W. Karyotypes of squirrel monkeys (Saimiri sciureus) from different geographic regions. *Am. J. Phys. Anthropol.* **38**, 269–277 (1973).
14. Ma, N. S. & Jones, T. C. Added heterochromatin segments in chromosomes of squirrel monkeys (Saimiri sciureus). *Folia Primatol. (Basel)* **24**, 282–292 (1975).
15. Lau, Y. F. & Arrighi, F. E. Studies of the squirrel monkey, Saimiri sciureus, genome. Cytological characterizations of chromosomal heterozygosity. *Cytogenet. Genome Res.* **17**, 51–60 (1976).
16. Cambefort, Y. & Moro, F. Cytogenetics and taxonomy of some South Bolivian monkeys. *Folia Primatol. (Basel)* **29**, 307–314 (1978).
17. 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).
18. Alkan, C. *et al.* Organization and evolution of primate centromeric DNA from whole-genome shotgun sequence data. *PLoS Comput. Biol.* **3**, 1807–1818 (2007).
19. Prakhongcheep, O. *et al.* Two types of alpha satellite DNA in distinct chromosomal locations in Azara's owl monkey. *DNA Res.* **20**, 235–240 (2013).
20. Plohl, M., Meštrović, N. & Mravinac, B. Satellite DNA evolution in Repetitive DNA (ed. Schmid, M.) 126–152 (Karger, 2012).
21. Dover, G. Molecular drive: A cohesive mode of species evolution. *Nature* **299**, 111–117 (1982).
22. Kuhn, G. C. S., Sene, F. M., Moreira-Filho, O., Schwarzacher, T. & Heslop-Harrison, J. S. Sequence analysis, chromosomal distribution and long-range organization show that rapid turnover of new and old pBuM satellite DNA repeats leads to different patterns of variation in seven species of the Drosophila buzzatii cluster. *Chromosom. Res.* **16**, 307–324 (2008).
23. Bachmann, L., Schibel, J. M., Raab, M. & Sperlich, D. Satellite DNA as a taxonomic marker. *Biochem. Syst. Ecol.* **21**, 3–11 (1993).
24. Baicharoen, S. *et al.* Locational diversity of alpha satellite DNA and intergeneric hybridization aspects in the Nomascus and Hylobates genera of small apes. *PLoS One* **9**, e109151, https://doi.org/10.1371/journal.pone.0109151 (2014).
25. Picariello, O., Feliciello, I., Bellinello, R. & Chinali, G. S1 satellite DNA as a taxonomic marker in brown frogs: Molecular evidence that Rana graeca graeca and Rana graeca italica are different species. *Genome* **45**, 63–70 (2002).
26. Garrido-Ramos, M. A. *et al.* Evolution of centromeric satellite DNA and its use in phylogenetic studies of the Sparidae family (Pisces, Perciformes). *Mol. Phylogenet. Evol.* **12**, 200–204 (1999).
27. Kirov, I. V., Kiseleva, A. V., Van Laere, K., Van Roy, N. & Khrustaleva, L. I. Tandem repeats of Allium fistulosum associated with major chromosomal landmarks. *Mol. Genet. Genomics* **292**, 453–464 (2017).
28. Alves, G., Seuánez, H. N. & Fanning, T. Alpha satellite DNA in neotropical primates (Platyrrhini). *Chromosoma* **103**, 262–267 (1994).
29. Cellamare, A. *et al.* New insights into centromere organization and evolution from the white-cheeked gibbon and marmoset. *Mol. Biol. Evol.* **26**, 1889–1900 (2009).
30. Malfoy, B. *et al.* Nucleotide sequence of an heterochromatic segment recognized by the antibodies to Z-DNA in fixed metaphase chromosomes. *Nucleic Acids Res.* **14**, 3197–3214 (1986).
31. Valeri, M. P., Dias, G. B., Pereira, V. D. S., Kuhn, G. C. S & Svartman, M. An eutherian intronic sequence gave rise to a major satellite DNA in Platyrrhini. *Biol. Lett.* **14**, 20170686, https://doi.org/10.1098/rsbl.2017.0686 (2018).
32. Stanyon, R. *et al.* Fluorescence *in situ* hybridization (FISH) maps chromosomal homologies between the dusky titi and squirrel monkey. *Am. J. Primatol.* **50**, 95–107 (2000).
33. Kugou, K., Hirai, H., Masumoto, H. & Koga, A. Formation of functional CENP-B boxes at diverse locations in repeat units of centromeric DNA in New World monkeys. *Sci. Rep.* **6**, 27833, https://doi.org/10.1038/srep27833 (2016).
34. Silva, B. T. F. *et al.* Protein electrophoretic variability in Saimiri and the question of its species status. *Am. J. Primatol.* **29**, 183–193 (1993).
35. Carneiro, J., Rodrigues-Filho, L. F. S., Schneider, H. & Sampaio, I. Molecular data highlight hybridization in squirrel monkeys (Saimiri, Cebidae). *Genet. Mol. Biol.* **39**, 539–546 (2016).
36. 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).
37. Willard, H. F. Chromosome-specific organization of human alpha satellite DNA. *Am. J. Hum. Genet.* **37**, 524–532 (1985).
38. Stanyon, R. *et al.* Primate chromosome evolution: Ancestral karyotypes, marker order and neocentromeres. *Chromosome Res.* **16**, 17–39 (2008).
39. Moore, C. M., Harris, C. P. & Abee, C. R. Distribution of chromosomal polymorphisms in three subspecies of squirrel monkeys (genus Saimiri). *Cytogenet. Cell Genet.* **53**, 118–122 (1990).
40. García, F. *et al.* Chromosomal homologies between humans and Cebus apella chromosomes revealed by ZOO-FISH. *Mamm. Genome* **11**, 399–401 (2000).
41. IUCN 2019. The IUCN Red List of Threatened Species. Version 2019-2. Downloaded on 18 July 2019. http://www.iucnredlist.org.
42. Stanyon, R. & Galleni, L. A rapid fibroblast culture technique for high resolution karyotypes. *Bolletino di Zool.* **58**, 81–83 (1991).
43. Seabright, M. A rapid banding technique for human chromosomes. *The Lancet* **298**, 971–972 (1971).
44. Sumner, A. T. A simple technique for demonstrating centromeric heterochromatin. *Exp. Cell Res.* **75**, 304–306 (1972).
45. Capozzi, O., Archidiacono, N., Lorusso, N., Stanyon, R. & Rocchi, M. The 14/15 association as a paradigmatic example of tracing karyotype evolution in New World monkeys. *Chromosoma* **125**, 747–756 (2016).
46. Yonenaga Yassuda, Y. & Chu, T. H. Chromosome banding patterns of Saimiri vanzolinii ayres, 1965 (Primates, Cebidae). *Pap. Avulsos Zool.* **36**, 165–168 (1985).
47. Novák, P., Neumann, P. & Macas, J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* **11**, 378, https://doi.org/10.1186/1471-2105-11-378 (2010).
48. 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).
49. 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).
50. Edgar, R. C. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792–1797 (2004).
51. Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: Molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
52. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* **47**, W256–W259, https://doi.org/10.1093/nar/gkz239 (2019).
53. Dixon, P. VEGAN, a package of R functions for community ecology. *J. Veg. Sci.* **14**, 927–930 (2003).
54. RStudio Team. RStudio: Integrated Development for R [Computer software]. RStudio, Inc.; http://www.rstudio.com/ (2015).

Acknowledgements

Author contributions

M.P.V. and G.B.D. carried out bioinformatic, cytogenetic and molecular analyses, participated in data analyses, designed the study and drafted the manuscript; M.P.V. prepared all figures; R.S. carried out cytogenetic analyses, participated in data analyses and helped drafting the manuscript; C.N.M., Y.Y. and R.S. obtained the materials for molecular and cytological analyses; G.B.D., G.C.S.K. and M.S. conceived and coordinated the study, and helped drafting the manuscript. All authors gave final approval for publication.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41598-020-64620-1.

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.

Supplementary Information for

Characterization of Satellite DNAs in Squirrel Monkeys genus *Saimiri* (Cebidae, Platyrrhini)
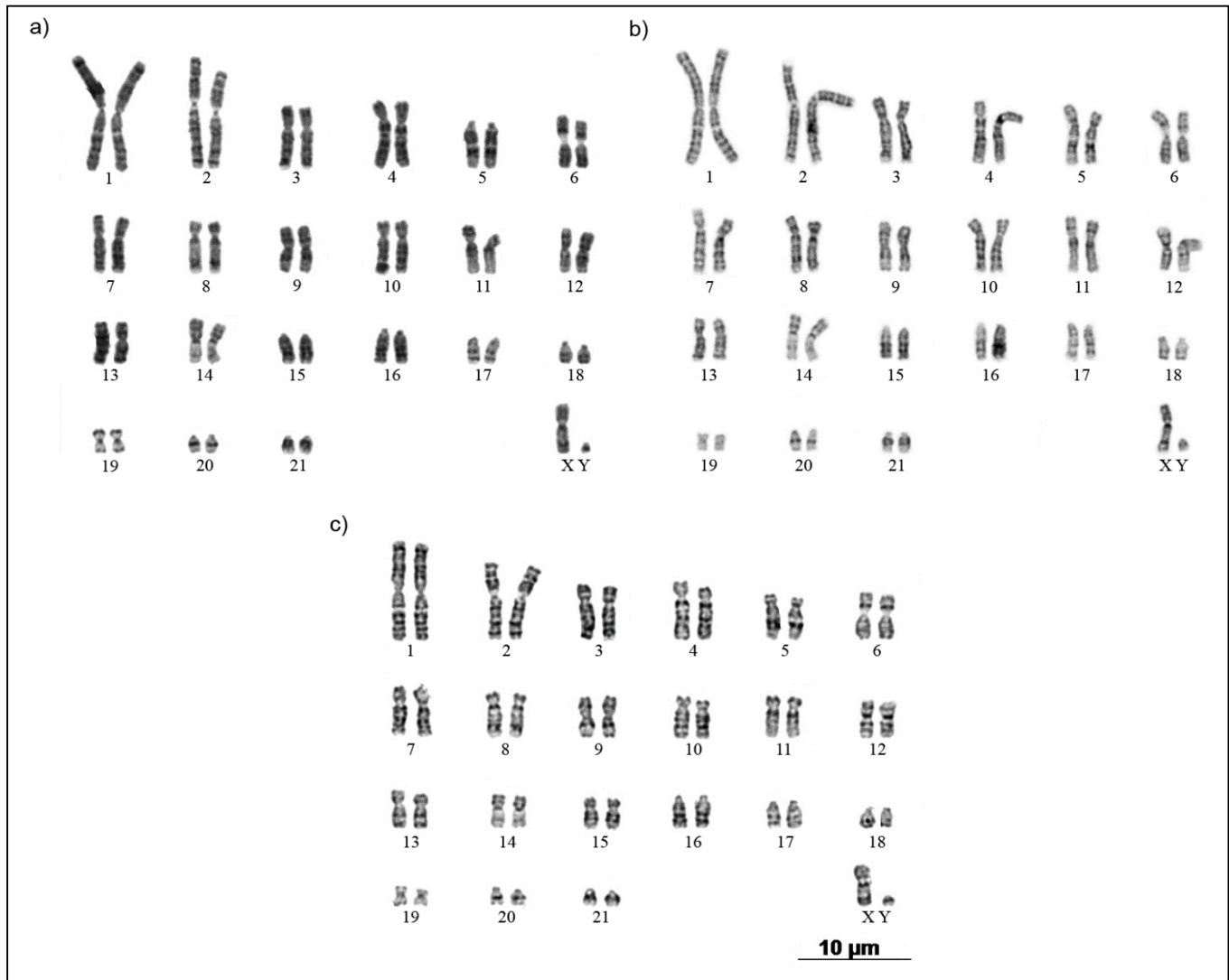
Supplementary Figure S1. Representative *Saimiri* karyotypes of (a) *S. sciureus* with five acrocentric pairs, chromosomes 5 and 15 are submetacentric (FN=78) from cell line SSC 782; (b) *S. boliviensis* and *S. vanzolinii* with six acrocentric pairs, chromosome 5 is submetacentric and 15 is acrocentric (FN=76) from cell line SVA 322; (c) *S. ustus* with seven acrocentric pairs, chromosomes 5 and 15 are acrocentric (FN=74) from cell line SUS 739.
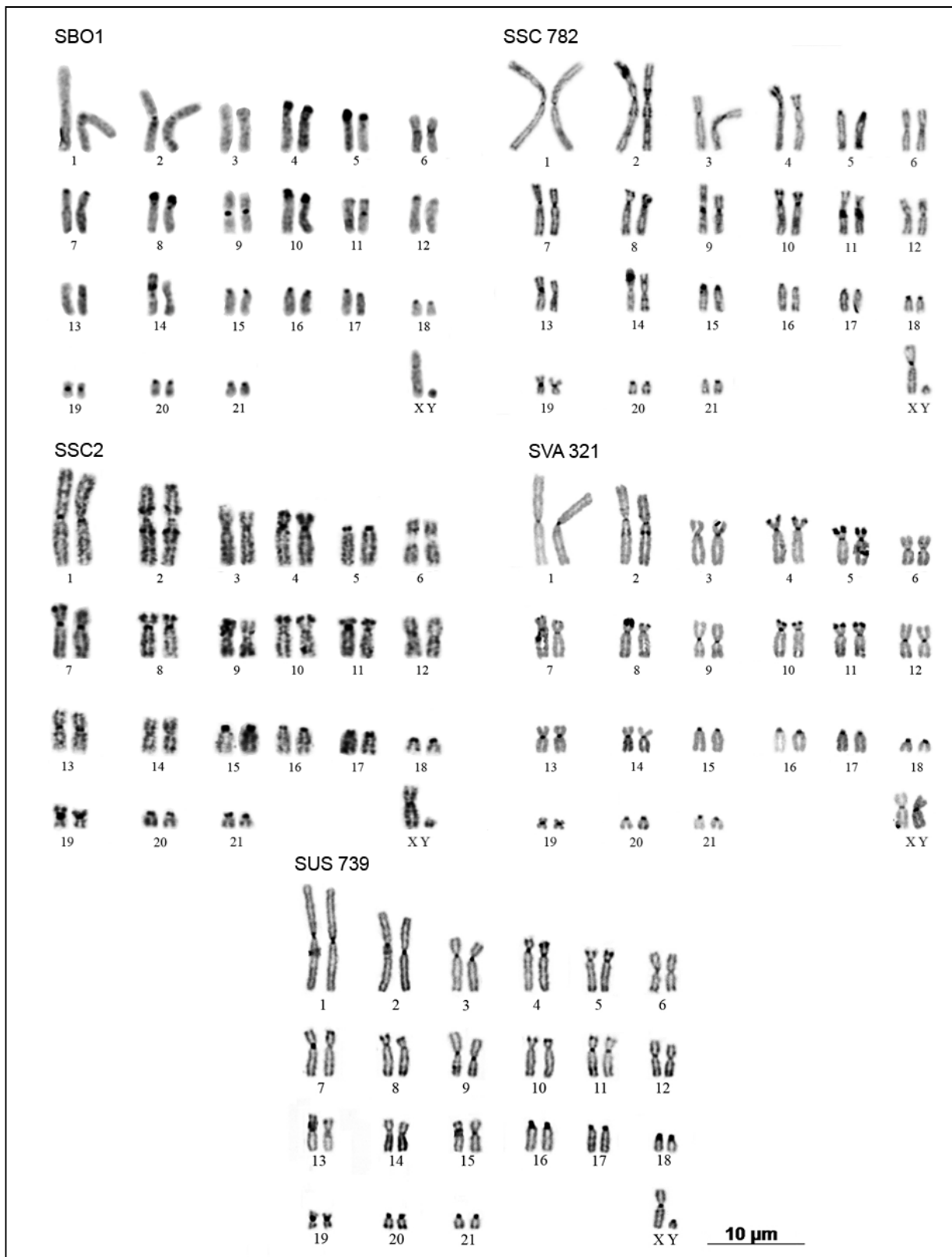
Supplementary Figure S2. CBG-banded *Saimiri* karyotypes of *S. boliviensis* (SBO1), *S. sciureus* (SSC 782 and SSC2), *S. vanzolinii* (SVA 322) and *S. ustus* (SUS 739).
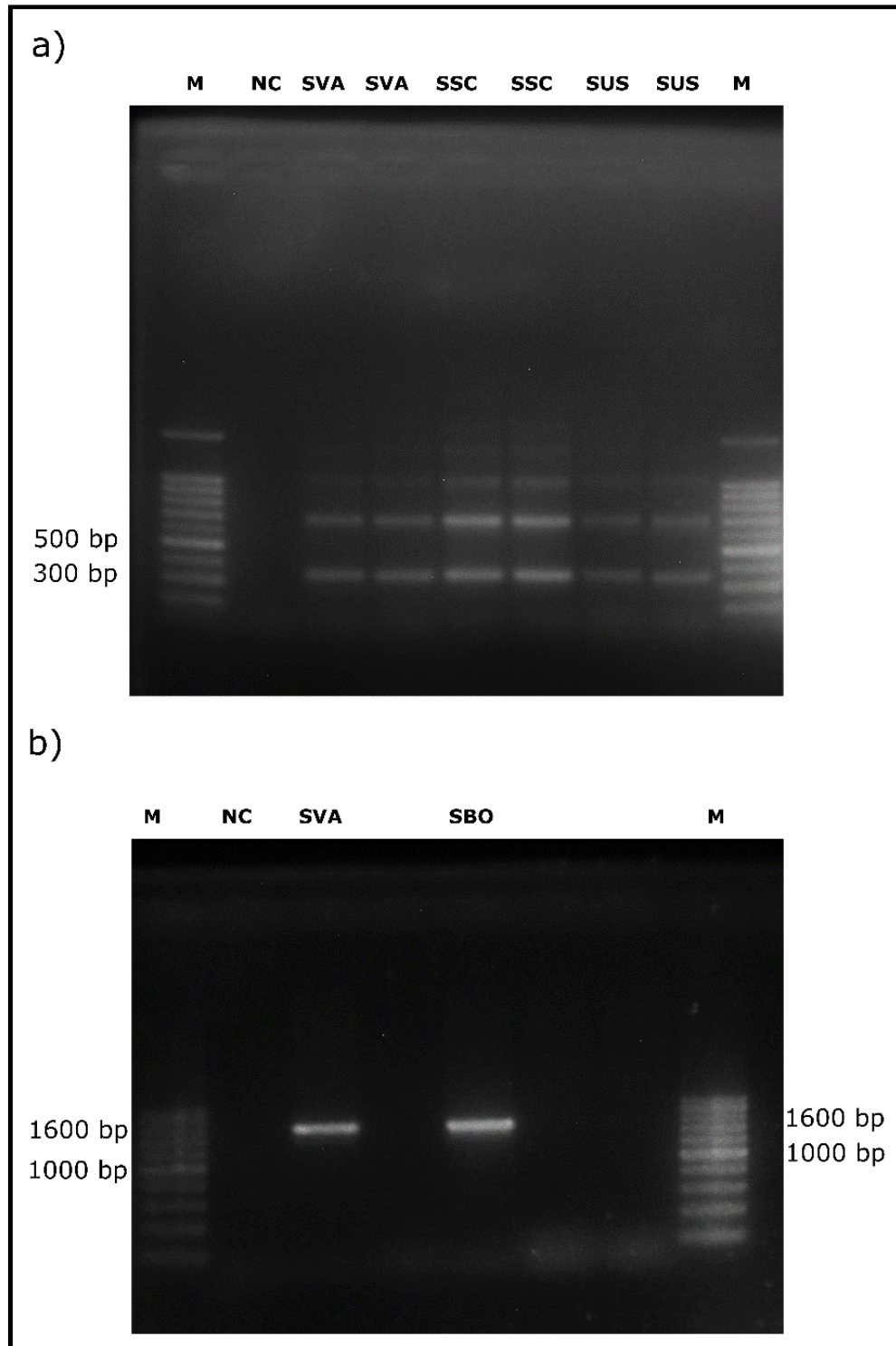
Supplementary Figure S3. Agarose gel (1%) showing the PCR products of *Saimiri* species using a) Alpha and b) CapA primers. M - DNA size marker (a) 100 bp ladder; b) 200 bp ladder); NC – Negative control (no DNA); SBO - *Saimiri boliviensis;* SSC - *Saimiri sciureus; SVA - Saimiri vanzolinii; SUS - Saimiri ustus.*
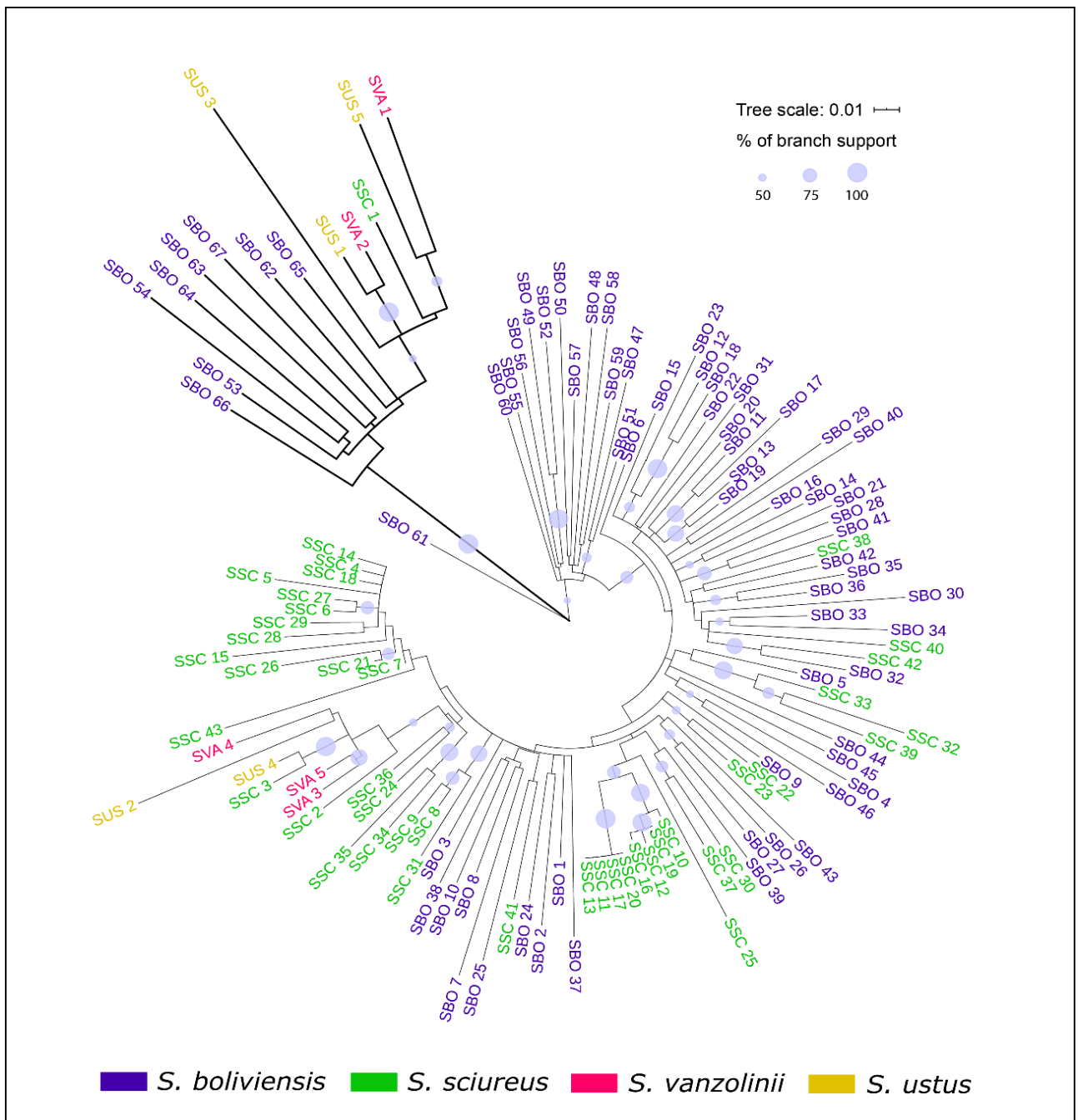
Supplementary Figure S4. Phylogenetic analysis of alpha sequences of *S. boliviensis* (SBO), *S. sciureus* (SSC), *S. vanzolinii* (SVA) and *S. ustus* (SUS) inferred by the Maximum Likelihood method (Tamura 3-parameter model) from 500 replicates using MEGA X. Tree visualized in iTOL v4.3.3 (https://itol.embl.de/).

Supplementary Table S1. Clusters retrieved by RepeatExplorer from a sample of the sequencing reads of *Saimiri boliviensis* (SRA accession: SRR317821).

Cluster	Read Number	Genome Proportion [%][a]	Proportion of Similarity Hits to Other Clusters[b]	Annotation[c]
1	108845	4.880	3.4000	SINE.Alu
2	68906	3.090	9.3000	SINE.Alu
3[d]	49193	2.210	0.000028	Unclassified
4	23622	1.060	0.4400	LINE.L1/Satellite
5[e]	22193	0.995	0.0012	Satellite.centr
6	15798	0.708	0.1300	LINE.L1
7	11074	0.496	1.1000	LINE.L1/ Satellite
8	9773	0.438	2.5000	LINE.L1/ Satellite
9	8954	0.401	1.5000	LINE.L1
10	7535	0.338	0.0072	LTR.ERVL.MaLR
11	7338	0.329	0.6900	LINE.L1
12	6147	0.276	0.0470	LINE.L1
13	4939	0.221	0.0000	Unclassified
14	4783	0.214	0.0000	Unclassified
15	4723	0.212	0.0000	Unclassified
16	4567	0.205	0.0000	Unclassified
17	2612	0.117	0.0000	DNA.hAT.Charlie
18	2038	0.091	0.0500	LINE.L1
19	1847	0.083	0.0300	LTR.ERVL.MaLR
20	1205	0.054	81.0000	SINE.Alu
21	946	0.042	0.0000	Unclassified
22	861	0.039	0.0000	LTR.ERV1
23	709	0.032	42.0000	SINE.Alu/ SINE.B4
24	656	0.029	0.0260	Unclassified
25	639	0.029	0.5800	LTR.ERV1
26	575	0.026	0.0000	LTR.ERVL
27	497	0.022	1.3000	LINE.L1
28	462	0.021	0.5000	DNA.hAT.Charlie/LTR.ERVL.MaLR
29	454	0.020	0.0000	Unclassified
30	451	0.020	0.0000	Unclassified
31	409	0.018	4.2000	SINE.7SL/srpRNA/SINE.Alu
32	408	0.018	0.6900	LINE.L1/Satellite/LTR.ERVK
33	380	0.017	0.9500	LINE.L1
34	362	0.016	0.0000	LTR.ERVL
35	315	0.014	0.0690	LTR.ERVL.MaLR
36	308	0.014	0.0000	LTR.ERVL.MaLR
37	284	0.013	0.0000	LINE.L1
38	245	0.011	0.0000	LTR.ERVL.MaLR
39	244	0.011	0.0000	LINE.L1
40	232	0.010	0.0000	LTR.ERV1/SINE.MIR/DNA.hAT.Blackjack
41	227	0.010	0.0000	LTR.ERV1

[a] Only clusters with an estimated genome proportion of at least 0.01% of the genome are included.

[b] This proportion indicates whether a significant number of reads from a given cluster also have similarity with reads from other clusters. It is a measure of uniqueness/redundancy of the repeat family.

[c] Annotation is given as displayed in the RepeatExplorer output. Clusters with at least 3% of matching similarity hits with known repeats are annotated accordingly (based on the mammalian repeat library of Repbase). Anonymous clusters are marked as "Unclassified".

[d] This cluster represents the CapA satellite DNA.

[e] This cluster represents the alpha satellite DNA, a well-known component of primate centromeres.

4. CAPÍTULO 2

Este capítulo é composto pelo artigo intitulado "First Description of a Satellite DNA in Manatees' Centromeric Regions", publicado no periódico *Frontiers in Genetics* (DOI: 10.3389/fgene.2021.694866). Nele descrevemos pela primeira vez um satDNA no genoma de *T. manatus*, o TMAsat, e o caracterizamos em outras espécies de Sirenia.

frontiers
in Genetics

First Description of a Satellite DNA in Manatees' Centromeric Regions

Mirela Pelizaro Valeri[1], Guilherme Borges Dias[2], Alice Alves do Espírito Santo[1], Camila Nascimento Moreira[3], Yatiyo Yonenaga-Yassuda[3], Iara Braga Sommer[4], Gustavo C. S. Kuhn[1] and 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, Brazil, [2]Department of Genetics and Institute of Bioinformatics, University of Georgia, Athens, GA, United States, [3]Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil, [4]Centro Nacional de Pesquisa e Conservação da Biodiversidade Marinha do Nordeste, Instituto Chico Mendes de Conservação da Biodiversidade, Brasília, Brazil

Trichechus manatus and *Trichechus inunguis* are the two Sirenia species that occur in the Americas. Despite their increasing extinction risk, many aspects of their biology remain understudied, including the repetitive DNA fraction of their genomes. Here we used the sequenced genome of *T. manatus* and TAREAN to identify satellite DNAs (satDNAs) in this species. We report the first description of TMAsat, a satDNA comprising ~0.87% of the genome, with ~684 bp monomers and centromeric localization. In *T. inunguis*, TMAsat showed similar monomer length, chromosome localization and conserved CENP-B box-like motifs as in *T. manatus*. We also detected this satDNA in the *Dugong dugon* and in the now extinct *Hydrodamalis gigas* genomes. The neighbor-joining tree shows that TMAsat sequences from *T. manatus*, *T. inunguis*, *D. dugon*, and *H. gigas* lack species-specific clusters, which disagrees with the predictions of concerted evolution. We detected a divergent TMAsat-like homologous sequence in elephants and hyraxes, but not in other mammals, suggesting this sequence was already present in the common ancestor of Paenungulata, and later became a satDNA in the Sirenians. This is the first description of a centromeric satDNA in manatees and will facilitate the inclusion of Sirenia in future studies of centromeres and satDNA biology.

Keywords: tandem repeats, *Trichechus manatus*, *Trichechus inunguis*, chromosome mapping, fluorescent *in situ* hybridization, TAREAN

INTRODUCTION

The order Sirenia encompasses four extant herbivorous aquatic mammals. The Dugongidae family includes the *Dugong dugon* and the Steller's sea cow *Hydrodamalis gigas*, the latter now extinct due to overhunting, and the Trichechidae family includes three manatee species: *Trichechus manatus*, *Trichechus inunguis*, and *Trichechus senegalensis* (Domning, 2018). *Dugong dugon* occurs across coastal waters in the Indo-West Pacific Ocean, and *T. senegalensis* is restricted to the west coast of Africa, making *T. manatus*, the West Indian manatee, and *T. inunguis*, the Amazonian manatee, the only sirenians to occur in the Americas. The West Indian manatee occurs in Caribbean waters and the Atlantic coast ranging from Florida to the northeast of Brazil, and *T. inunguis* is found along the Amazon River basin (Bonvicino et al., 2020).

All extant sirenians are considered as vulnerable by the International Union for Conservation of Nature and Natural Resources (IUCN; Deutsch et al., 2008; Keith, 2015; Marmontel et al., 2016; Marsh and Sobtzick, 2019).

The West Indian manatee has two recognized subspecies: *Trichechus manatus latirostris* (Florida Manatee), found in the United Estates and Gulf of Mexico coasts, and *T. m. manatus* (Antillean manatee), found in the Caribbean, Central and South America. Recent morphological and genetic analyses suggest the need for a revision in the *T. manatus* taxonomy considering the influence of the Amazon River as a barrier to gene flow. These studies showed that the *T. m. manatus* populations from the Caribbean and up to the Amazon River mouth are phylogenetically closer to the populations of *T. m. latirostris* from the United States than to the Brazilian *T. m. manatus* populations south of the Amazon River mouth (Vianna et al., 2006; Barros et al., 2017; Lima et al., 2019, 2021). Hybrids between *T. manatus* and *T. inunguis* have also been reported on the sympatric area at the Amazon River mouth (Vianna et al., 2006; Lima et al., 2019; Luna et al., 2021).

Satellite DNAs (satDNAs) are a type of repetitive DNA found in most eukaryotic genomes. They are arranged as long arrays of tandem repeats with variable unit length, number of copies and chromosome organization. SatDNAs are usually associated with chromosome landmarks such as centromeres, telomeres, and heterochromatic regions. Despite the fact that satDNAs do not encode proteins, they are associated with important biological functions such as formation and maintenance of heterochromatin at telomeres and centromeres, and maintenance of chromosome integrity and genome stability (reviewed in Shapiro and von Sternberg, 2005; Biscotti et al., 2015; Shatskikh et al., 2020). SatDNAs can form higher-order repeat (HOR) units made of multimers with a number of diverged monomers that are tandemly repeated as a set (reviewed in Plohl et al., 2012; Vlahović et al., 2016). HOR organization has been found in several satDNAs, including the alfa centromeric satDNA in humans, and may be relevant to the centromeric function (Sujiwattanarat et al., 2015; Sullivan et al., 2017). In addition, satDNAs monomer sequences can present internal repetitions, which may be related with secondary structures relevant to centromeric function (Kasinathan and Henikoff, 2018). Centromeric satDNAs in mammals usually present the CENP-B box, a conserved 17 bp region known to be the DNA-binding domain for the centromeric protein B (CENPB), with nine nucleotides (nTTCGnnnnAnnCGGGn) composing the most evolutionarily conserved domain (ECD; Muro et al., 1992; Masumoto et al., 2004; Alkan et al., 2011; Kasinathan and Henikoff, 2018). Most satDNAs are under concerted evolution, a process by which new mutations within monomers are quickly homogenized across the repeat family and fixed in reproductively isolated populations, resulting in intraspecific repeat homogeneity but interspecific divergence (Dover, 1982; Plohl et al., 2012; Smalec et al., 2019). Moreover, according to the library model, related species may share a collection of satDNAs sequences with mostly quantitative interspecies differences due to expansion or contraction (even elimination) during the evolution (Fry and Salser, 1977; Meštrović et al., 1998).

Another aspect of satDNAs evolution is their relationship with mobile elements, since there are several examples of satDNAs derived from transposons and retrotransposons in plants and animals (reviewed in Meštrović et al., 2015).

The repetitive DNA fraction of manatees' genomes has been poorly studied, especially in the case of satDNAs. We used the sequenced genome of *T. manatus* and the TAREAN (Novák et al., 2017, 2020) pipeline to explore the satDNAs present in this genome. Herein, we describe for the first time the centromeric satDNA of the West Indian manatee, which we found to be also present in the Amazonian manatee, the dugong, and in the extinct Steller's sea cow. We characterized this sequence *in silico* and mapped it in *T. manatus* and *T. inunguis* chromosomes. In addition, we investigated the presence of the TMAsat sequence in mammals outside the order Sirenia, which allowed us to establish a rough timeline for its origin.

MATERIALS AND METHODS

De novo Identification of Satellite DNAs

In order to identify satDNAs in manatees, we used whole-genome sequencing data from *T. m. latirostris* (accession number SRR328416) available in the *National Center for Biotechnology Information –* NCBI and the TAREAN pipeline (Novák et al., 2017). The first step of this pipeline is a graph-based clustering, which performs all to all similarity comparisons of DNA sequencing reads, resulting in clusters of those reads derived from repetitive elements. Then, it examines the presence of circular or globula-like graph structures to identify potential tandem repeats, classified as putative high or low confidence satellites. The raw Illumina reads (~100 bp long) used in this analysis were randomly sampled by TAREAN, comprising ~2.4% (870,965 reads) of the 3.67 pg estimated genome size (Kasai et al., 2013). The reads that make up each cluster are partially assembled into contigs that were used for repeat annotation with the CENSOR web server (Kohany et al., 2006) that contains a collection of Mammalia repeats from RepBase, updated in 08-24-2020 (Bao et al., 2015). The single potential tandem repeat cluster (13) with globula/ring-like structure was analyzed in detail through similarity searches against the *T. manatus* reference genome (accession GCA_000243295.1) using the BLASTn tool with default parameters (Altschul et al., 1990) to verify if the sequence is a tandem repeat. In addition to the annotation using the CENSOR web server, this cluster was annotated through BLASTn similarity searches against the whole nucleotide collection (nr/nt).

The identified satDNA sequence was characterized regarding its genome proportion, monomer length, AT content, and presence of internal direct or inverted duplications. The satDNA genome proportion was estimated by TAREAN. TAREAN tries to improve the assembly process by applying a k-mer-based approach to obtain a less fragmented monomer consensus, but it restricts itself to the 50% most prevalent k-mers in a cluster. For this reason, we chose the whole-genome assembly resource as a more representative sample of TMAsat diversity. The most common sequence (MCS) of TMAsat was generated using Geneious (prime version 2020.2.4) with a 25% threshold and 66 monomeric sequences retrieved from the reference genome, previously aligned with the muscle aligner

implemented in MEGA X. The MCS was used to estimate monomer length, AT content, and presence of internal repetitions. The last feature was also conducted in the Geneious software using the diagonal plot method in high sensitive mode, with window size of 50 bp and identity threshold of 60%.

We searched for the presence of TMAsat in the two other Sirenia species with a sequenced genome available in NCBI, *D. dugon* (under accession numbers of assembled genome GCA_015147995.1 and raw Illumina reads DRR251525) and the extinct *H. gigas* (under accession numbers of assembled genome GCA_013391785.1 and raw Illumina reads SRR12067498). First, we used TMAsat sequence as query in BLASTn similarity searches against these assembled genomes. In addition, we also used the raw Illumina reads (~150 bp long) and TAREAN to identify TMAsat in these genomes. The analyzed reads were randomly sampled by TAREAN totalizing 1,038,927 in *D. dugon* and 570,097 in *H. gigas*. The MCS of TMAsat in *D. dugon* and *H. gigas* were generated using monomeric sequences retrieved from the reference genome after BLAST searches, totalizing 50 sequences from *D. dugon* and 40 from *H. gigas*. The TMAsat MCS in *T. inunguis* was obtained using the five cloned sequences obtained by PCR. All MCS were generated as described previously for *T. manatus*.

Biological Samples

We used biological samples of *T. manatus* and *T. inunguis* to determine TMAsat chromosomal distribution and investigate its presence in *T. inunguis*, whose genome has not been sequenced. Skin sample from a male *T. manatus* captured at Porto de Pedras/AL, Brazil (−9.164167 and −35.294444) in 2019 was provided by CEPENE/ICMBio (SISBIO 60829-2) and used for fibroblast culture. A fibroblast cell line from a male *T. inunguis* established in 1998 was provided by Dr. Yatiyo Yonenaga-Yassuda, from the University of São Paulo, and was previously analyzed by Assis et al. (1998). Chromosome spreads from fibroblast cultures were obtained according to Stanyon and Galleni (1991) and genomic DNAs were extracted with the Wizard Genomic DNA Purification Kit (Promega).

PCR Amplification, Cloning, and Sequencing of Satellite DNAs

TMAsat was amplified by PCR from the *T. inunguis* genomic DNA using primers designed from the satDNA consensus sequence (estimated by TAREAN) as follow: TMAsat-F CTCCTTCAAGCTGCTTAACT and TMAsat-R GGGAACTTACACTTGCTGCT. The PCR cycling conditions were as follows: 95°C – 3 min, 35 cycles: 95°C – 30 s; 55°C for 30 s; 72°C – 1 min; and 72°C – 3 min for final elongation. The PCR product corresponding to monomer size was excised from the agarose gel and purified using the Illustra GFX PCR DNA and Gel Band Purification Kit. The purified products were ligated into the pGEM-T Easy vector (Promega) and used in the transformation of *Escherichia coli* XL1-BLUE strain electrocompetent cells (Phoneutria). Five recombinant colonies of TMAsat were sequenced (access numbers MW272776– MW272780) with the ABI3130 platform (Applied Biosystems).

Fluorescence *in situ* Hybridization

Fluorescent *in situ* hybridization (FISH) was performed using the TMAsat cloned (MW272776) sequence as probe on metaphase spreads of *T. manatus* and *T. inunguis*. FISH was performed with 200 ng of biotin-labelled probes, following (Valeri et al., 2020). The analyses and image acquisition were performed under a Zeiss Axioimager 2 epifluorescence microscope equipped with a CCD camera and with the AxioVision software (Carl Zeiss MicroImaging, Jena, Germany), respectively.

In silico Characterization of satDNAs

DNA polymorphisms and nucleotide diversity along the satDNA sequences were analyzed using the software DnaSP 6.12.03 (Rozas et al., 2017) with the same monomer sequences used to generate the MCS from *T. manatus*, *D. dugon*, and *H. gigas*. In this analysis, the monomer sequences were previously aligned with the muscle method (Edgar, 2004) implemented in MEGA X and the window length and step size were set for 10 and 1 bp, respectively. Windows were classified as conserved or variable if they exhibited more than two SDs bellow or above the nucleotide average variability, respectively.

Monomer sequences of TMAsat from *T. manatus*, *T. inunguis*, *D. dugon*, and *H. gigas* were aligned with the muscle method implemented in MEGAX and used for the construction of a neighbor-joining tree. These sequences were the same used to obtain the MCS, totalizing 161 sequences, including 66 from *T. manatus*, five from *T. inunguis*, 50 from *D. dugon* and 40 from *H. gigas*. The neighbor-joining tree was obtained using MEGA X with 500 bootstrap replicates and the final tree was visualized in iTOL v4.3.3[1] (Letunic and Bork, 2019). We also used the same set of sequences to estimate the inter- and intra-specific nucleotide divergence (number of base substitutions per site), as well as the average nucleotide divergence over all pairwise sequence comparisons using MEGA X.

We searched for any putative CENP-B box in TMAsat MCS from *T. manatus*, *T. inunguis*, *D. dugon*, and *H. gigas* using the 17 bp sequence containing the ECD (nTTCGnnnnAnnCGGGn; Masumoto et al., 2004) and CENP-B box sequences of *Loxodonta africana/Dasypus novemcintus* (CTTTGCCGAGAACGGAG; Alkan et al., 2011). This search was conducted in the Geneious software in global pairwise alignment mode and 51% similarity cost matrix.

To investigate the presence of TMAsat in other mammals, we utilized the MCS from *T. manatus* as query in BLASTn similarity searches against Mammalia (NCBI:txid40674) wgs database excluding Sirenia (NCBI:txid9774; search date 06-07-2021). The flanking regions of TMAsat similarity hits were analyzed with the CENSOR web server (Kohany et al., 2006) containing the Mammalia RepBase library (updated in 06-14-2021; Bao et al., 2015). To better analyze these hits with the TMAsat consensus sequence, we compared their sequences using dotplots and pairwise alignments in the Geneious software.

---

[1]https://itol.embl.de/

RESULTS

In silico Identification and satDNA Analyses

The only potential satDNA identified (with low confidence) by TAREAN in the *T. manatus* genome was represented by the cluster 13. This sequence was analyzed in detail through similarity searches against the *T. manatus* reference genome (accession GCA_000243295.1) using the BLASTn tool with default parameters (Altschul et al., 1990). Despite being classified by TAREAN with low confidence, we verified this sequence tandemly repeated at least 25 times on assembled contigs of *T. manatus*. These repeats comprise 0.87% of the genome of *T. manatus* with monomer length of ~684 bp estimated by
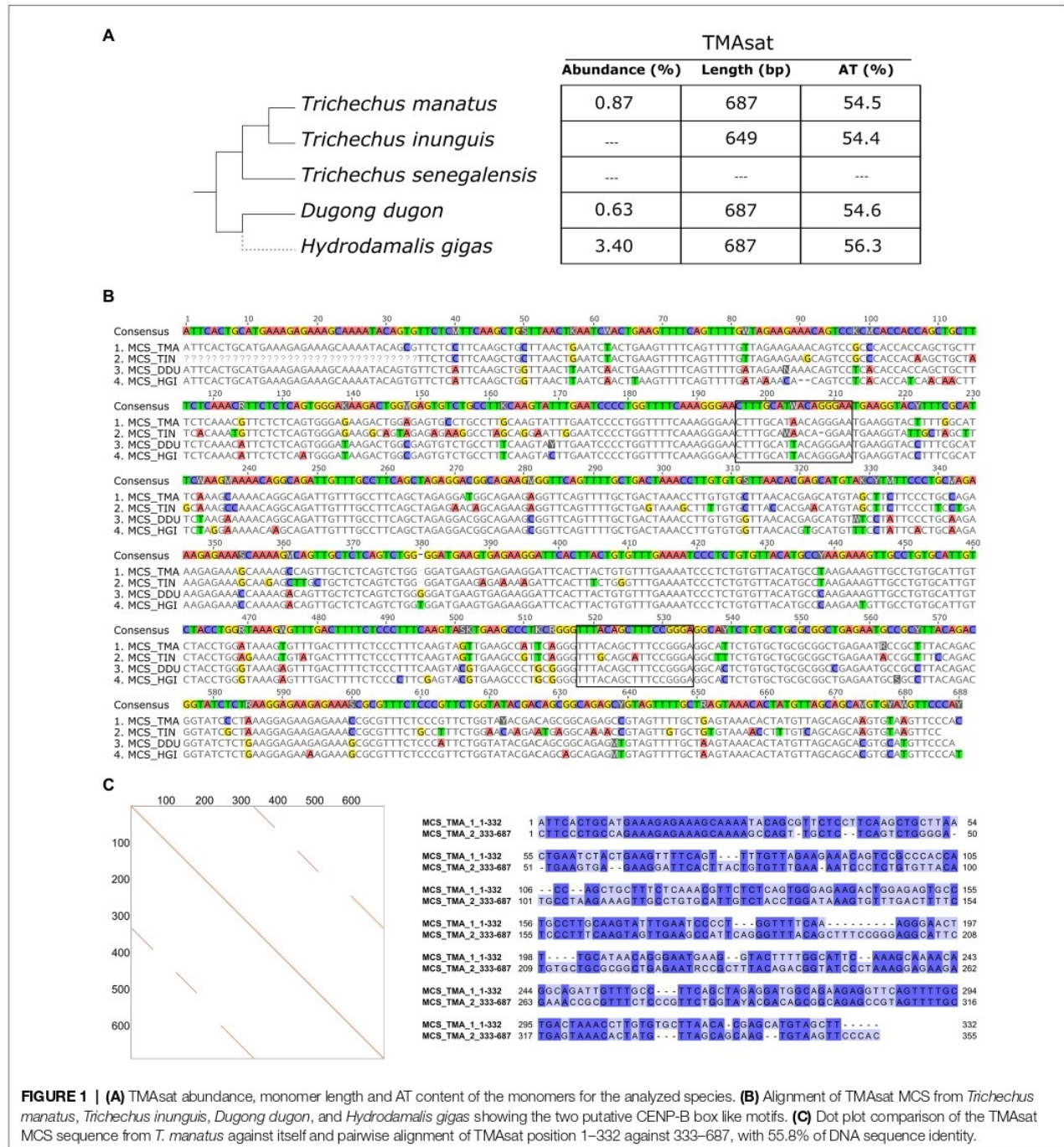
FIGURE 1 | **(A)** TMAsat abundance, monomer length and AT content of the monomers for the analyzed species. **(B)** Alignment of TMAsat MCS from *Trichechus manatus*, *Trichechus inunguis*, *Dugong dugon*, and *Hydrodamalis gigas* showing the two putative CENP-B box like motifs. **(C)** Dot plot comparison of the TMAsat MCS sequence from *T. manatus* against itself and pairwise alignment of TMAsat position 1–332 against 333–687, with 55.8% of DNA sequence identity.
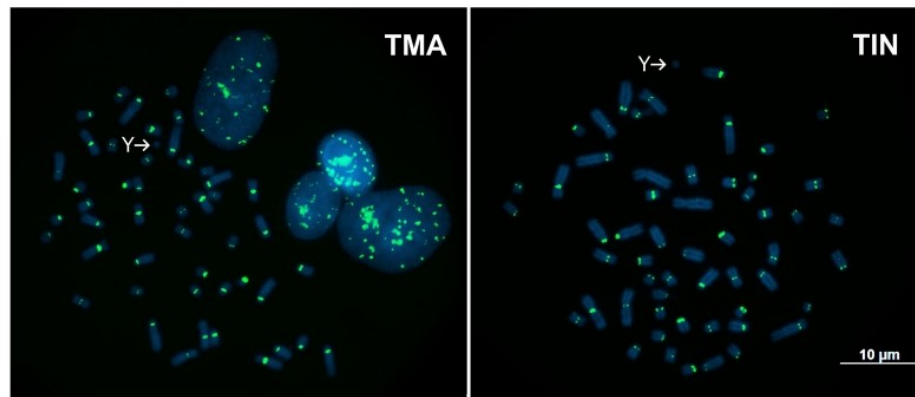
FIGURE 2 | Metaphases of *T. manatus* (TMA) and *T. inunguis* (TIN) after FISH with the TMAsat probe. Y chromosomes without signals of TMAsat are indicated.

TAREAN. The consensus sequence generated by TAREAN (**Supplementary Figure 1**) did not show similarity with any known repetitive DNA from the mammalian RepBase collection (Bao et al., 2015). We named this new satDNA as TMAsat (for *T. manatus* satellite).

The TMAsat MCS from *T. manatus* was generated from an alignment of 66 monomers manually isolated from the assembled reference genome (**Supplementary Figure 2; Supplementary File 1**). It showed monomer length of 687 bp and 54.5% of AT content (**Figures 1A,B**). The dotplot of TMAsat against itself revealed a segment repeated twice inside TMAsat, from position 1 to 332 and 333 to 687 (**Figure 1C**), A pairwise alignment of the two segments of TMAsat, 1–332 and 333–687 bp, showed that they are related but quite divergent, with only 55.8% identity (**Figure 1C**). A detailed investigation in the assembled contigs showed that the TMAsat unit of ~687 bp is organized in higher-order structure, mostly alternating the segments TMAsat1 (1–332) and TMAsat2 (333–687). However, we found one case of TMAsat1 dimer (accession NW_004443969.1 position 56,989–75,023 bp), few cases of TMAsat2 dimer (accessions NW_004444053.1; NW_004444936.1; and NW_004444425.1) and in one contig (accession NW_004443969.1) three, six and 10 tandemly repeated units of TMAsat2.

Genomic Distribution of TMAsat in the Genus *Trichechus*

TMAsat was amplified by PCR from *T. inunguis* genomic DNA, and the resulting PCR products showed a similar monomer length of ~647 bp (**Supplementary Figure 3**). The PCR product was cloned and sequenced in order to confirm that it was indeed homologous to TMAsat. The MCS based on the cloned sequences showed similar AT content and 89.6% of identity (**Figures 1A,B**; **Supplementary Figure 4**). A selected TMAsat cloned sequence was labeled and used as probe in FISH on chromosomes of both *T. manatus* and *T. inunguis*. TMAsat showed centromeric localization in *T. manatus* (2n = 48) and *T. inunguis* (2n = 56), mapping to the centromeres of all chromosomes, except the Y (**Figure 2; Supplementary Figure 5**).

TMAsat localization is compatible with the CBG-banding pattern in both species, which reveals centromeric heterochromatin in all chromosomes (Assis et al., 1998; Gray et al., 2002), with the exception of the Y.

TMAsat in Other Sirenia

Besides *T. manatus*, there are two additional Sirenia species with sequenced genomes available: *D. dugon* and the extinct *H. gigas*, both belonging to the Dugongidae family. A search for TMAsat sequences on the assembled contigs of these species revealed the presence of tandemly repeated TMAsat sequences. TAREAN returned with high confidence one cluster of a putative satDNA with 685 bp length in both species, cluster 8 in *D. dugon* and cluster 3 in *H. gigas*, which contained homologous sequences to TMAsat (**Figures 1A,B**). In *D. dugon*, cluster 8 represents 0.63% of the genome and the MCS generated from the 50 monomers retrieved from the assembled genome is 687 bp long with 54.6% of AT content (**Supplementary Figure 6; Supplementary File 2**). In addition, we found evidence of other HOR configurations rather than alternating TMAsat1 and TMAsat2 in *D. dugon*: a dimer of TMAsat1 (BMBL01107524.1 and BMBL01079760.1), a dimer of TMAsat2 (BMBL01112453.1 and BMBL01093845.1), four (BMBL01013125.1), five (BMBL01107524.1) and six (BMBL01055248.1) tandemly repeated copies of TMAsat2.

In *H. gigas*, cluster 3 comprises 3.4% of the genome and the MCS based on 40 monomers from the reference genome is 687 bp long and has 56.3% of AT content (**Supplementary Figure 7; Supplementary File 3**). In this species, the most frequent TMAsat organization is the alternating segments of TMAsat1 and TMAsat2, and we only found one dimer of TMAsat2 (JACANZ010402190.1).

The sliding window analysis of nucleotide variability of this satDNA in *T. manatus*, *D. dugon*, and *H. gigas* revealed the presence of conserved and variable regions within the monomers (**Figures 3A–C**). However, we did not have access to biological samples of *D. dugon* or *H. gigas* to map TMAsat on their chromosomes. The monomeric TMAsat sequences from *T. manatus*, *T. inunguis*, *D. dugon*, and *H. gigas* were aligned and

FIGURE 3 | Identification of conserved and variable TMAsat segments of **(A)** *T. manatus*, **(B)** *D. dugon* and **(C)** *H. gigas* by sliding window analysis (sliding window of 10 bp and step size of 1 bp). Average nucleotide diversity (Pi) is indicated by the red line, while average nucleotide diversity is indicated by the solid line, and average diversity ± 2 SD is indicated by the dashed line. **(D)** Neighbor-joining tree with TMAsat sequences of *T. manatus* (TMA), *T. inunguis* (TIN), *D. dugon* (DDU), and *H. gigas* (HGI). Bootstrap values generated from 500 replicates.

TABLE 1 | The two putative CENPB box-like motifs identified in the MCS of TMAsat from *T. manatus* (TMA), *T. inunguis* (TIN), *D. dugong* (DDU), and *H. gigas* (HGI).

	Position 196–212 bp	Position 518–534 bp
ECD	NTTCGNNNNANNCGGGN	NTTCGNNNNANNCGGGN
L. africana and *D. novemcinctus*	CTTTGCCGAGAACGGAG	CTTTGCCGAGAACGGAG
T. manatus	CTTTGCATAACAGGGAA	TTTACAGCTTTCCGGGAA
T. inunguis	CTTTGCAWAACAGGAAT	TTTGCAGCATTCCGGGA
D. dugon	CTTTGCATTACAGGGAA	TTTACAGCTTTCCGGGA
H. gigas	CTTTGCATTACAGGGAA	TTTACAGCTTTCCGGGA

Conserved nucleotides in the evolutionarily conserved domain (ECD) are shown in red/highlighted, and conserved nucleotides compared with L. africana and D. novemcinctus motif (Alkan et al., 2007) other than the ECD domain are shown in blue.

used to construct a neighbor-joining tree, which did not reveal any species-specific clustering (**Figure 3D**). We also estimated the inter and intraspecific nucleotide divergence (**Supplementary Table 1**), as well as the average divergence over all sequence pairs ($d = 0.34$). As expected from the Neighbor Joining results, TMAsat intraspecific diversity was not lower than interspecific diversity, except in *T. inunguis* ($d = 0.06$). The low diversity of TMAsat sequences in *T. inunguis* may be due to the low number of sequences used in the analysis and the use of PCR.

CENP-B Box Is Present in TMAsat

The CENP-B box is a 17 bp region conserved among mammalian centromeric satDNAs and known to be the DNA-binding domain for the centromeric protein CENPB. We searched for putative CENP-B box-like motifs within TMAsat MCS from *T. manatus*, *T. inunguis*, *D. dugon*, and *H. gigas*, and found two putative motifs MCS (**Figure 1B**). The first is located in position 196 to 212 bp, matching best with the CENP-B box sequence found in *L. africana* and *D. novemcinctus*. The second putative motif was found in position 518–534 bp. Both motifs display 5–6 identical nucleotides to ECD out of nine in all Sirenia species (**Table 1**). The two putative CENP-B box motifs were present in a conserved segment as indicated by the sliding window analysis of nucleotide variability among satDNA monomers from *T. manatus* and *D. dugon* (**Figures 3A,B**). In *H. gigas*, these motifs spanned both conserved and variable regions of the TMAsat monomer (**Figure 3C**).

The CENP-B box-like motifs found in positions 196–212 bp of TMAsat from *T. manatus* (CTTTGCATAACAGGGAA) and *T. inunguis* (CTTTGCAWAACA-GGAAT) shared 14 out of the 17 nucleotides with each other. In *D. dugon* and in *H. gigas* the CENPB-box-like motif was the same (CTTTGCATTACAGGGAA) and shared 15 out of 17 nucleotides with *T. manatus*. Six bases in *T. manatus*, *D. dugon* and *H. gigas* and five in *T. inunguis*, out of the nine from the ECD were conserved. The second putative motif (position 518–534 bp) showed six out of nine identical bases to the ECD in the four analyzed species. *T. manatus*, *D. dugon*, and *H. gigas* shared an identical second motif (TTTACAGCTTTCCGGGA), whereas *T. inunguis* differed in two nucleotides (TTT**G**CAGC**A**TTCCGGGA).

TMAsat in Other Mammals

We investigated the presence of TMAsat in other mammals using the MCS from *T. manatus* as query in similarity searches against Mammalia sequences in the wgs database from NCBI excluding Sirenia. The total number of returned hits was 13 distributed in four species (**Supplementary Table 2**). With a cut off for query cover equal or greater than 30%, we found four hits in the African elephant (*L. africana*) and four hits in the Asian elephant (*Elephas maximus*). In addition to African and Asian elephants, *Procavia capensis* and *Heterohyrax brucei* appeared in the hits with query covers smaller than 30%. Looking closer into these contigs from *L. africana*, *E. maximus*, *Procavia capensis*, and *Heterohyrax brucei*, we verified few sequences in tandem (maximum of 18) with the repetition unit comprising roughly one TMAsat HOR monomer. The small number of hits found

suggests that this sequence is not a typical satDNA in these taxa, but is instead a repetition related to a transposable element. Indeed, nine out of 10 TMAsat arrays were flanked by LINE-1 in *L. africana*, and 15 out of 16 in *E. maximus*.

DISCUSSION

The TMAsat, reported herein for the first time, was the only putative satDNA found in our analysis, comprising less than 1% of the *T. manatus* genome and mapping to the centromeric regions of all chromosomes, except the Y. The TMAsat could be absent or undetectable by FISH due to low copy number or sequence divergence on the Y chromosome. In *T. inunguis*, we confirmed the presence of TMAsat by PCR and FISH and despite the two species having different karyotypes ($2n = 48$ and $2n = 56$, respectively), TMAsat displayed the same chromosome localization (**Figure 2**). This could be related to the recent ~1.34 million years ago (Mya) divergence time between the species (de Souza et al., 2021).

We also detected the TMAsat in *D. dugon* and *H. gigas* with similar monomer length, comprising 0.63 and 3.4% of the genomes, respectively. The different genome proportion found in *T. manatus* (Illumina HiSeq; 150x genome coverage), *D. dugon* (Illumina Novaseq6000; 64x genome coverage) and *H. gigas* (Illumina NovaSeq; 11x genome coverage) could be due to different genome coverage and/or sequencing platforms used for each species, and may not reflect real interspecific variation. This is especially true in the case of the extinct *H. gigas*, whose DNA source for genome sequencing is a petrous bone from a specimen who probably died during the 1760s (Sharko et al., 2021), and thus the abundance estimates need to be taken with caution.

Although there are slight differences within the MCS from each species, the Neighbor Joining analysis does not indicate intraspecific homogeneous monomers. Only the monomers from *T. inunguis* were grouped together, probably due to the low number of sequences used in the analysis or biased PCR amplification with the selected primers. Nevertheless, we cannot discard a species-specific TMAsat sequence in *T. inunguis* since some mutations are present in all or almost all five sequences and are absent or present in just few monomers outside the species. West Indian and Amazonian manatees present a recent divergence time (de Souza et al., 2021) and an incomplete reproductive isolation (Vianna et al., 2006; Lima et al., 2019), which could contribute to the TMAsat high interspecific homogeneity observed. Overall, the species-specific mutations of the group are probably not yet fixed, despite the ~46.83 Mya estimated split of Trichechidae and Dugongidae, thus lacking species-specific sequences as reflected in the neighbor joining tree (**Figure 3D**), which disagrees with the predictions of concerted evolution. This process, which has been described for many satDNAs, promotes fast sequence homogenization within a species or population, resulting in much higher interspecific than intraspecific differences (Plohl et al., 2012). Although interspecific satDNA sequence conservation is unexpected according to the concerted evolution model, interspecific homogeneity of centromeric satDNAs was observed in other mammalian groups, like in rodents from the *Peromyscus* genus

(Smalec et al., 2019), in four squirrel monkeys (*Saimiri* genus; Valeri et al., 2020) and in two species of two-toed sloths from the genus *Choloepus* (Sena et al., 2020). In all these cases, a possible centromeric function was hypothesized. Moreover, the library model of satDNA evolution relies on the preexistence of a satDNA collection in related species, with the differences observed among the species mostly due to amplification-contraction events of these sequences pool, and does not imply in rapid sequence changes (Plohl et al., 2009). These could be the case of TMAsat evolution if considering the monomer variants as independent amplification-contraction units.

In addition to the centromeric localization in *T. manatus* and *T. inunguis*, we detected the CENPB-box like motif, another centromeric feature, twice in the TMAsat sequences of all four Sirenia species. In *T. manatus* and in *D. dugon*, both putative CENPB-boxes were located in conserved segments of TMAsat. Even though the CENPB-box-like motif found in TMAsat does not present all the nine nucleotides of the ECD, we cannot exclude its functional activity. Among *Peromyscus* species, the CENPB-box-like motifs found in the centromeric satDNA had between four and six conserved bases out of nine ECD nucleotides. It has been suggested that a divergent motif sequence may be required for functional activity in this group (Smalec et al., 2019), which could also be the case for manatees and the dugong. Divergent motif sequences have also been observed in the centromeric satDNAs of the African elephant (*L. africana*), nine-banded armadillo (*D. novemcintus*; Alkan et al., 2011) and in the two-toed sloths of the genus *Choloepus* (Sena et al., 2020).

The only genomes outside Sirenia in which the TMAsat sequence was found were those of the Order Proboscidea (elephants) and Hyracoidea (hyraxes), that together with Sirenia are reunited in Paenungulata, a subgroup of the Superorder Afrotheria (Foley et al., 2016). With only a few hits (with the short arrays mostly flanked by the transposable element L1), the TMAsat sequence is probably not a typical satDNA in these species. TMAsat in Sirenian probably evolved from these ancestral sequences still found in elephants and hyraxes, which could be the basis for both TMAsat1 and TMAsat2.

In the tree sirenians with sequenced genome, the most frequent organization of TMAsat arrays was the alternating TMAsat1 and TMAsat2 form. In the few exceptions, we found more consecutive TMAsat2 units than TMAsat1. Other satDNAs were found organized as a composite of two related units, mostly in the alternating form as TMAsat. This is the case of S1a-S1b in European brown frogs (Feliciello et al., 2006) and Tcast1a-Tcast1b in the red flour beetle *Tribolium castaneum* (Feliciello et al., 2011, 2014), in which the rolling circle amplification followed by substitutions by homologous recombination were proposed to explain the origin of the composite a-b arrays.

The sequenced genomes we used were generated from short reads (average 100–150 bp) that do not cover the total length of the monomeric unit of TMAsat, resulting in an assembly that may not represent well the long satDNA arrays. Further analyses with Southern blot and dot blot experiments as well as long-reads sequencing may help clarify the overall organization of repeats in the genome and within the long satDNA arrays. As an example, Vondrak et al. (2020) using ultra-long nanopore reads found nine

out of 11 putative satDNA sequences derived from short tandem arrays located within LTR-retrotransposons that occasionally expanded in length, and just two organized in long arrays typical of satDNA. In addition, the long-reads sequencing approach proved a valuable contribution in determining the origin of the satDNAs. Several satDNAs from plants and animals derived from tandem amplification of internal segments of TEs (Dias et al., 2015; Meštrović et al., 2015; Vondrak et al., 2020), as was the case of TMAsat described herein, that could be L1 related.

In conclusion, we reported for the first time the centromeric satDNA in the West Indian manatee, which seems to be present across Sirenia, a group with all extant species under threat of extinction. TMAsat monomers from *T. manatus*, *T. inunguis*, *D. dugon*, and *H. gigas* lack species-specific sequences, contradicting the predictions of concerted evolution. The TMAsat-like ancestral sequence is present in other Paenungulata, such as elephants and hyraxes, suggesting that TMAsat suffered an expansion within Sirenia less than ~69 Mya (Foley et al., 2016; de Souza et al., 2021), after the divergence of Sirenia from Proboscidea and Hyracoidea.

DATA AVAILABILITY STATEMENT

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: https://www.ncbi.nlm.nih.gov/genbank/, MW272776–MW272780.

ETHICS STATEMENT

The animal study was reviewed and approved by SISBio/ICMBio permit 60829-2.

AUTHOR CONTRIBUTIONS

MV and GD conceived and designed the experiments, analyzed the data, and contributed to writing – original draft preparation. MV and AE performed the experiments. CM, YY-Y, and IS obtained the materials for molecular and cytological analyses. MV, GD, AE, GK, and MS contributed to writing – review and editing. GK and MS contributed to supervision and project administration. MS contributed to funding acquisition. All authors contributed to the article and approved the submitted version.

FUNDING

ACKNOWLEDGMENTS

REFERENCES

Alkan, C., Cardone, M. F., Catacchio, C. R., Antonacci, F., O'Brien, S. J., Ryder, O. A., et al. (2011). Genome-wide characterization of centromeric satellites from multiple mammalian genomes. *Genome Res.* 21, 137–145. doi: 10.1101/gr.111278.110

Alkan, C., Ventura, M., Archidiacono, N., Rocchi, M., Sahinalp, S. C., and Eichler, E. E. (2007). Organization and evolution of primate centromeric DNA from whole-genome shotgun sequence data. *PLoS Comput. Biol.* 3, 1807–1818. doi: 10.1371/journal.pcbi.0030181

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Assis, M. F. L., Best, R. C., Barros, R. M. S., and Yonenaga-Yassuda, Y. (1998). Cytogenetic study of *Trichechus inunguis* (Amazonian manatee). *Braz. J. Genet.* 11, 41–50.

Bao, W., Kojima, K. K., and Kohany, O. (2015). Repbase update, a database of repetitive elements in eukaryotic genomes. *Mob. DNA* 6:11. doi: 10.1186/s13100-015-0041-9

Barros, H. M. D. R., Meirelles, A. C. O., Luna, F. O., Marmontel, M., Cordeiro-Estrela, P., Santos, N., et al. (2017). Cranial and chromosomal geographic variation in manatees (Mammalia: Sirenia: Trichechidae) with the description of the Antillean manatee karyotype in Brazil. *J. Zool. Syst. Evol. Res.* 55, 73–87. doi: 10.1111/jzs.12153

Biscotti, M. A., Olmo, E., and Heslop-Harrison, J. S. (2015). Repetitive DNA in eukaryotic genomes. *Chromosom. Res.* 23, 415–420. doi: 10.1007/s10577-015-9499-z

Bonvicino, C. R., Viana, M. C., de Oliveira, E. H. C., Emin, R., Silva Junior, J. D. S., de Sousa, M. E. M., et al. (2020). Distribution of south American manatees, *Trichechus manatus* Linnaeus, 1758 and *T. inunguis* (Natterer, 1883) (Sirenia: Trichechidae). *Bol. Mus. Para. Emílio Goeldi, Ciênc. Nat.* 15, 573–599. doi: 10.46357/bcnaturais.v15i3.246

de Souza, É. M. S., Freitas, L., da Silva Ramos, E. K., Selleghin-Veiga, G., Rachid-Ribeiro, M. C., Silva, F. A., et al. (2021). The evolutionary history of manatees told by their mitogenomes. *Sci. Rep.* 11:3564. doi: 10.1038/s41598-021-82390-2

Deutsch, C. J., Self-Sulivan, C., and Mignucci-Giannoni, A. (2008). Data from: Trichechus manatus, West Indian Manatee. *IUCN Red List Threat. Species*, e.T22103A9356917. doi: 10.2305/IUCN.UK.2008.RLTS.T22103A9356917.en

Dias, G. B., Heringer, P., Svartman, M., and Kuhn, G. C. S. (2015). Helitrons shaping the genomic architecture of drosophila: enrichment of DINE-TR1 in α- and β-heterochromatin, satellite DNA emergence, and piRNA expression. *Chromosom. Res.* 23, 597–613. doi: 10.1007/s10577-015-9480-x

Domning, D. P. (2018). "Sirenian evolution," in *Encyclopedia of Marine Mammals*. eds. B. Würsig, J. G. M. Thewissen and K. M. Kovacs (London, United Kingdom: Academic Press), 856–859.

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

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

Feliciello, I., Akrap, I., Brajkovi, J., Zlatar, I., and Ugarkovic, D. (2014). Satellite DNA as a driver of population divergence in the red flour beetle Tribolium castaneum. *Genome Biol. Evol.* 7, 228–239. doi: 10.1093/gbe/evu280

Feliciello, I., Chinali, G., and Ugarković, D. (2011). Structure and population dynamics of the major satellite DNA in the red flour beetle Tribolium castaneum. *Genetica* 139, 999–1008. doi: 10.1007/s10709-011-9601-1

Feliciello, I., Picariello, O., and Chinali, G. (2006). Intra-specific variability and unusual organization of the repetitive units in a satellite DNA from Rana dalmatina: molecular evidence of a new mechanism of DNA repair acting on satellite DNA. *Gene* 383, 81–92. doi: 10.1016/j.gene.2006.07.016

Foley, N. M., Springer, M. S., and Teeling, E. C. (2016). Mammal madness: is the mammal tree of life not yet resolved? *Philos. Trans. R. Soc. B Biol. Sci.* 371:20150140. doi: 10.1098/rstb.2015.0140

Fry, K., and Salser, W. (1977). Nucleotide sequences of HS-α satellite DNA from kangaroo rat dipodomys ordii and characterization of similar sequences in other rodents. *Cell* 12, 1069–1084. doi: 10.1016/0092-8674(77)90170-2

Gray, B. A., Zori, R. T., McGuire, P. M., and Bonde, R. K. (2002). A first generation cytogenetic ideogram for the Florida manatee (Trichechus manatus latirostris) based on multiple chromosome banding techniques. *Hereditas* 137, 215–223. doi: 10.1034/j.1601-5223.2002.01657.x

Kasai, F., O'Brien, P. C. M., and Ferguson-Smith, M. A. (2013). Afrotheria genome; overestimation of genome size and distinct chromosome GC content revealed by flow karyotyping. *Genomics* 102, 468–471. doi: 10.1016/j.ygeno.2013.09.002

Kasinathan, S., and Henikoff, S. (2018). Non-B-form DNA is enriched at centromeres. *Mol. Biol. Evol.* 35, 949–962. doi: 10.1093/molbev/msy010

Keith, L. D. (2015). Data from: Trichechus senegalensis (African Manatee). *IUCN Red List Threat. Species*, e.T22104A97168578. doi: 10.2305/IUCN.UK.2015-4.RLTS.T22104A81904980.en

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

Letunic, I., and Bork, P. (2019). Interactive tree of life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi: 10.1093/nar/gkz239

Lima, C. S., Magalhães, R. F., Marmontel, M., Meirelles, A. C., Carvalho, V. L., Lavergne, A., et al. (2019). A hybrid swarm of manatees along the guianas coastline, a peculiar environment under the influence of the Amazon River plume. *An. Acad. Bras. Cienc.* 91:e20190325. doi: 10.1590/0001-3765201920190325

Lima, C. S., Magalhães, R. F., and Santos, F. R. (2021). Conservation issues using discordant taxonomic and evolutionary units: a case study of the American manatee (Trichechus manatus, Sirenia). *Wildl. Res.* 48:385. doi: 10.1071/WR20197

Luna, F. O., Beaver, C. E., Nourisson, C., Bonde, R. K., Attademo, F. L. N., Miranda, A. V., et al. (2021). Genetic connectivity of the west Indian manatee in the southern range and limited evidence of hybridization with Amazonian manatees. *Front. Mar. Sci.* 7, 1–15. doi: 10.3389/fmars.2020.574455

Marmontel, M., de Souza, D., and Kendall, S. (2016). Data from: Trichechus inunguis. *IUCN Red List Threat. Species*, e.T22102A43793736. doi: 10.2305/IUCN.UK.2016-2.RLTS.T22102A43793736.en

Marsh, H., and Sobtzick, S. (2019). Data from: Dugong Dugon (amended version of 2015 assessment). *IUCN Red List Threat. Species* 2019, e.T22104A97168578. doi: 10.2305/IUCN.UK.2015-4.RLTS.T22104A81904980.en

Masumoto, H., Nakano, M., and Ohzeki, J. I. (2004). The role of CENP-B and α-satellite DNA: de novo assembly and epigenetic maintenance of human centromeres. *Chromosom. Res.* 12, 543–556. doi: 10.1023/B:CHRO.0000036593.72788.99

Meštrović, N., Mravinac, B., Pavlek, M., Vojvoda-Zeljko, T., Šatović, E., and Plohl, M. (2015). Structural and functional liaisons between transposable elements and satellite DNAs. *Chromosom. Res.* 23, 583–596. doi: 10.1007/s10577-015-9483-7

Meštrović, N., Plohl, M., Mravinac, B., and Ugarković, D. (1998). Evolution of satellite DNAs from the genus palorus-experimental evidence for the "library" hypothesis. *Mol. Biol. Evol.* 15, 1062–1068. doi: 10.1093/oxfordjournals.molbev.a026005

Muro, Y., Masumoto, H., Yoda, K., Nozaki, N., Ohashi, M., and Okazaki, T. (1992). Centromere protein B assembles human centromeric α-satellite DNA at the 17-bp sequence, CENP-B box. *J. Cell Biol.* 116, 585–596. doi: 10.1083/jcb.116.3.585

Novák, P., Neumann, P., and Macas, J. (2020). Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nat. Protoc.* 15, 3745–3776. doi: 10.1038/s41596-020-0400-y

SUPPLEMENTARY MATERIAL

Novák, P., Robledillo, L. Á., Koblížková, A., Vrbová, I., Neumann, P., and Macas, J. (2017). TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Res.* 45:e111. doi: 10.1093/nar/gkx257

Plohl, M., Meštrović, N., and Mravinac, B. (2012). "Satellite DNA evolution," in *Repetitive DNA. Genome Dyn. Vol. 7.* ed. M. A. Garrido-Ramos (Basel, Karger), 126–152.

Plohl, M., Petrović, V., Luchetti, A., Ricci, A., Šatović, E., Passamonti, M., et al. (2009). Long-term conservation vs high sequence divergence: the case of an extraordinarily old satellite DNA in bivalve mollusks. *Hereditas* 104, 543–551. doi: 10.1038/hdy.2009.141

Rozas, J., Ferrer-Mata, A., Sanchez-DelBarrio, J. C., Guirao-Rico, S., Librado, P., Ramos-Onsins, S. E., et al. (2017). DnaSP 6: DNA sequence polymorphism analysis of large data sets. *Mol. Biol. Evol.* 34, 3299–3302. doi: 10.1093/molbev/msx248

Sena, R. S., Heringer, P., Valeri, M. P., Pereira, V. S., Kuhn, G. C. S., and Svartman, M. (2020). Identification and characterization of satellite DNAs in two-toed sloths of the genus Choloepus (Megalonychidae, Xenarthra). *Sci. Rep.* 10, 1–11. doi: 10.1038/s41598-020-76199-8

Shapiro, J. A., and von Sternberg, R. (2005). Why repetitive DNA is essential to genome function. *Biol. Rev.* 80, 227–250. doi: 10.1017/S1464793104006657

Sharko, F. S., Boulygina, E. S., Tsygankova, S. V., Slobodova, N. V., Alekseev, D. A., Krasivskaya, A. A., et al. (2021). Steller's sea cow genome suggests this species began going extinct before the arrival of Paleolithic humans. *Nat. Commun.* 12, 1–8. doi: 10.1038/s41467-021-22567-5

Shatskikh, A. S., Kotov, A. A., Adashev, V. E., Bazylev, S. S., and Olenina, L. V. (2020). Functional significance of satellite DNAs: insights from drosophila. *Front. Cell Dev. Biol.* 8:312. doi: 10.3389/fcell.2020.00312

Smalec, B. M., Heider, T. N., Flynn, B. L., and O'Neill, R. J. (2019). A centromere satellite concomitant with extensive karyotypic diversity across the Peromyscus genus defies predictions of molecular drive. *Chromosom. Res.* 27, 237–252. doi: 10.1007/s10577-019-09605-1

Stanyon, R., and Galleni, L. (1991). A rapid fibroblast culture technique for high resolution karyotypes. *Bolletino di Zool.* 58, 81–83. doi: 10.1080/11250009109355732

Sujiwattanarat, P., Thapana, W., Srikulnath, K., Hirai, Y., Hirai, H., and Koga, A. (2015). Higher-order repeat structure in alpha satellite DNA occurs in New World monkeys and is not confined to hominoids. *Sci. Rep.* 5, 1–10. doi: 10.1038/srep10315

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

Valeri, M. P., Dias, G. B., Moreira, C. N., Yonenaga-Yassuda, Y., Stanyon, R., Silva Kuhn, G. C., et al. (2020). Characterization of satellite DNAs in squirrel monkeys genus Saimiri (Cebidae, Platyrrhini). *Sci. Rep.* 10, 1–11. doi: 10.1038/s41598-020-64620-1

Vianna, J. A., Bonde, R. K., Caballero, S., Giraldo, J. P., Lima, R. P., Clark, A., et al. (2006). Phylogeography, phylogeny and hybridization in trichechid sirenians: implications for manatee conservation. *Mol. Ecol.* 15, 433–447. doi: 10.1111/j.1365-294X.2005.02771.x

Vlahović, I., Glunčić, M., Rosandić, M., Ugarković, Đ., and Paar, V. (2016). Regular higher order repeat structures in beetle *Tribolium castaneum* genomex. *Hereditas* 104, 543–551. doi: 10.1093/gbe/evw174

Vondrak, T., Ávila Robledillo, L., Novák, P., Koblížková, A., Neumann, P., and Macas, J. (2020). Characterization of repeat arrays in ultra-long nanopore reads reveals frequent origin of satellite DNA from retrotransposon-derived tandem repeats. *Plant J.* 101, 484–500. doi: 10.1111/tpj.14546

Supplementary Material

```
>TMAsat
ATTCACCACATGAAAGAGAAAGCAAAATACAGCGTTCTCCTTCATGCTGCTTAACTGAATCTACTGAAGTTT
TCAGTTTTGTTAGAAGAAAGTGTCCGCCCACCACAAGCTGCTCTCTCAAATGTTCTCTCAGTGGGAGAAG
GCAGTAGAGTGAGGGCCTAGCAGGAATTTGAATCCCCTGGTTTTCAAAGGGAACTTTGCATATGAGGAAT
GAAGGTATTGCTAGCTTACATAGCCAAACAGGCAGATTGTTTGCCTTCAGCTAGAGAACAGCAGAAGAGG
TTCAGTTTTGCTAACTAAAGCTTGTGTGCTTAACACGAGCATGTAGCTTCTTCCCTGCCTGAAAGAGAAAG
CAAGAGCTTGCTGCTCTCAGTCTGGGGATGAAGGGAAAAAGATTCACTTTCTGGGTTTGAAAATCCCTCT
GTGTTACATGCCTAAGAAAGTTGCCTGTGCATTGTCTACCTGGATAAAGTGTATGACTTTTCTCCCTTTCAA
GTAGTTGAAGGCGTACAGGGTTTCAGCATTAAGGGAGGCATTCTGTGCTGCGCGGCTGAGAATACCGCTT
TACAGACGGTATCGCTACAGGAGAAGAGAAACCGCGTTTCTGCCTTTCTGGAACACGAATGAGGCAAAAC
CGTAGTTGTGCTGTGTAAACACTTTGTCAGCAGCAAGTGTAAGTTCCGC
```

Supplementary Figure 1. TMAsat consensus sequence, in fasta format, generated by TAREAN.

Supplementary Figure 2. Alignment of 66 monomers manually isolated from the *Trichechus manatus* assembled reference genome (accession GCA_000243295.1). Part one of eight.

Supplementary Figure 2. Alignment of 66 monomers manually isolated from the *Trichechus manatus* assembled reference genome (accession GCA_000243295.1). Part two of eight.
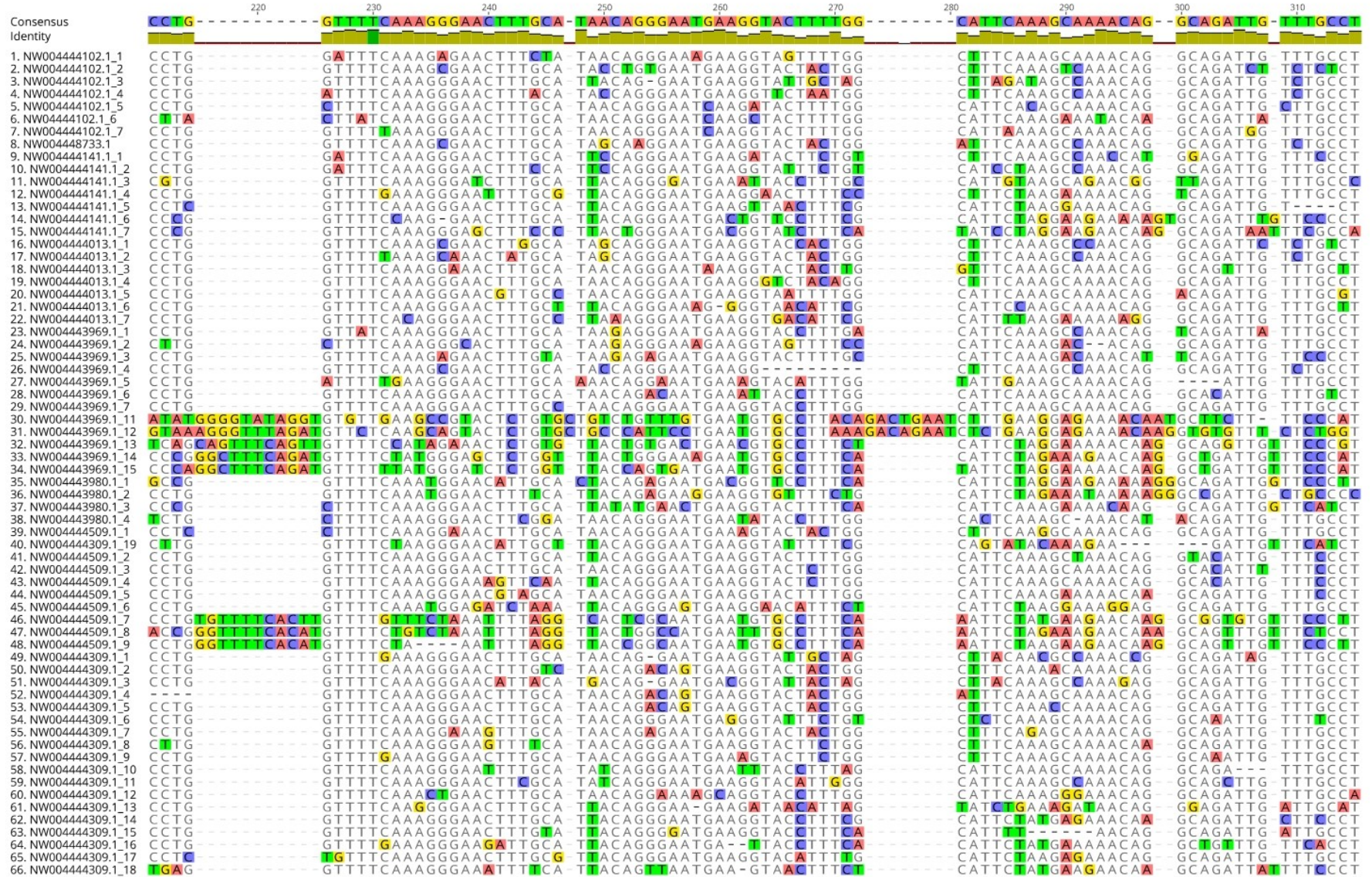
Supplementary Figure 2. Alignment of 66 monomers manually isolated from the *Trichechus manatus* assembled reference genome (accession GCA_000243295.1). Part three of eight.
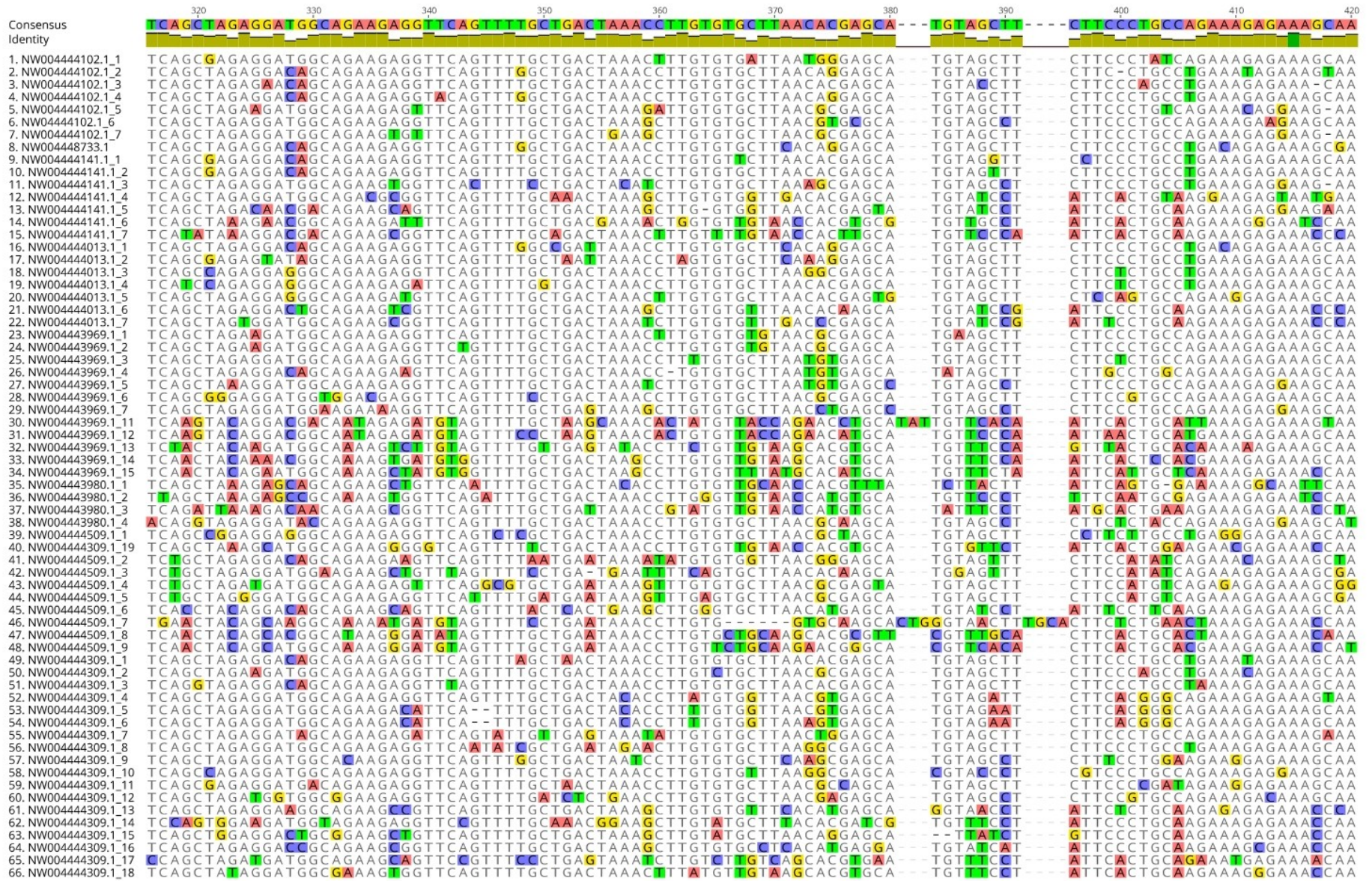
Supplementary Figure 2. Alignment of 66 monomers manually isolated from the *Trichechus manatus* assembled reference genome (accession GCA_000243295.1). Part four of eight.
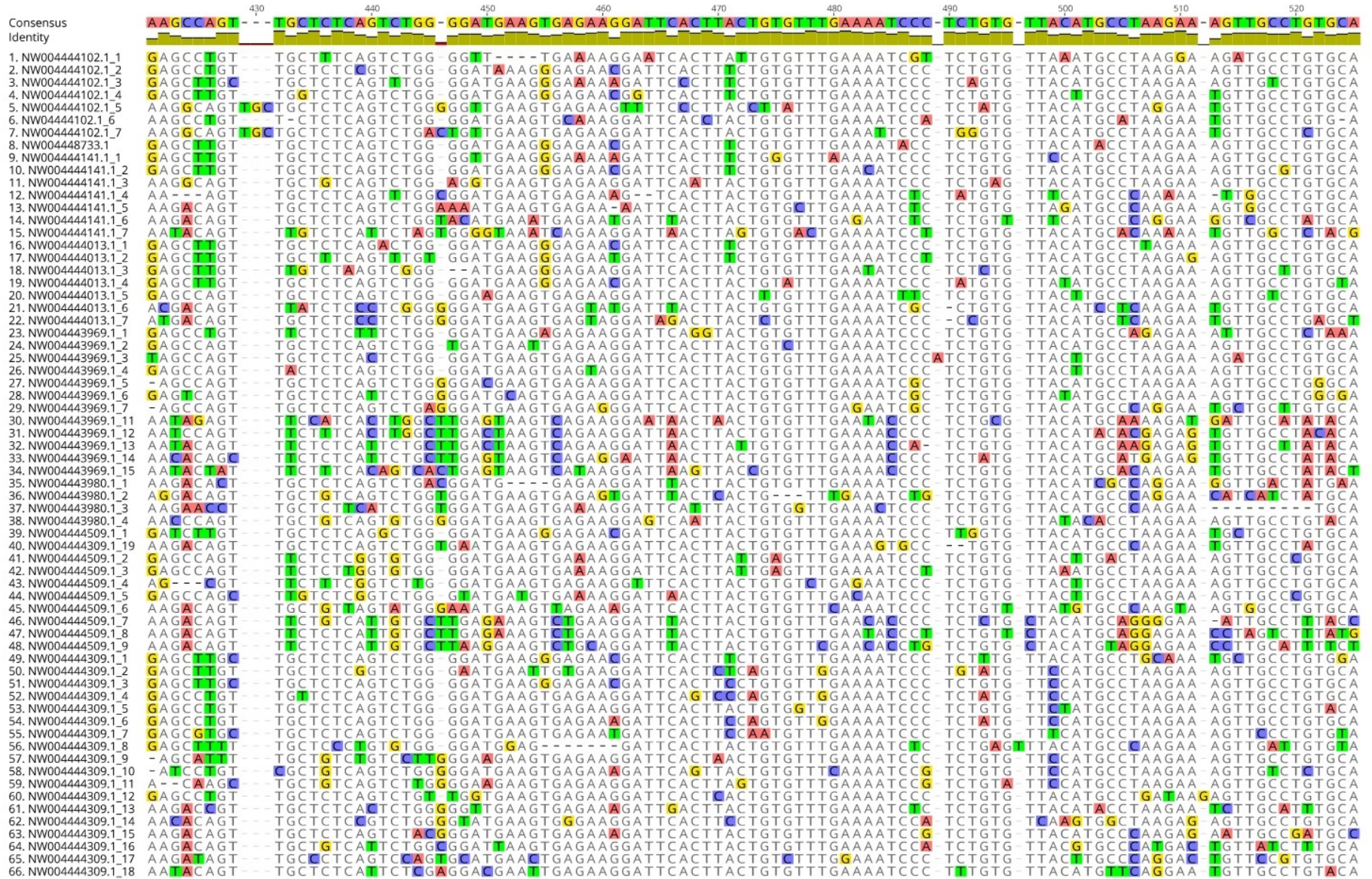
Supplementary Figure 2. Alignment of 66 monomers manually isolated from the *Trichechus manatus* assembled reference genome (accession GCA_000243295.1). Part five of eight.

Supplementary Figure 2. Alignment of 66 monomers manually isolated from the *Trichechus manatus* assembled reference genome (accession GCA_000243295.1). Part six of eight.
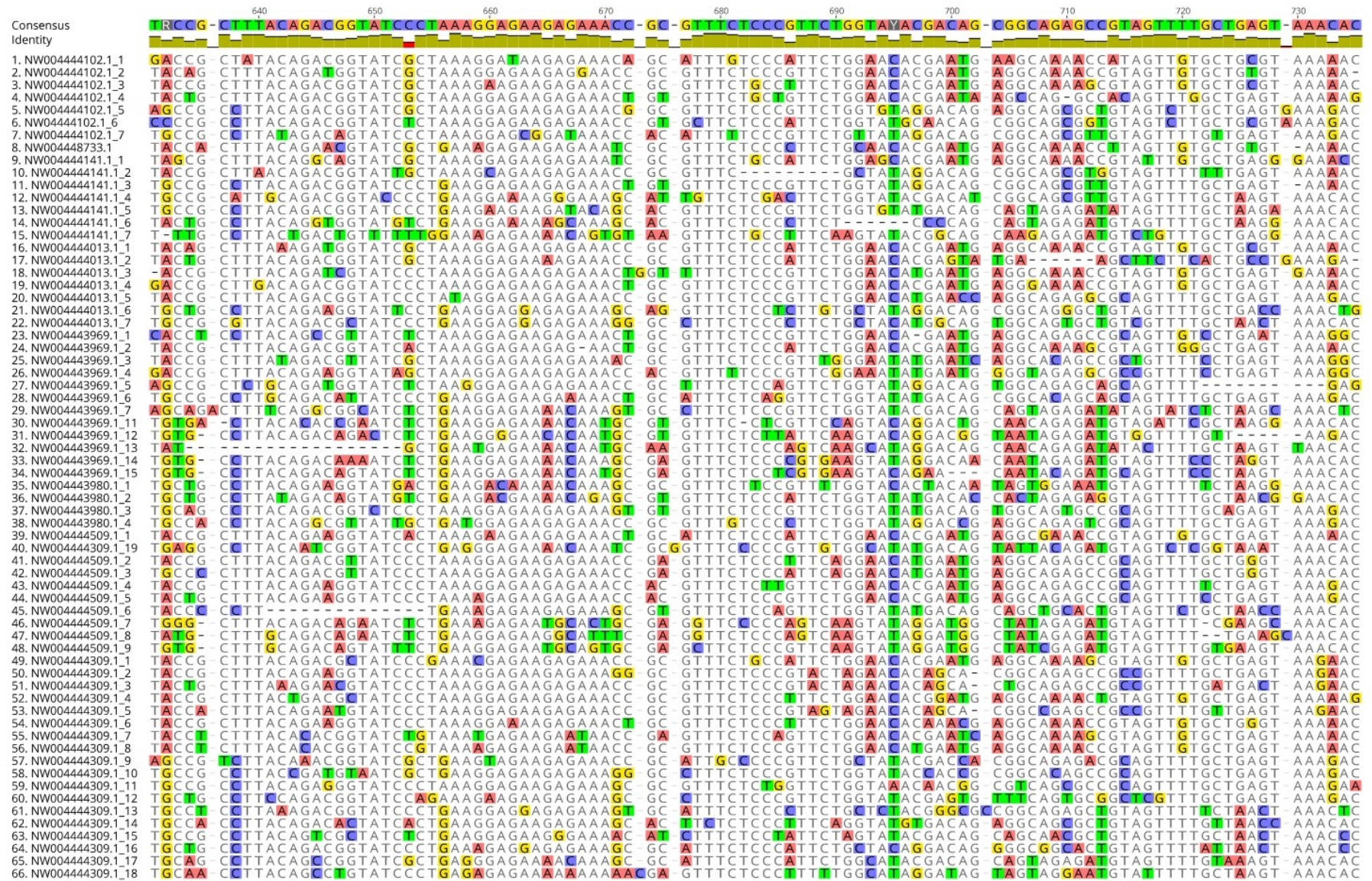
Supplementary Figure 2. Alignment of 66 monomers manually isolated from the *Trichechus manatus* assembled reference genome (accession GCA_000243295.1). Part seven of eight.
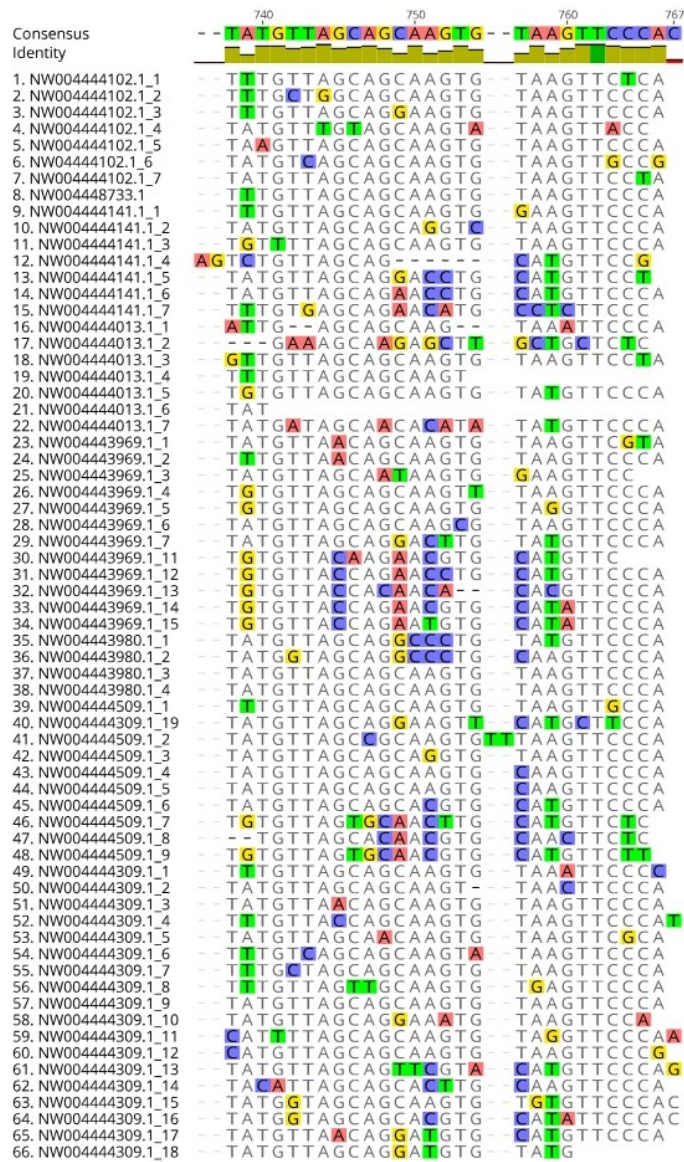
Supplementary Figure 2. Alignment of 66 monomers manually isolated from the *Trichechus manatus* assembled reference genome (accession GCA_000243295.1). Part eight of eight.
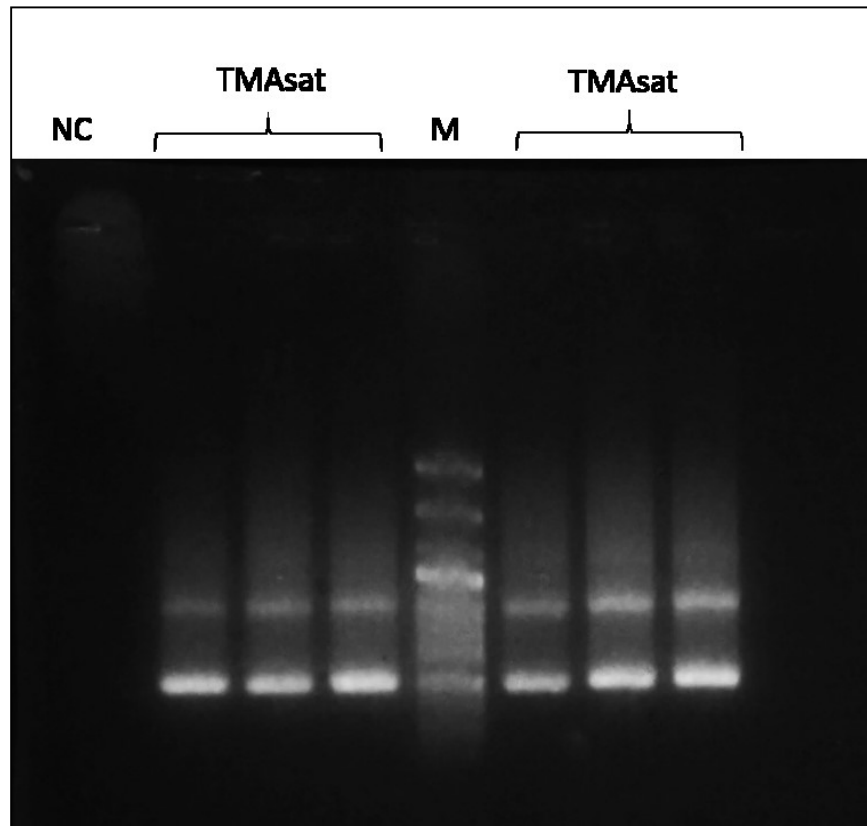
Supplementary Figure 3. Agarose gel (1%) showing the PCR products of *Trichechus inunguis* using the TMAsat primers. M - DNA size marker of 100 bp ladder; NC – Negative control (no DNA).
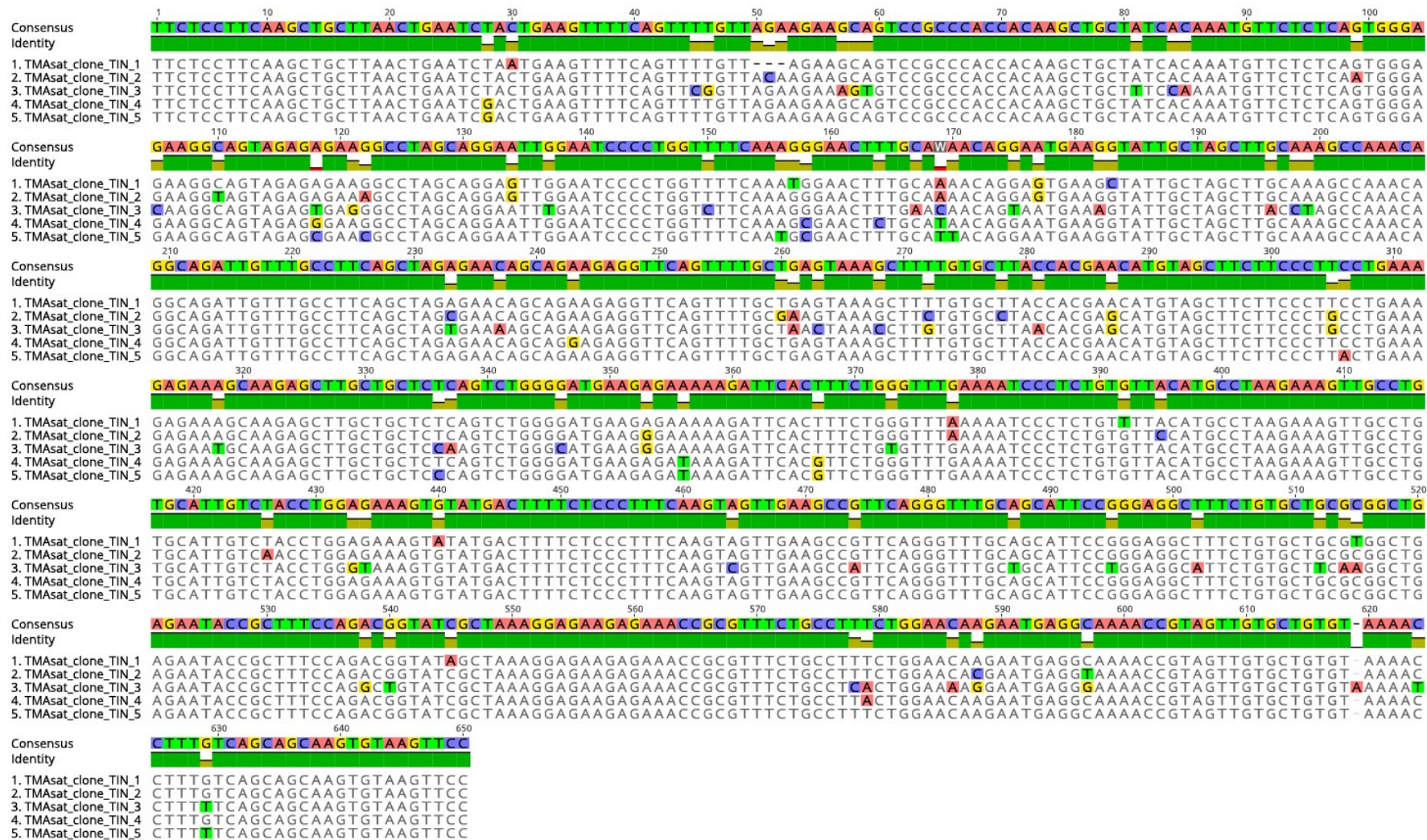
Supplementary Figure 4. Alignment of five TMAsat monomers cloned and sequenced from the *Trichechus inunguis* genome (accession MW272776- MW272780).
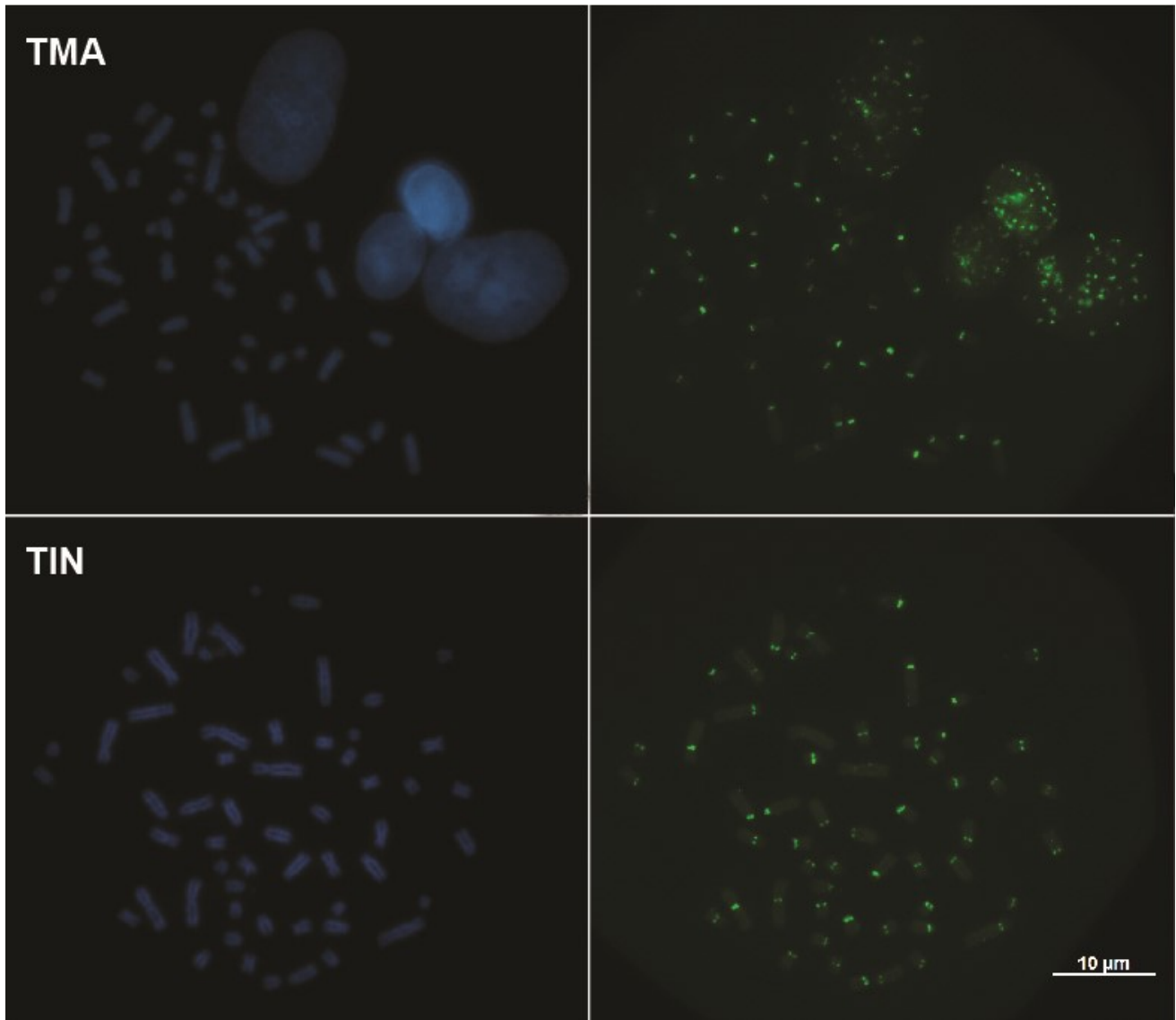
Supplementary Figure 5. Metaphases of *T. manatus* (TMA) and *T. inunguis* (TIN) after FISH with the TMAsat probe. Channels are presented separately, DAPI on left in blue and TMAsat probe on right in green.
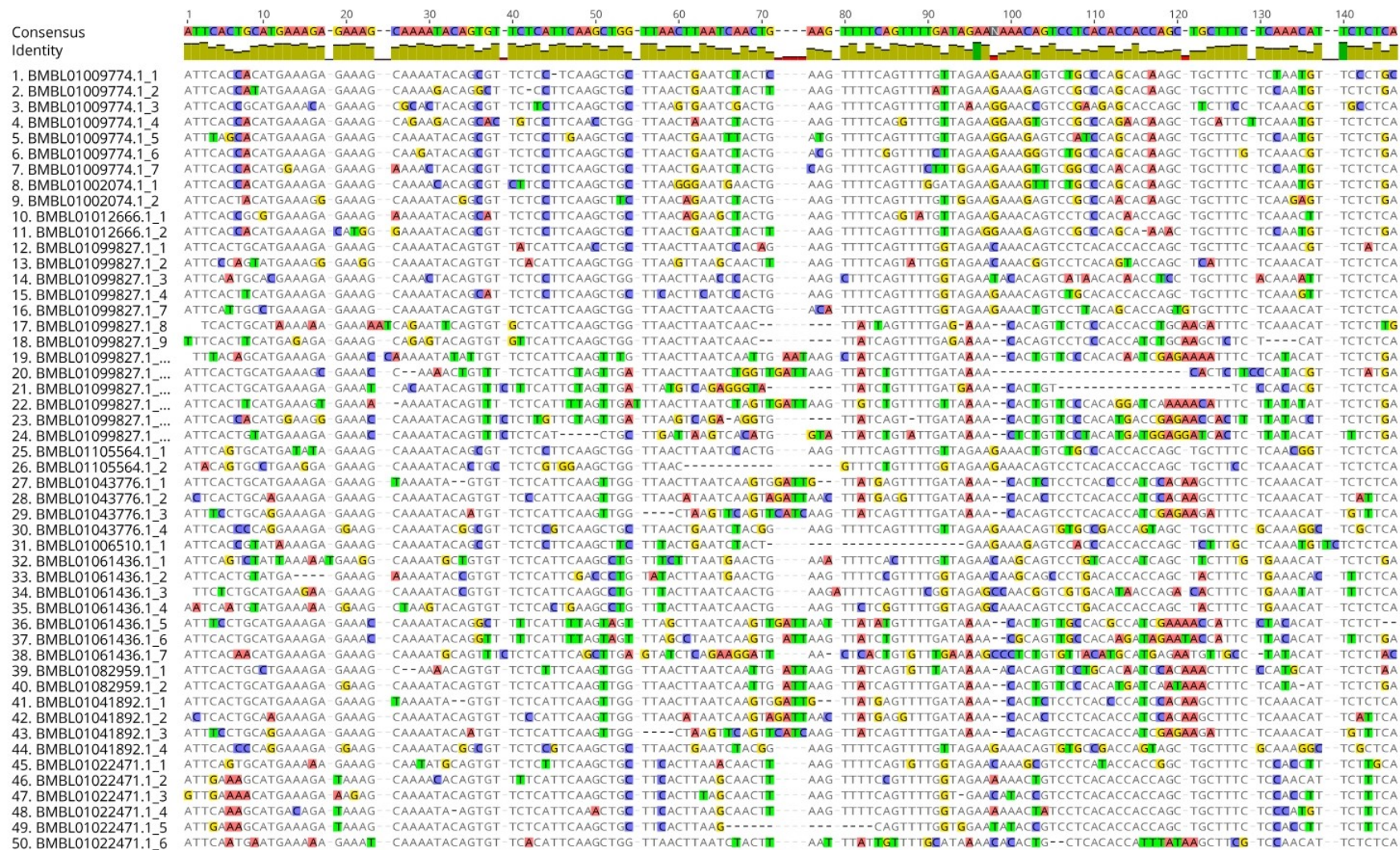
Supplementary Figure 6. Alignment of 50 monomers manually isolated from the *Dugong dugon* assembled reference genome (accession GCA_015147995.1). Part one of six.
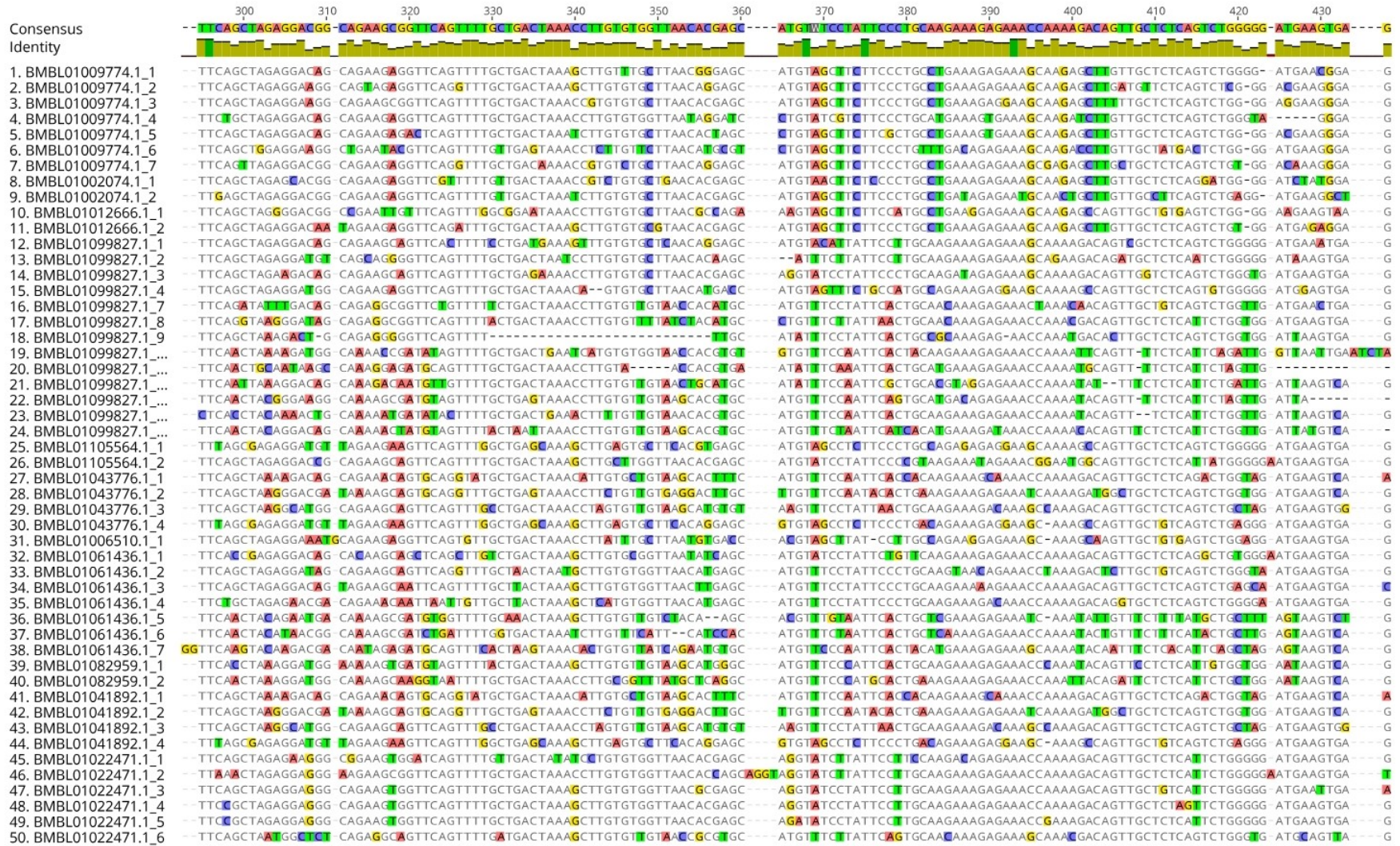
Supplementary Figure 6. Alignment of 50 monomers manually isolated from the *Dugong dugon* assembled reference genome (accession GCA_015147995.1). Part two of six.

Supplementary Figure 6. Alignment of 50 monomers manually isolated from the *Dugong dugon* assembled reference genome (accession GCA_015147995.1). Part three of six.
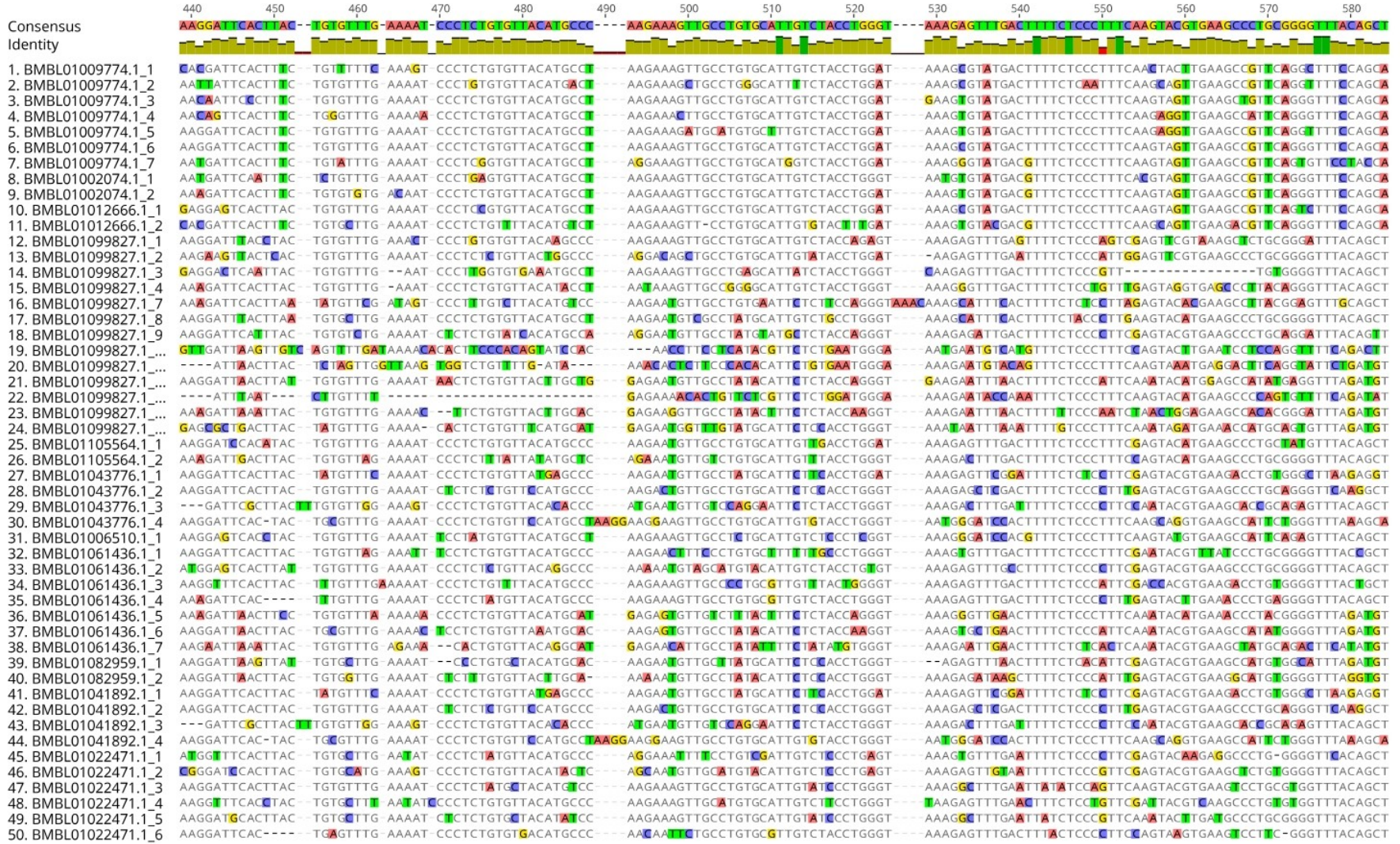
Supplementary Figure 6. Alignment of 50 monomers manually isolated from the *Dugong dugon* assembled reference genome (accession GCA_015147995.1). Part four of six.
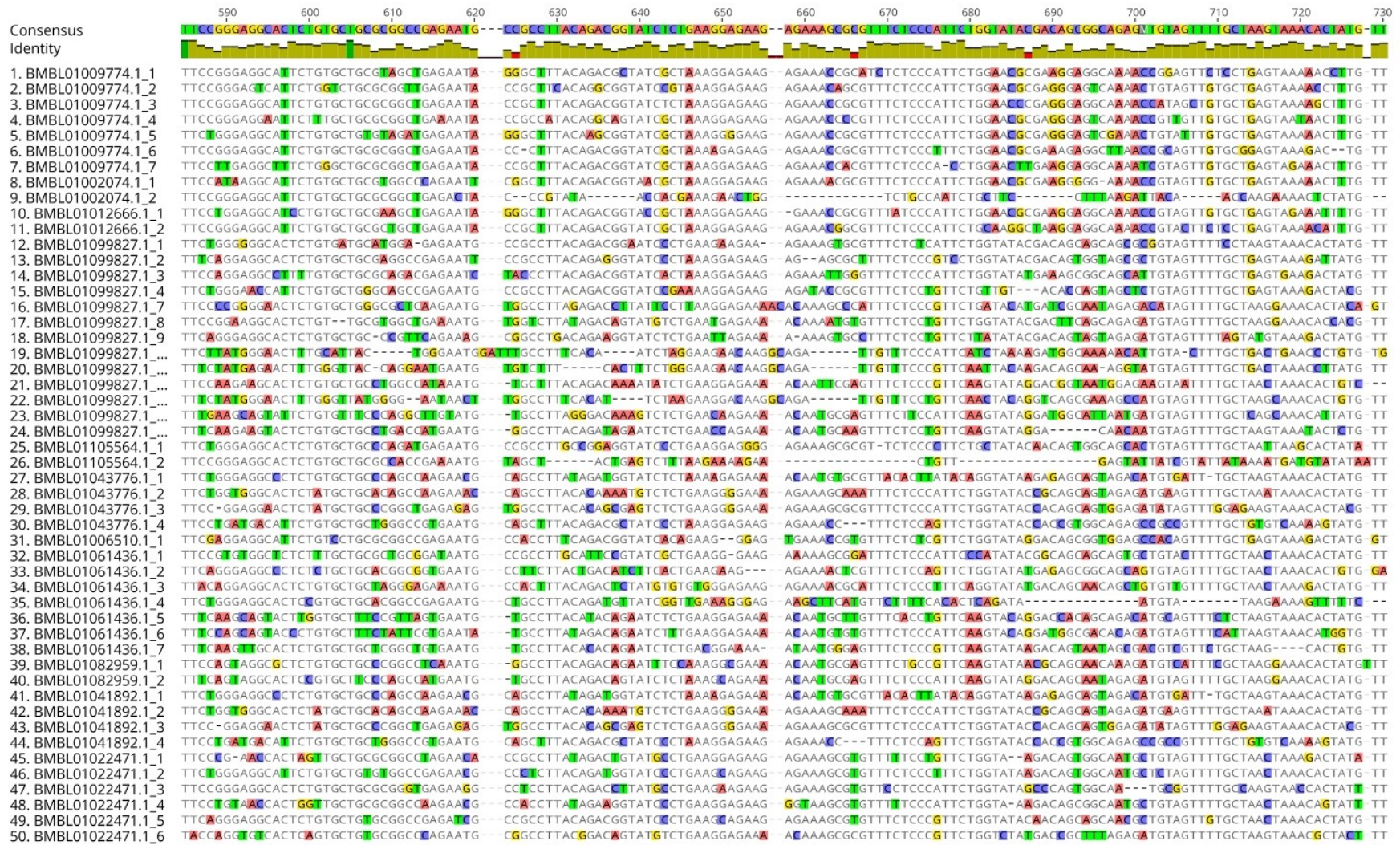
Supplementary Figure 6. Alignment of 50 monomers manually isolated from the *Dugong dugon* assembled reference genome (accession GCA_015147995.1). Part five of six.
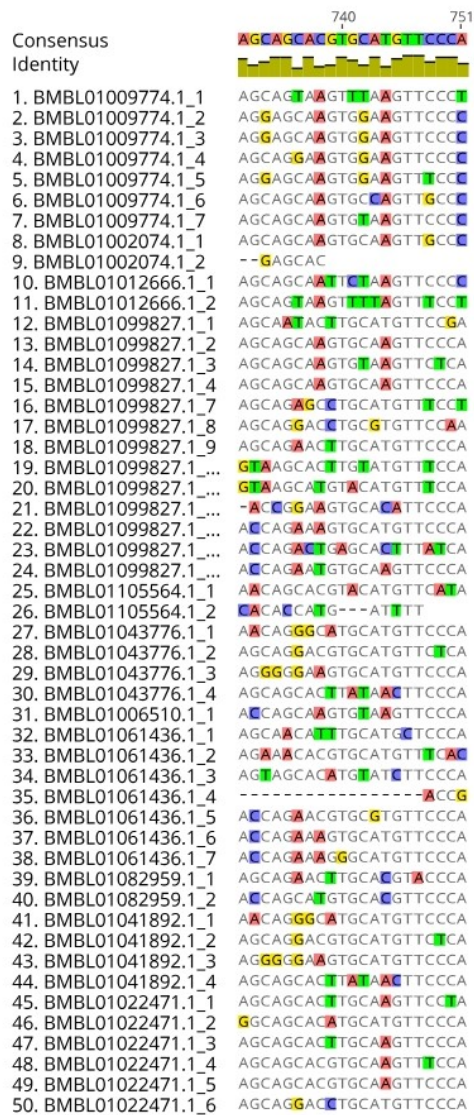
Supplementary Figure 6. Alignment of 50 monomers manually isolated from the *Dugong dugon* assembled reference genome (accession GCA_015147995.1). Part six of six.
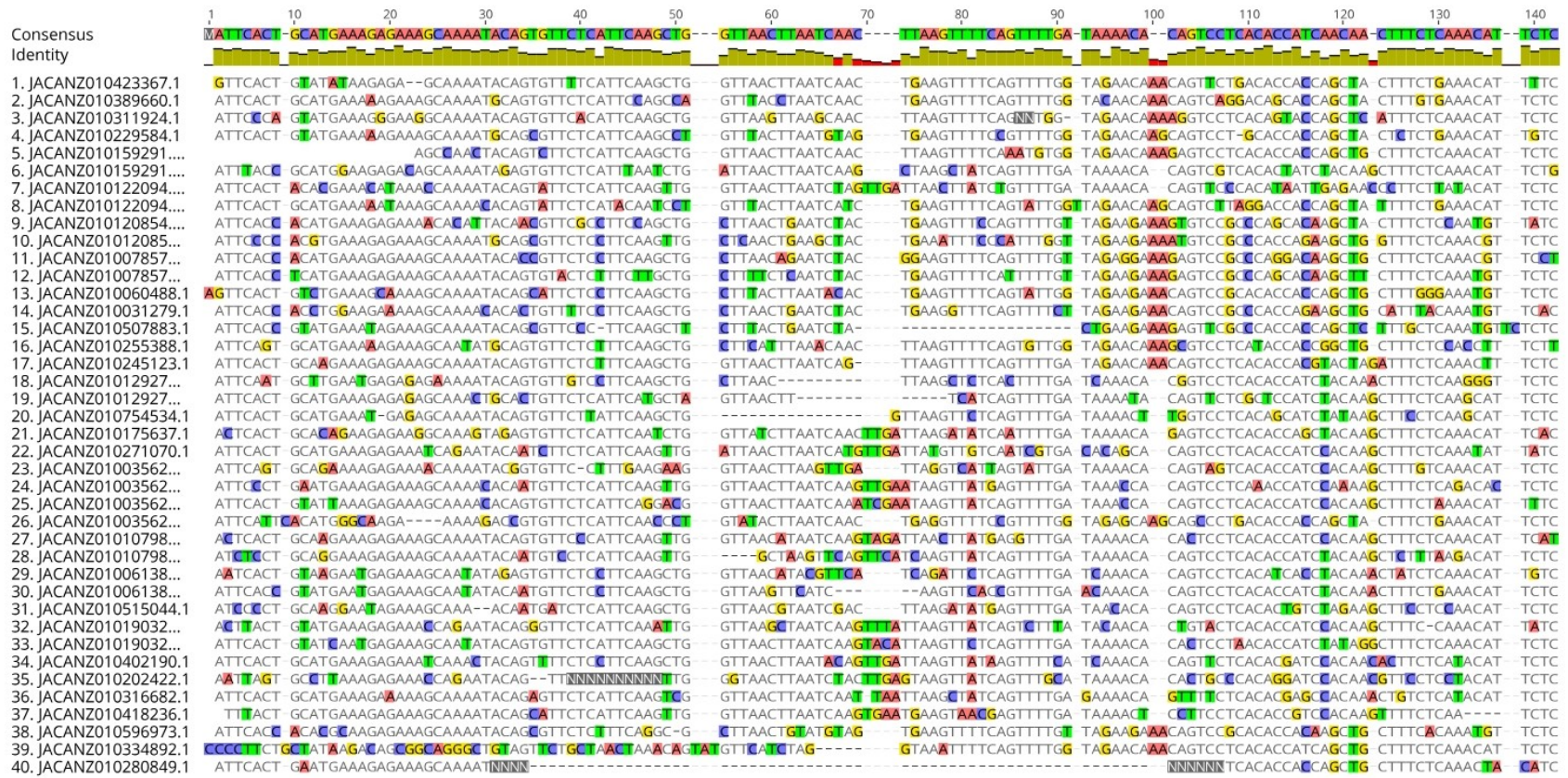
Supplementary Figure 7. Alignment of 40 monomers manually isolated from the *Hydrodamalis gigas* assembled reference genome (accession GCA_013391785.1). Part one of six.
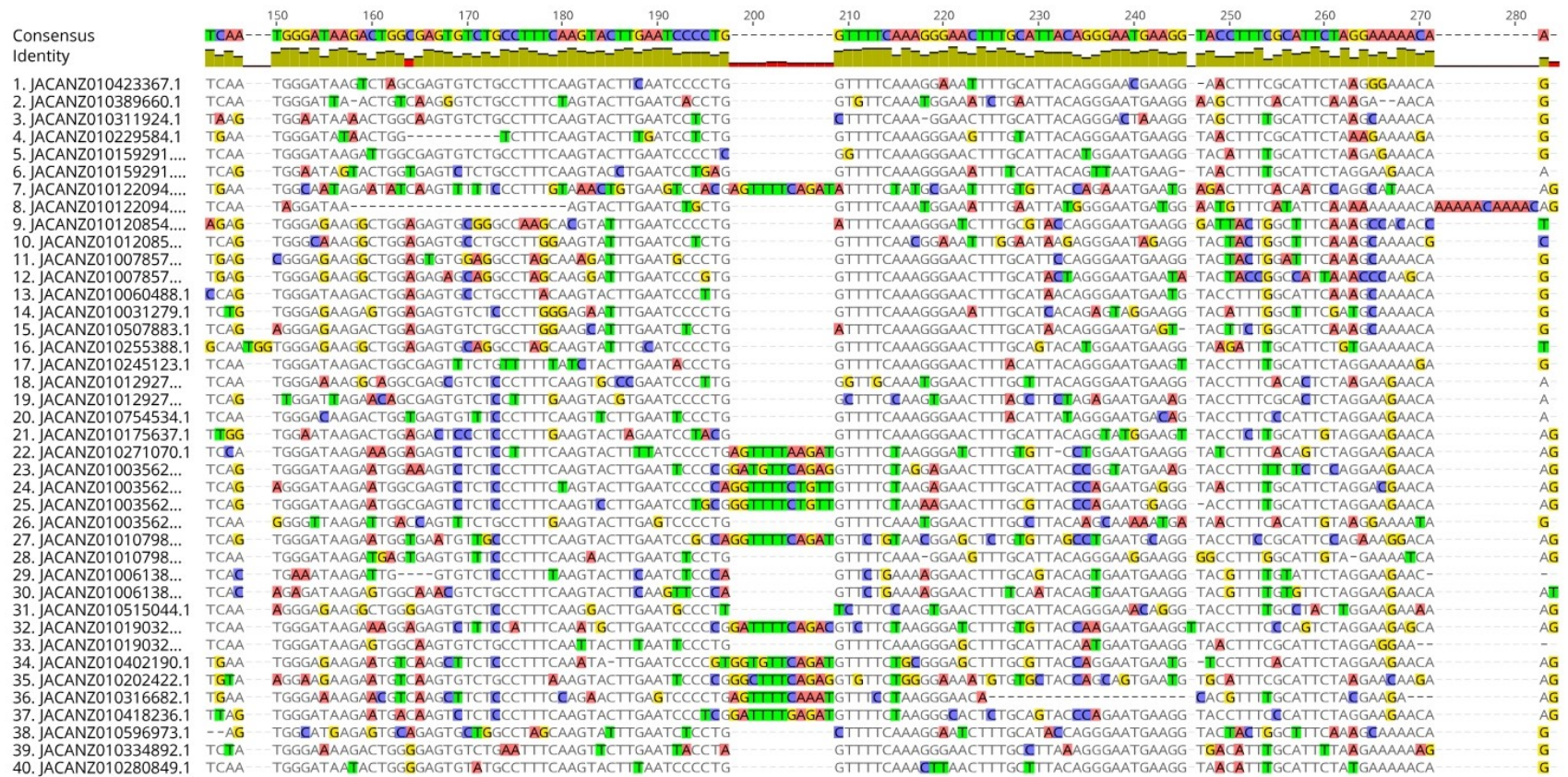
Supplementary Figure 7. Alignment of 40 monomers manually isolated from the *Hydrodamalis gigas* assembled reference genome (accession GCA_013391785.1). Part two of six.
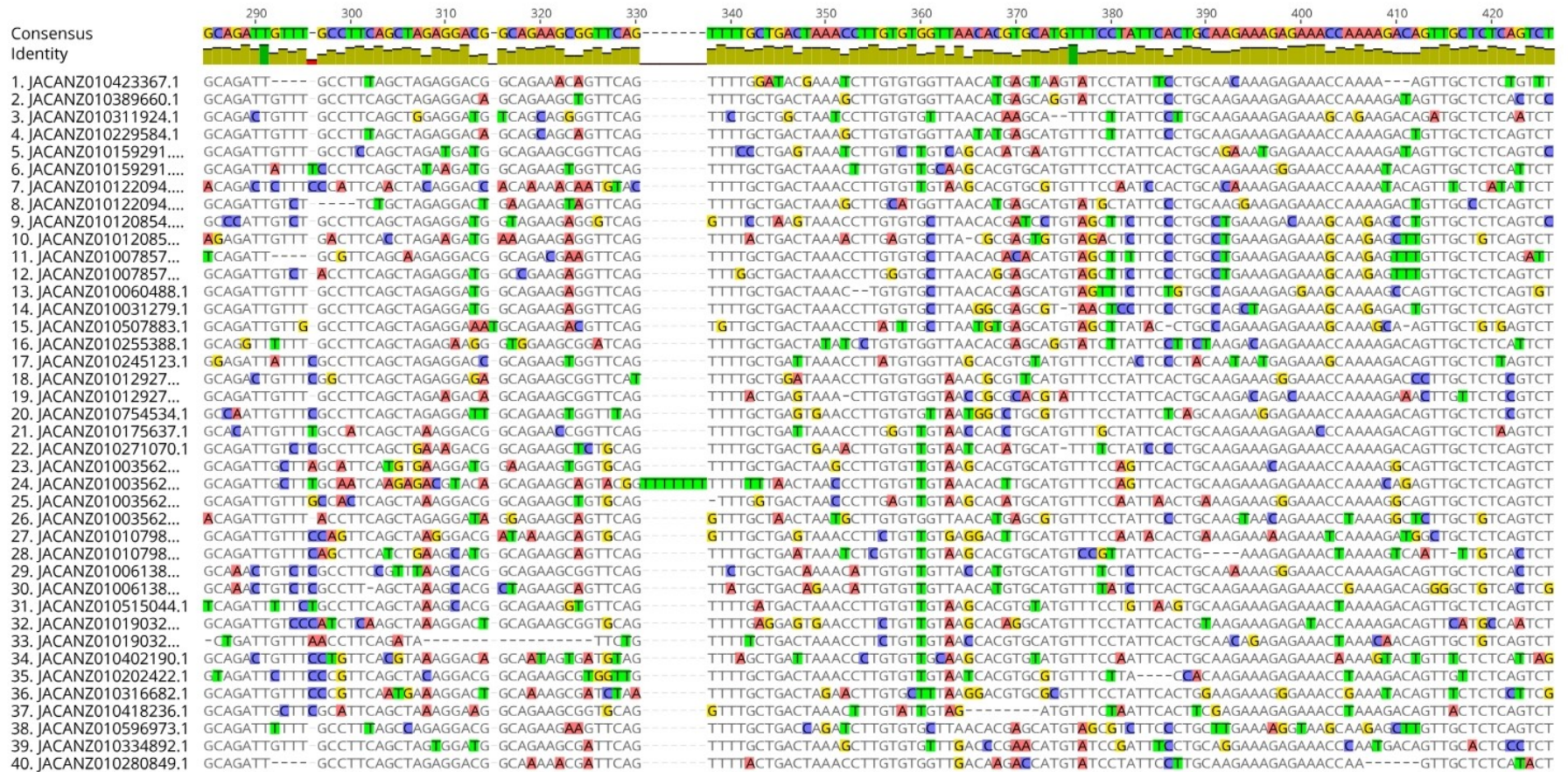
Supplementary Figure 7. Alignment of 40 monomers manually isolated from the *Hydrodamalis gigas* assembled reference genome (accession GCA_013391785.1). Part three of six.

Supplementary Figure 7. Alignment of 40 monomers manually isolated from the *Hydrodamalis gigas* assembled reference genome (accession GCA_013391785.1). Part four of six.
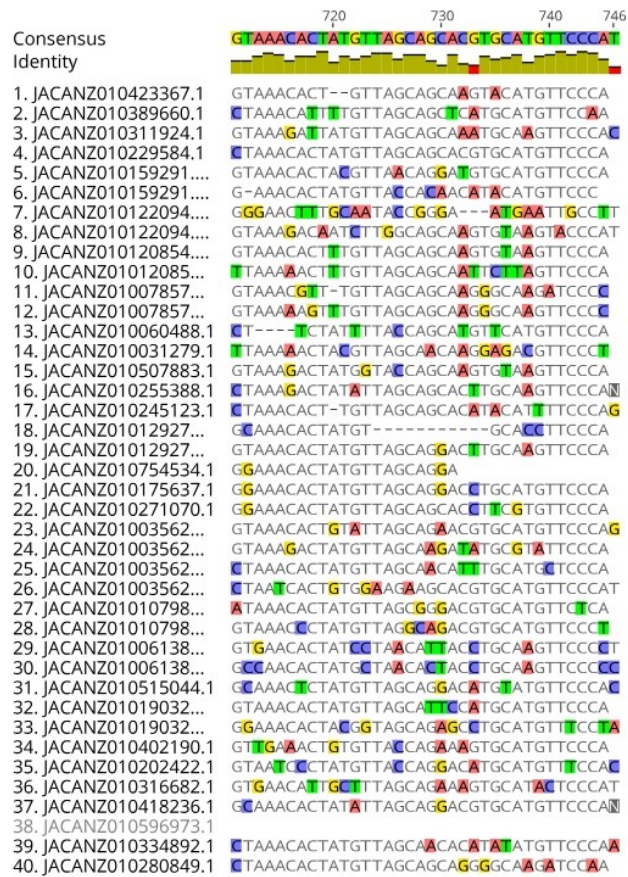
Supplementary Figure 7. Alignment of 40 monomers manually isolated from the *Hydrodamalis gigas* assembled reference genome (accession GCA_013391785.1). Part five of six.

Supplementary Figure 7. Alignment of 40 monomers manually isolated from the *Hydrodamalis gigas* assembled reference genome (accession GCA_013391785.1). Part six of six.

Supplementary Table 1. Estimates of average nucleotide divergence (number of base substitutions per site) over all sequence pairs within and between species.

	TMA	TIN	DDU	HGI
TMA	0.32			
TIN	0.285	0.06		
DDU	0.354	0.357	0.37	
HGI	0.342	0.356	0.355	0.33

TMA: *T. manatus,* TIN: *T. inunguis*, DDU: *D. dugon*, HGI: *H. gigas*

Supplementary Table 2. Blast searches using the TMAsat Most Common Sequence (MCS) from *Trichechus manatus* as query against the GenBank Mammalia (NCBI:txid40674) wgs database excluding Sirenia (NCBI:txid9774).

Species	Description	Coverage (%)	Identity [%]	E-value	Accession number
Loxodonta africana	isolate ISIS603380 cont3.14052, whole genome shotgun sequence	60	68.94	7,00E-10	AAGU03014053.1
Loxodonta africana	isolate ISIS603380 cont3.64816, whole genome shotgun sequence	62	68.29	8,00E-09	AAGU03064817.1
Elephas maximus	isolate Icky scaffold_1324, whole genome shotgun sequence	60	68.51	3,00E-08	JABTCH010005241.1
Elephas maximus	isolate Icky scaffold_65, whole genome shotgun sequence	35	70.22	1,00E-07	JABTCH010002495.1
Elephas maximus	isolate Icky scaffold_751, whole genome shotgun sequence	58	65.02	1,00E-07	JABTCH010005233.1
Loxodonta africana	isolate ISIS603380 cont3.85517, whole genome shotgun sequence	58	65.02	1,00E-07	AAGU03085518.1
Elephas maximus	isolate Icky scaffold_1495, whole genome shotgun sequence	47	68.00	1,00E-06	JABTCH010003728.1
Loxodonta africana	isolate ISIS603380 cont3.56850, whole genome shotgun sequence	47	68.00	1,00E-06	AAGU03056851.1
Procavia capensis	isolate Dallas Zoo Isis #01D731 contig_177334, whole genome shotgun sequence	13	75.53	4,00E-06	ABRQ02177336.1
Heterohyrax brucei	isolate US064 HetBruBak_scaffold_45440, whole genome shotgun sequence	6	89.36	1,00E-05	PVJQ01022728.1
Loxodonta africana	isolate ISIS603380 cont3.91379, whole genome shotgun sequence	26	70.17	5,00E-05	AAGU03091380.1
Elephas maximus	isolate Icky scaffold_624, whole genome shotgun sequence	13	74.19	5,00E-05	JABTCH010004173.1
Loxodonta africana	isolate ISIS603380 cont3.14053, whole genome shotgun sequence	7	84.91	5,00E-05	AAGU03014054.1

Supplementary Sequence File 1. TMAsat monomeric sequences (in fasta) manually isolated from the *Trichechus manatus* assembled reference genome (accession GCA_000243295.1).

>NW004444102.1_1
TTTCACCGCATGAAAGAGACAGCAAAATACAGCGTTCTCCTTCAAGCTGCTTAACTGAATCTACAGAAGTTTTCAATT
TTTTTCGAAGGAACAGTCCACCAACCAGAAGCTGCCTTCTGAAATCTTCTCTCAGATGGAGAAGACTGGAGAGTAC
CTGCTTTGCAAGTAATTGAATCCCCTGGATTTCAAAGAGAACTTTCTATAACAGGGAAAGAAGGTAGTTTTGGCTTT
CAAAGCAAAACAGGCAGATTGTTTGCCTTCAGCGAGAGGATGGCAGAAGAGGTTCAGTTTTGCTGACTAAACTTTG
TGTGATTAATGGGAGCATGTAGCTTCTTCCCTATCAGAAAGAGAAAGCAAGAGCCTGTTGCTTTCAGTCTGGGGTTT
GAAAAGGAATCACTTATTGTGTTTGAAAATCGTTCTGTGTTAAATGCCTAAGGAAGATGCCTGTGCATTGTTTACATG
GATAAAGTGTATGACTTTTCACCCTTTCAAGTCCTTGAAGCCAATCAGACTTTCCAGCAATCCGGGAGGCATTCCGT
GCTGCGGGGCTGAGAAGACCGCTATACAGACGGTATCGCTAAAGGATAAGAGAAACAGCATTTGTCCCATTCTGG
AACACGAATGAAGCAAAACCATAGTTGTGCTGCGTAAAAACTTTGTTAGCAGCAAGTGTAAGTTCTCA
>NW004444102.1_2
ATTCACCACATGAAAGAGAAAACAAGATACAGCGGTCTCCTTCAAGCTGCTTAACTGAATCAACTGAAGTTTTCAGT
TTTGTTAGAAGAAAGAGTCCGCCCATCATAAGCTGCTTTCTCAAATGTTCTCTCAGTGGGAGAAGGCTGGAGAGTG
CAGGTCTAGCAAGTATTTGAATACCCTGGTTTTCAAAGCGAACTTTGCATACCTGTGAATGAAGGTACTACTGGCTT
TCAAAGTCAAACAGGCAGATCTTCTCTCTTCAGCTAGAGGACAGCAGAAGAGGTTCAGTTTGGCTGACTAAACCTT
GTGTGCTTAACAGGAGCATGTAGCTTCTTCCTGCCTGAAATAGAAAGTAAGAGCCTGTTGCTCTCCGTCTGGGGAT
AAAGGGAGAACGATTCACTTTCTGTGTTTGAAAATCCCTCTGTGTTACATGCCTAAGAAAGTTGCCTGTGCATGGTC
TACCTGGATAATTTGTATGACTTATCCCCTTTCAAGTAGTTGAAGCCGTTCAGGGTTTCCAGCATTTCGGGAGGCAT
TCTGTGTTGCGTGGCTGAGAATACAGCTTTACAGATGGTATCGCTAAAGGAGAAGAGGAACCGCGTTTCTCCCGTT
CTGGAACACGAATGAGGCAAAACCGTAGTTGTGCTGTGTAAAAATTTGCTGGCAGCAAGTGTAAGTTCCCA
>NW004444102.1_3
ATTCACCCCATGAAAGAGAAAGCAAAATACAGCGTTCTCCTTCAAGCTGTTTAACTGAATCTACTGAAGTTTTCAGTT
TTGTTAGAAGAAAGAGTCCGCACACTACAAGCTGCTTTCTCAAATGTTGTCTCAGTGGGAGAAGGCAGCAGAGTGA
AGGCCTAGCAAGAATTTGAATGCCCTGGTTTTCAAAGGGAACTTTGCATTACAGGAATGAAGGTATTGCTAGCTTAG
ATAGCCAAACAGGCAGATTGTCTGCCTTCAGCTAGAGAACAGCAGAAGAGGTTCAGTTTTGCTGACTAAACCTTGT
GTGCTTAACACGAGCATGTACCTTCTTCCCAGCCTGAAAGAGAAACAAGAGCTTGCTGCTCTCAGTTTGGGGATGA
AGGGAAAAAGATTCCCTTTCTGTGTTTGAAAATCCCTCAGTGTTACATGCCTAAGAAAGTTGTCTGTGCATTGTCTA
CCTTGATAAAGTGTTTGACTTTTCTCCATTTGAAGTAGTTGAAGCCGTTCAGGGTTTGCAGCATTCCGGGAGGCATT
CTGTGCTGCGCGGCTGAGAATACCGCTTTACAGACGGTATCGCTAAAGAAGAAGAGAAACCGCGTTTCTGCCTTTC
TGGAACACGAATGAGGCAAAAGCGTAGTTGTGCTGCGTAAAAACTTTGTTAGCAGGAAGTGTAAGTTCCCA
>NW004444102.1_4
ATTCACCAAATGAAAAAGAAAGCAAAATACAGCGTTCTCCTTCAAGCTGCTTACTGAATCTACTGAAGTTTTCAGTTT
TGTTAGAAGAAAGAGTCAGCCCACCCCAAGCTGCTTTCTCAAATGTTCTCTCAGTGGGAGAAGGCTGGTGAGTGCA
GGCCTTGCAAGTATTTGAATCTCCTGATTTTCAAAGGGAACTTTACATACCAGGGAATGAAGGTTCTAATGGCTTTC
AAAGCCAAACAGGCAGATTGTCTGCCTTCAGCTAGAGGACAGCAGAAGAGGTACAGTTTGGCTGACTAAACCTTGT
GTGCTTAACAGGAGCATGTAGCTTCTTCCCTGCCTGAAAGAGAAAGCAAGAGCTTGTTGGTCTCAGTCTGGGGATG
AAGGGAGAACGGTTCACTTTCTGTGTTTGAAAATCCCTCTGTGTTACTTGCCTAAGAATGTTGCCTGTGCATTGTCT
ACCTCTGTAAAGTGTATGACTCTTCTGCGGTTCAAATAGTTGATACCTTTCAGTGCTTACAGAGTTCCGGTAGGGAG
TGTCTGCTGCCCAGCTGAGAATACTGCTTTACAGACGGTATCGCTAAAGGAGAAGAGAAACTGTGTTTCTGCTGTT
CTGGAACACGAATAAGCCAGGCCACAGTTTGGCTGAGTAAAAAGTATGTTTGTAGCAAGTATAAGTTACC
>NW004444102.1_5
ATTCAGTGCATGAAAGAGAAAGCAATATACAGCTTTCTCCTCAAGTTGCATAACTTCATCCACTGAAGTTTTCAGTTT
TGGTAGAAGAAACACTCCTCACAACACAAGCTGACTTCTCAAAGGTTCTCTCAGTGGGATAAGACTGGAGAGTGCC
TGCCTTGCAAATACTTGAATCCCCTGCTTTTCAAAGGGAACTTTGCATAACAGGGAATCAAGATACTTTTGGCATTC
ACAGCAAAACAGGCAGATTGCTTGCCTTCAGCTAGAAGATGGCAGAAGAGTTTCAGTTTTGCTGACTAAAGATTGT
GTGCTTAACGCGAGCATGTAGCTTCTTCCCTGTCAGAAACAGGAAGAAAAGGCAGTTGCTGCTCTCAGTCTGGGG
GTTGAAGTGAGAAGTTTTCCCTTACCTTATTTGAAAATCCCTCTATGTTACATGCCTAGGAATGTTGCCTGTGCATTG
TCTACCTGGATAAAGGGTTTGATTTTTCTCCTGTTCGAGTAGCTGAAGCCATGCGGGGTTTAAAGCTTTCCAGGAGT
CATTCTGTGCTGTGCAGCCGAGAAAGCCGCCTTACAGACGGTATCGCTAAAGGAGAAGAGAAACGGCGTTTCTCC
CGTTCTGGTGTAGGACAGAGGCAGCGCTGTAGTCTTGCTGAGTGAAAGACTAAGTTAGCAGCAAGTGTAAGTTCCC
A
>NW04444102.1_6
ATTCACTGCATGAAAAAGGAAGCAAAATACAGCGTTCTCCTTGACGCTGCTTAACTGCATCCACTGAAGTTTTCAGT
TTTGGTAGAAGAAACAGTCCACTCAACAGCAGCTGCTTTCTCAAAAGTTCTCTCAGTGAGATAAGACTTGAGAGTGC
CTGCCTTGCAAGTACTTGAATCCCTTACTTATCAAAGGGAACTTTGCATAACAGGGAATCAAGCTACTTTTGGCATT
CAAAGAAATACAAGCAGATTATTTGCCTTCAGCTAGAGGATGGCAGAAGAGGTTCAGTTTTGCTGACTAAAGCTTGT
GTGCTTAACGTGCGCATGTAGCCTCTTCCCTGCCAGAAAGAAGAAGCAAAAGCCTGTTCTCTCAGTCTGGGGATGA
AGTGCAAAGGATTCACCTACTGTGTTTGAAAATCCATCTGTGTTACATGCATAAGAATGTTGCCTGTGACTGTGTAC
CTGGGTAAAGGGTTGGATTTTTCTCCCGTTTGAGTAGCTAAAGCCCTGCTGGATTTACAGCTTTAGGGAGGCATTCT
ACGTTGTTCAGCCGAGAACCCCGCCTTACAGACGGTATCTCTAAAGGAGAAGAGAAACCGTGCTTCTCCCATTCTG
GTATGCAACAGCGGCAGCGGTGTAGTCTTGCTGCGTAAAAGACTATGTCAGCAGCAAGTGTAAGTTGCCG
>NW004444102.1_7
ATTCACTGCATGAAAGAGAAAGCAAAATACAGCATGCTCCTTCAAGCTGCTTAACTTCTTCCACTGATGTTTTCAGTT
TTGGTAGAAGAAACAGTCCGTACACCACCAGATGACTTCTCAAATGTTCTCTCAGTGGGATAAGACTGGAGAGTGC
CTGCCTTGCAAACACTTGAATCCCCTGGTTTTTAAAGGGAACTTTGCATAACAGGGAATCAAGGTACTTTTGGCATT

AAAAGCAAAACAGGCAGATGGTTTGCCTTCAGCTAGAGGATGGCAGAAGTGTTTCAGTTTTGCTGACTGAAGCTTG
TGTGCTTAACGCGAGCATGTAGCTTCTTCCCTGCCAGAAAGAGGAAGAAAAGGCAGTTGCTGCTCTCAGTCTGACT
GTTGAAGTGAGAAGGATTCACTTACTGTGTTTGAAATTCCCTGGGTGTTACATGCCTAAGAATGTTGCCTCTGCATT
TTCTACGTGAGTAAAGGGTTTAACTTTTCTCCCATTAGAGTAGCTGAAGCCCTGCGGATTTTACAGCTTTCCGGGAG
GCATTCTGTACTGTGCCGCTGAGAATGCCGCCTTATAGACAGTATCCCTAAAGGAGCGGATAAACCACATTTTTCC
CGTTCTGTTATAGGACAGCGGCAGCGTTGTAGTTTTGTTGAGTAAAGACTATGTTAGCAGCAAGTGTAAGTTCCTA
>NW004448733.1
ATTCACCACATGAAAGAGAAAACAAGCTACAGCGTTCTCCTTCAAGCTGCTTAACTGAATCTACTGAAGATTTCAGT
TTTGTTAGAAGAAAGAGTCCGCCCAACAGAAGCTGCTTTCTCAAATGTTCTCTCAGTGGGAGAAGGCTGGAGAGTG
CAGGCCTAGCAAGTATTTGAATACCCTGGTTTTCAAAGCGAACTTTGCATAGCAAGGAATGAAGGTACTACTGGATT
TCAAAGCCAAACAGGCAGATTGTCTGCCTTCAGCTAGAGGACAGCAGAAGAGGTTCAGTTTGGCTGACTAAACCTT
GTGTGCTTCACAGGAGCATGTAGCTTCTTCCCTGCCTGACAGAGAAAGCGAGAGCTTGTTGCTCTCAGTCTGGGGA
TGAAGGGAGAACGATTCACTTTCTGTGTTTGAAAATACCTCTGTGTTACATACCTAAGAAAGTTGCCTGTGCATTGT
CTACCTGGATAAAGTGTATGACTTTTCCCCCTTTCAAGTCGTTGAAGCTGTTCAGGGTTTCCAGCATTCCCGGAGG
GATTCTGTGCTGCGCGGCTGGGAATACCACTTTACAGAACGTATCGCTGAAAGAGAAGAGAAATCGCGTTTCTCCC
CTTCTGCAACACGAATGAGGCAAAACCTTAGTTGTGCTGTGTAAAACTTTGTTAGCAGCAAGTGTAAGTTCCCA
>NW004444141.1_1
GTTCACCACGTGAAAGAGAAAGCAAACTACAGCGTTCTCCTTCATGCTGCTTAACTGAATCTATTGAAGTTGTCAGT
TTTGTTCGAAGAAAGAGTCCCCCCAACACAAGCTGCTTTCTCTAATGTTCTCTCAGTGGGAGAAGGCTGGAGAGTG
CGGGCCTTGCAAGGATTACAATCCCCTGGATTTCAAAGGGAACTTTGCATTCCAGGGAATGAAGATACTTCTGTCTT
TCAAAGCCAACCATGGAGATTGTTTCCCTTCAGCGAGAGGACAGCAGAAGAGGTTCAGTTTTGCTGACTAAACCTT
GTGTTCTTAACACGAGCATGTAGGTTCCTCCCTGCCTGAAAGAGAAAGCAAGAGCTTGTTGCTCTCAGTCTGGGGGT
TGAAGGGAAAAGATTCACTTTCTGGGTTTAAAAATCCCTCTGTGTTCCATGCCTAAGAAAGTTGCCTGTGCATTGT
CTACCTGGATAAAGTGTATAACTTTTCTCCCTTTCAAGTAGTTGAAGCCGTTCAGGGTTTCCAGCATTCCGGGAGGC
ATTGTGTGCTGCGCGTCTGAGAATAGCGCTTTACAGGCAGTATCGCTAAAGGAGAAGAGAAATCGCGTTTCTGCCA
TTCTGGAGCACGAATGAGGCAAAACCGTATTTGTGCTGAGGGAAACCTTTGTTAGCAGCAAGTGGAAGTTCCCA
>NW004444141.1_2
TTTCACCACGTGAAAGAGAAAGCAAACCACAGCCTTCTCCTTCGAGCTGCTTATCTGAATCTACTGAAGTTTTTAGT
TTTGTTAGAAGAAAGGGTCCCCCCAACACAAGCTGCTTTCTCTAATGTTCTCTCAGTGGGGAAGGCTGGAGAGTGC
AGGCCTAGCAAGTATTTCAATCCCCTGGATTTCAAAGGGAACTTTCCATTCCAGGGAATGAAGGTATTTCTGTCATC
CTAAGCCAAACAGGCAGATTGTTTGCCTTCAGCGAGAGGACAGCAGAAGAGGTTCAGTTTTGCTGACTAAACCTTG
TGTGCTTAACACGAGCATGTAGTTTCTTCCCTGCCTGAAAGAGAAAGCAAGAGCTTGTTGCTCTCAGTCTGGGGAT
GAAGGGAGAACGATTCACTTTCTGTGTTTGAACATCCCTCTGTGTTACATGCCTAAGAAAGTTGCGTGTGCATTGTC
TACCTGGATAAAGTGTATGACTTTTCTCCCTTTCAAGTAGTTGAAGCCGTTCTGGGTTTACAGCATTCCGGGAGGCT
TTCTGTGCTGCGCGGCTGAGAATACCGCTTAACAGACGGTATTGCTAAAGCAGAAGAGAAACCGCGTTTCGCTATA
GGACAGCGGCAGCGTGGTAGTTTTTTTGAGTAAAAACTATGTTAGCAGCAGGTCTAAGTTCCCA
>NW004444141.1_3
AATCACTGCATGAAATAGAGAGCAAAAAACAGCGTTTTCCTTCAAGCTGCATAACTTCATCCACTGAAGTTTTCAGTT
TTGCTGGAAAAAAACTTCCACACACCACCAGCTGACCTCTCAAATGTTCTCTCAGTGGGATAAGACTGGTGAGTGT
GTGCCGTGCAAGTACTTGAATCCCGTGGTTTTCAAAGGGATCTTTGCATTACAGGGGATGAAATTACCTTTGCCATT
GTAAGCAGAACGGTTAGATTGTTTGCCCTCAGCTAGAGGATGGCAGAAGTGGTTCACTTTTCCTGACTACATCTTGT
GTGCTTAAAGCGAGCATGTAGCCTCTTCCCTGCCTGAAAGAGGAAGAAAAGGCAGTTGCTGTCAGTCTGGAGGTG
AAGTGAGAAGGATTCAATTACTGTGTTTGAAAATCCCTCTGAGTTACATGCCTAAGAAAGTTGCCTGTGCATTGTCT
ACCGGGGTAAGGGGTTTACTTTTCTCCTGTTCAAGTAGCTGAAGCCCTGCGCGTTTTACCGCTTTCCTGGAAGCAT
TCTGTTCTGCGCGGCCGAGAATGCCGCCTTACAGACGGTATCCCTGAAGGAGAAGAGAAACTGTGTTTCTCCCGTT
CTGGTATAGGACAGCGGCAGCGTTGTAGTTTTGCTGAGTAAAACTGTTTTAGCAGCAAGTGTAAGTTCCCA
>NW004444141.1_4
AATCACTGCATGAAAAAGAATGCAAAACAGTGTGTTCCCTTTCAAGCTGATTAACTTTGTCAACTGAAGTTTTCAGTG
TTAATAGAACAAACAGTCCTGACACCGCCAGCTACTTTCTGAAAATTTCTCTCAATGGGGAAAGACTGGGGATTGTC
TGCCTTTCTAGTACTTGAATCCCCTGGTTTTGAAAGGGAATTTTGCGTTACAGGGAATGAAGGAACTTTTCCCTTTCT
AAGAAAAACAGTCAGATTGTTTGCCTTCAGCTAGAGGATGGCAGACGCGGTTCAGTTTTGCAAACTAAAGCTTGTGT
GGTTGACACGAGCATGTATCCTATTCACTGTAAGGAAGAGTAATGAAAAAGTTGCTCTCAGTTTGGCGGATGAAGT
GAGAAAGTCACTTACTGTGTTTGAAATCTCTATGTGTTTCATGCCCAAAAATTGGCCTGTGCATTTTCTACCTGGGT
GAAGATTTTGACTATTCTCCCGCTTGAGTACTTAAAGCCCTGCGGGGTTTACAGCTTTCCTGGGAGGCACTCTGTG
CTGCATGAACCAGAATGCCGCATTGCAGACGGTACCCCTGAAGGAAAAGGGAAAGCATTGTTCTCGACTTCTGGTA
TACGACATCGGCAGCGTTGTAGTTTTGCTAAGTAAACACAGTCTGTTAGCAGCATGTTCCG
>NW004444141.1_5
ATTCACTGCATGAAAAGAAAGAAAAATACAGGGTTCTCATTCAGGCTGGTTAAGTGAATCAACTGAAGTTTTCAGTTT
TGGTTGAACAAACAGTCCTGACACCACCTGCTACTTTCAGAACATTCTCTCAATGGGGTAATACTGGTGAATGTCT
GCCTTTCAAGTACTTGAATCCCCTCGTTTTCAAAGGGAACTTTGCATTACAGGGAATGAAGTTAACTTTCGCATTCTA
AGGAAAACAGGCAGATTGTCTTCAGCTAGACAACGACAGAAGCAGTTCAGTTTTGCTGACTAAAGCTTGGTGGTTA
ACACGAGTATGTATCCTATTCCCTGCAAGAAAGAGGAAGAAAAGACAGTTGCTCTCAGTCTGGAAAATGAAGTGA
GAAAATTCACTTACTGTGCTTGAAATCTCTCTGTGTTAGATGCCCAAGAAAGTGGCCTGTGCATTGCCTACCTCGG
TAAAGGGTTTGACTTTTCTCCCATTCGAGTATGTGAAGCCCTGCGGGGTCTACGGCTTTCCAGGAGGCAGTGTGTG
CTGCGCACCCGAGAATGCCGCCTTACAGACGGTATCCCTGAAGAAGAAGATACAGCACGTTTCTCCCGTTCTGGT
GTATGACAGCAGTAGAGATATAGTTTTGCTAAGAAAACACTATGTTAGCAGGACCTGCATGTTCCT
>NW004444141.1_6
ATTCACTGCATGAAAGAGAAAACAAAATACAGTGTTCTCATTCAAGCTGGTTAACTTAATCAATTTGATTATCAATTTT
GATAAAGCGCAGTCCTCACACCATCTACAAGCGGTCTCAAACATTCTCTCCATGGGATAAGATTTGTGAGTTTCTCC
GTTTCAAGTACTTGAATCCCCGGTTTTCCAAGGAACTTTGCATTACAGGGAATGACTGTTCCTTTCGCATTCTAGG

AAGAAAAAGTGCAGATTTGTTCCCCTTCAGCTAAAGAACGGCAGAAGATTTTCAGTTTTGCTGACGAAAACTGGTGT
TGTAACCACGTGCGTGTTGCCTATTCACTGCAAGAAAGGGAATCCAAAAGACAGTTGCTCTCAGTCTGGTACATGA
AATGAGAATGATTTACTTACTGTGTTTGAGAATCTCTCTGTTTTTCATGCCCAGGAAGGTCGCCTATGCATTCTCCAC
CTGGGTAAAGAGTTTGATTTTTCTCCTATCCGAGTGCGTGAGCTCTGCGGGGTTTACAGTTTTCTGGGAGGCACTC
CGTGTTGCACAGCCAAGAATACTGCCTTACAGGTGGTATGTCTGAAGGAAAAGCAAAGCACGTTTCTCCCCTTCT
CCCCAGCAGTAGAGATGTAGTTTTGCTAAGGAAACACTATGTTAGCAGAACCTGCATGTTCCCA
>NW004444141.1_7
ATTCACTGCAAGAAAGAGAAAGCAAAATACAGTGTTCTCATTCCTTAAACAAGTTGGTTTCCTTAAACAAGTTGACTT
ATCCTTTTTGATAATACAGATTCCTGACAACATCTACAAGCTTTCTAAAACATTCCCTCAGTTGAAAAGACTGGAGCA
TCTCCCCCTTTCAAGGTAGTTGAATCCCCCGGTTTTCAAAGGGAGCTTTCCCTTACTGGGAATGACGGTATCTTTCA
TATCCTAGGAAGAACAAGGCAGATAATTTCGCCATCATATAAAGGACGACAGAAGCGGTTCAGTTTTGCAGACTAAA
CTTTGTTTTGTAACCACTTGCATGTTCCCAATTCACTGCAAGAAAGAGAAACCCAAATACAGTTTGTCTCATTCTAGT
GGGGTAAATCAGAAGGATTAACTTAGTGTGACTGAAATCCTTCTGTGTTACATGCACAAAATGTTGGCTCTACGT
TCTCCACCTTGGTAAAGAGATGACCTTTTCTCCCATTTGAGTACGTGAAGCCATGTGGGGTTTATGTGTTTTGAGTA
GGCGCTCTGTCCTGCCCGGCCGAGAATTTGCCTTACTGACTGTTTTTTGGAAAGAGAAACAGTGTAAGTTTCTGC
CTTTCAAGTATACGGCAGCAAGAGAGATGTCTGTTTGCTGAGGAAACACTTTGTGAGCAGAACATGCCTCTTCCC
>NW004444013.1_1
ATTCACCACATGAAAGAGAAAACAAGCCACAGCGTTCTCGTTCAAGCTGCTTAACTGAATTTACTGAAGTTTTCAGT
TTTGTTAAAAGAAAGAGTCCGCCCAACACAAGCTGCTTTCTCCAATGTTCTCTCAGAGGGAGAAGGCTGGAGAGTG
CAGGCCTAGCAAGTATTTGACTACCCTGGTTTTCAAAGCGAACTTGGCATAGCAGGGAATGAAGGTACCACTGGCT
TTCAAAGCCCAACAGGCAGATTCTCTGTCTTCAGCTAGAGGACAGCAGAAGAGGTTCAGTTTGGCCGATTAAACCT
TGTGTGCTTCACAGGAGCATGTAGCTTCTTCCCTGCCTGACAGAGAAAGCAAGAGCTTGTTGCTCTCAGACTGGGG
ATGAAGGGAGAACGATTCACTTTCTGTGTTTGAAAATCCCTCTGTGTTACATGCCTTAGAAAGTTGCCTGTGCATTG
TCTACCTGGATAAAGTGTATGACTTATGCCATTTCAAGTAGTTGAAGCCGTTCAGGGTTTGCAGCATTCCGGGAGG
CATTCTGTGCTGCGTGGCTGAGAATACAGCTTTAAAGATGGTATCGCTAAAGGAGAAGAGAAACCGCGTTTCTCCC
ATTCTGGAACACGAATGAGGCAAAACCGTAGTTGTGCTGCGTAAAAACATTGAGCAGCAAGTAAATTCCCA
>NW004444013.1_2
ATTCACCACATGAAAGAGAAAGCAAAATACAGCGTTCTGCTTCAAGCTGCTTAACTAAATCTACTGAACTTTTCAGTT
TTGTGAGAAGAAAGAGTCCGCCCACCACAAGCTGCTTTCTCAAATGTTCTCTCAGTGGGAGAAGGCTGGAGAGTGC
AGGCCTAGCAAGTATTTGAATCCCCTGGTTTTTAAAGCAAACTATGCATAGCAGGGAATGAAGGTACTACTGGCTTT
CAAAGCCAAACAGGCAGATTGTCTGCCTTCAGCGAGAGTATAGCAGAAGAGGTTCAGTTTTGCTAATTAAACCTAGT
GTGCTTCAAAGGAGCATGTAGCTTCTTCCCTGCCTGAAAGAGAAAGCAAGAGCTTGTTGCTTTCAGTTTGTGGATG
AAGGGAGAATGATTCACTTTCTGTGTTTGAAAATCCTTCTGTGTTACATGCCTAAGAGAGTTGCCTGTGCATTGTCT
ACCTGGATACAGCATATGACTTTTCAACCTTTCAAGTCGTCGAAGCAGTTCAGGGCTTCCAGCATTCCGGGAGGCA
TTCTGTGCTGCGCGGCTGAGAATACTGCTTTACAGACGGTATCGCTAAAGGAGAAAAGAAACCGCGTTTCTCCCAT
TCTGGAACACGAGTATGAAGCTTCTTCACTGCCTGAAAGAGAAAGCAAGAGCTTGCTGCTCTC
>NW004444013.1_3
ATTCACCACATGCAAGAGAAAGCAAAATACAGCCTTCTCCTTTAAGCTCCTTAACTGAGTCTACTGACGTTTTTAGTT
TTGTTAGAAGAAACAGTCCGCGGGACACAAGTTGTTTTCTCAAATGTTCTCTCAGTGGGAGAAGACTGGAGAGTGA
GGGCCTAGCATCTATTTGAATCTCCTGGTTTTCAAAGGAAACTTTGCATAACAGGGAATAAAGGTACTACTTGGTTT
CAAAGCAAAACAGGCAGTTTGTTTGCTTTCAGCCAGAGGAGGGCAGAAGAGGTTCAGTTTTGCTGACTAAACCTTG
TGTGCTTAAGGCGAGCATGTAGCTTCTTCTCTGCCTGAAAGAGAAAGCAAGAGCTTGTTTGTCTAAGTCGGGATGA
AGGGAGAAGGATTCACTTACTGTGTTTGAATATCCCTCTCTGTTACATGCCTAAGAAAGTTGCTTGTGCATTTTCTAC
CTGGATAAAGTGTATGACTTTTCTCCCTTTCAAGTAGTTGAAGTCGTTCAGGTTTGACAGCCTTCCGGGAGGCATTC
TGTGCTTCACGGCTGTGAAACCGCTTTACAGATCGTATCCCTAAAGGAGAAGAGAAACTGGTTTTTCTCCCGTTCTG
GAACATGAATGAGGCAAAACCGTAGTTGTGCTGAGTGAAACGTTGTTAGCAGCAAGTGTAAGTTCCTA
>NW004444013.1_4
ATTCACCCCATGAAAGAGAAAGGAAAATACAGCGTTCTCCTTCAAGCTGCTTAACTGAATCTACCGAAATTTTCAGT
TTTGTTAGAAGAAACAGTCCGCCCACTACAAGCTGCTTTCTCAAATATTCTCTCAGTGGGAGAAAACAGGAGAGTGC
CGGTCTAGCAGGTATTTCAATCCCCTGGTTTTCAAAGGGAACTTTGCATAACAGGGAATGAAGGGTCTACAGGCTT
TCAAAGCAAAACAGGCAGATTGTTTGCCTTCATCCAGAGGAGGGCAGAAGAGATTCAGTTTTGGTGACTAAACCTT
GTGTGCTTAACACGAGCATGTAGCTTCTTCTCTGCCTGAAAGAGAAAGCAAGAGCTTGTTGCTCTCAGTCTGGGGA
TGAAGGGAGAACGATTCACTTACTGTATTTGAAAATCCCTATGTGTTATATGCCTAAGAAAGTTGCCTGTGTATCGTA
TACCTGGATAAAGTGTATGACATTTCTCCCTTTCAAGTAGTTGAAGCCGTTCAGGGTTTCCAGCATTCTGGGAGGCA
TTCTGTGCTGCGCCGCTGAGAAGACCGCTTGACAGACGGTATCCCTAAAGGAGAAGAGAAACTGCGTTTCTCCCG
TTCTGGAACACGAATGAGGGAAAACCGTAGTTGTGCTGAGTAAAAACTTTGTTAGCAGCAAGT
>NW004444013.1_5
ATTCACTGTATCGGAGAGAAAGCAAAATACAGCGTTCTCCTTCAAGCTGCTTAACTGAATCTACTGAAGTTTTCACTT
TTGGTAGAGAAACAGTCCGCCAAACACCAGCTGCTTTCTCAGAAGTTCTCTCAGTGGGAGAAGACTGGGAGTGCCT
GCCTTGCAAGTATTTGAATCCCCTGGTTTTCAAAGGGAACGTTGCCTAACAGGGAATGAAGGTAATTTTGGCATTCA
AAGCAAAACAGACAGATTGTTTGCGTTCAGCTAGAGGAGGGCAGAAGATGTTCAGTTTTGCTGACTAAACTTTGTGT
GCTTAACACGAGTGTGTAGCTTCTCCAGTGCCAGAAGGAGAAAGCAAGAGCCAGTTGCTCTCAGTCTGGGGAAGA
AGTGAGAAGGATTCACTTACTTTGTTTGAAAATTTCTCTGTGTTACTTGCCTAAGAAAGTTGTCTGTGCATTGTCTAC
CTGGATAAAGTGTATGACTTTTCTCCCTTTCAAATAGTTGAAGCCATTCGGGGTTTACAGCATTACGGGAGGCATTC
TGTGCTTCGCGGCTGAGAATACCGCTTTACAGACGGTATCCCTATAGGAGAAGAGAAACCGCGTTTCTCCCGTTCT
GGAACATGAACCAGGCAGAGGCGCAGTTTTGCTGAGTAAAGACTGTGTTAGCAGCAAGTGTATGTTCCCA
>NW004444013.1_6
ATTCGTCGCATGAAAGAGAAAGCAACCTGCAGCATTCTCCTTCCAGCTGCTTAACTGAATCTACTGATGATTTCAGT
TTTGTTAGATGAAACACTCCGCCGACCACCAGCAGCTTTCTGAATCGTTCTCTCAGTGGGTAAGACTGGATAGTGC
CTGCTTTGCCAGTATTTGAATCCCCTGGTTTTCAAAGGGAACTTTGCTTTACAGGGAAAGGGGTAACATTCGCATTC

CAAGCAAAACAGGCAGATTGTTTGCTTTCAGCTAGAGGACTGCAGAAGTCGTTCAGTTTTGCTGACTAAAGCTTGTG
TGTTTAACACAAGCATGTATCCGATTCCCTGCAAGAAAGAGAAACCCAACGACAGTTTATCTCCCTCGGGGGGATG
AAGTGAGTATGATTTACTTACTGTGTTTGAAAATCGCCTGTGTTACATCCTCAAGAATGTTGCCTGTGCATTGTTTAG
CTGGGTAAAGACTTTGTCTTTTCTCCCTTTCGAGTACTTGAAGCCCTGCTGTGTTTACAGCTTTCCAGCAGGCACTC
TGTGCTGCTTGGCCTGCTGCCTTACAGAAGGTATTCCTGAAGGAGGAGAGAAAGCAGGTTTCTCCTCTTGTGCTAT
AGGACAGCGGCAGGGCTGTAGTTTTGCTGACCAAACTGTAT
>NW004444013.1_7
GTTCATCTAGGTAAGTTTTCAGTTTTGGTAGAACAAACAGTCCTCACACCATCAGCTGCTTTCTCAAACATTCTCTCT
ATGGGAAAGACTGGGAAGTGTCTGAATTTCAAGTTCTTGAATCCCCTGGTTTTCACAGGGAACTTTGCCTTAAAGG
GAATGAAGGTGACATTCGCATTTTAAGAAAAAAGGGCAGATTGTTTGCCTTCAGCTAGTGGATGGCAGAAGCGGTT
CAGTTTTGCTGACTAAATCTTGTGTGTTTGACCCGAGCATGTATCCGATTTCCTGCAAGAAAGAGAAACCCAATGAC
AGTTGCTCTCCCTCTGGGGGATGAAGTGAGTAGGATAGACTTACTCTGTTTGAAAATCCCCCGTGTTACATGCTCAA
GAATGTTGCCTGAGCTTTGTTCACCTGGGTAAAGACTTTAACTTTTCTGTCTTTAGAGTGCGTGAGGCCCTGCTATG
TTTACAGCTTTCCGGGAGGTACTCTGTGCTGCGCGGCCGAGAATGCCGCGTTACAGACGCTATCCCTGAAGGAGG
AGAGAAAGGGCCTTTCTCCCCTTCTGCTACATGGCAGTGGCAGTGCTGTCGTTTTGCTAACTAAACACTATGATAG
CAACACATATATGTTCCCA
>NW004443969.1_1
ATTCAGCACATGAAAGAGAAAGCAAAATACAGCGTTCTCCTTCAAGCTGCTTAACTGAATCTACTAAAGTTTTCAGTT
TTGTTAGAAGAAACCGTCCTCCCACCACCAGCTGCTTTCTGAAACGTTCCCTCAGTGGGAGAAGACTGGAGAGTGC
CTCCCTTGCAAGTATTTGAATCCCCTGGTTATCAAAGGGAACTTTGCATAAGAGGGAATGAAGGTACCTTTGACATT
CAAAGCCAAACAGTCAGATTATTTGCCTTCAGCTAGAAGATGGCAGAAGAGGTTCAGTTTTGCTGACTAAACTTTGT
GTGTGTAACGCGAGCATGAAGCTTCTTCCCTGCCAGAAAGAGAAAGCAAGAGCCTGTTTCTCTCTTTCTGGGGATG
AAGAGAGAAGGATTCAGGTACTGTGTTTGAAAATCCCTCTGTGTTACATGCCAGAGAAATTTGCCTCTAAATTGTCT
ACCTGGATAAAGTGTATGACTTTTCTCCCTTTCAAGTAGTTGAAGCCATTCAGGGTTTAAAGCATTCCGGGAGGCAT
TCTGTGCTGCCCAGCTGAGAACACCTCCTTACAGCCGTTATCTCTAAAGGAGAAGAGAAACTGCGTTTCTCCCGTT
CTGGAACAGAATGAGGCAGAGCCGCAGTTGTCCTGAATAAAGGCTATGTTAACAGCAAGTGTAAGTTCGTA
>NW004443969.1_2
ATTCAGCGCATGAGGGAGAAAGGAATATACAGCGTTCTCCTTCAAGCTGCTTAACTGAATCTACTGAAGTTTTCAGT
TTTGTTAGAAGAAACTGTCCACCCACCACCAGCTGCTTTCTCAAACATTCTCTCAGTGGGAGAAGACTGCAGAGTG
CCTGTCTTGCAAGTATTTGCATCCCTTGCTTTTCAAAGGGCACTTTGCATAAGAGGGAAAGAAGGTAGTTTTCCCAT
TCAAAGACACAGGCAGATTGTTTGCCTTCAGCTAGAAGATGGCAGAAGAGGTTCTGTTTTGCTGACTAAACCTTGTG
TGTGTAACGCGAGCATGTAGCTTCTTCCCTGCCAGAAAGAGAAAGCAAGAGCCAGTTGCTCTCAGTCTGGTGATGA
ATTGAGAAGGATTCACTTACTGTCTTTGAAAATCCCTCTGTGTTACATGCCTAAGAAAGTTGCCTGTGCATTGTCTAC
CTGTATAAAGTGTATGACTTTTCTCCCTTTCAAGTAGTTGAAGCTGTTCAGGGTTTGCAGCATTCTGGGAGGCATTC
TGGGCTGCGCGGCTGAGAATACCGCTTTACAGACGGTATCACTAAAGGAGAAGAGAACTGCGTTTCTCCCATTCTG
GAACACGAATGAGGCAAAAGCGTAGTTGGGCTGAGTAAAAACTTTGTTAACAGCAAGTGTAAGTTCCCA
>NW004443969.1_3
ATTCACCACATGAAAGAGAAAGCAAAATACAGCGTTCTCCTTCAAGCTGCTTAACTGAATCTACTGAAGTTTTCAGTT
TTGTTAGAAGAAACTGTCCGCCCACCACCAGCTGCATTCTCAGATCAGGAAGAAAGCAGAGTTCCTGCCTTGCAAG
TATTTGAATCCCCTGGTTTTCAAAGAGAACTTTGTATAAGAGAGAATGAAGGTACTTTTGCCATTCAAAGACAAACAT
TCAGATTGTTCCCCTTCAGCTAGAGGATGGCAGAAGAGGTTCAGTTTTGCTGACTAAACCTTTTGTGCTTAATGTGA
GCATGTAGCTTCTTCTCTGCCAGAAAGAGAAAGCAATAGCCAGTTGCTCTCACTCTGGGGATGAAGTGAGAAGGAT
TCACTTACTGTGTTTGAAAATCCCATCTGTGTTACTTGCCTAAGAAAGATGCCTGTGCATTGTCTACCTGGAGAAAG
TGTATGACTTTTGTCCCTTTCAAGTACTTGAAGCCATTCAGGGTTTACAGCATTCCTGGACGCATTCTGTACTGCGT
GGCTGAGAATACCGCTTTATAGACGTTATCGCTAAAGGAGAAGAGAAACGCGTTTCTCCCGTTTGGGAATATGAA
TCAGGCACAGCCGCTGTTTTCCTGAGTAAAGGCTATGTTAGCAATAAGTGGAAGTTCC
>NW004443969.1_4
ATTCAGCACATGAAAGAAAGCAAAATACAGTTTTCTCCTTCAAGCTGCTTAACTGAATCTACTGAAGTTTTCAGTTTG
GTTAGAAGAAACTGCCCACCACCACCTGCTTTCTAAAAGGTCGTCTCAGTGACAGAAGTCTGCAGAGTGCCTGCCT
TGCAAGTATTTGAATCCCCTGGTTTTCAAAGCGAACTTTGCATACCAGGGAATGAAGGCTTTCAAAGCCAAACAGGC
AGATTGTCTGCCTTCAGCTAGAGGACAGCAGAAGAAGTTCAGTTTTGCTGACTAAACCTGTGTGCTTAATGTGAGCA
TATAGCTTCTTGCCTGGCAGAAAGAGAAAGCAAGAGCCAGTTACTCTCAGTCTGGGGATGAAGTGAGTAGGATTCA
CTTACTGTGTTTGAAAATCCCTCTGTGTTACTTGCCTAAGAAAGTTGCCTGTGCATTGTCTACCTGGATAAAGTGTAT
GACTTTTCTCCCTTTCAAGTAGTTGAAGCCATTCAGGGTTTACAGCATTCCGTGAGGCATTCGGTACTACGTGGCTG
AGTAGACCGCTTTACAGAAGGTATAGCTAAAGGAGAAGAGAAACCACGTTTTTCCCGTTCGGAAATATGAATGGGG
TAGAGGCGCCGTTTTCCTGAGTAAAGGCTGTGTTAGCAGCAAGTTTAAGTTCCCA
>NW004443969.1_5
ATTCACCGCATGAAAGAGAAAGCAAAATACAGCGTTCTCCTTCAAGCTGCTTAACATAGTCCATTGAACTTTTCAGTT
ACATTAGAAGAAACAGTCCGCCCACCACCAGCTGCTTACTCATAGGTTCTCTGTGATGGATAAGACTGGGGAGTGC
TTGCCTTGGAAGTATTTAAATCCCCTGATTTTTGAAGGGAACTTTGCAAAACAGGAAATGAAAGTACATTTGGTATTG
AAAGCAAAACAGATTGTTTGCCTTCAGCTAAAGGATGGCAGAAGAGGTTCAGTTTTGCTGACTAAATCTTGTGTGCT
TAATGTGAGCCTGTAGCCTCTTCCCTGCCAGAAAGAGGAAGCAAAGCCAGTTGCTCTCAGTCTGGGGGACGAAGT
GAGAAGGATTCACTTACTGTGTTTGAAAATCGCTCTGTGTTACATGCCTAAGAAAGTTGCCTGGGCATTGTCTACCT
GGGTAAAGGGATCCACTTTTCTCCCTTTCAAGTCGGTGAAACCATTCAGTGTTTACAACATTCCGGGAGGCATTCTG
TGTTGCGCGGCCGAGAAAGCCGCTCTGCAGATGGTATCTCTAAGGGAGAAGAGAAACCGCTTTTCTCCAGTTCTG
GTATAGGACAGTGGCAGAGCAGCAGTTTTGAGTGTGTTAGCAGCAAGTGTAGGTTCCCA
>NW004443969.1_6
CTTCACTGCATGAAAGAGAAAACAAAATACAGTGTTTTTCCTTCAAGCTGCTTAACTCAATCTACTGAAGTTTTCAGTT
TTGTTAGAAGAAACAGTCCGCCCGCCACCAGCAGCTTTCTCAAAGGTTGTCTCAGTGGGATAAGACTGGAGAGTGC
CTGCCTTGCACATATTTGAATCCCTGGTTTTCAAAGGGAACTTTGCATAACAGACAATGAAATTACCTTTGGCATTCA

AAGCAAAACAGGCACATTGTTTGTCTTCAGCGGGAGGATGGTGGACGAGGTTCAGTTTTCCTGACTAAACCTTGTG
TGCTTAACGCGAGCATGTAGCTTCTTCCGTGCCAGAAAGAGAAAGCAAGAGTCAGTTGCTCTCATTCTGGGGGATG
CAGTGAGAAGGATTCACTTACTGTGTTTGAAAATCGCTCTGTGTTATATGCCTAAGAAAGTTGCCTGGGGATTTTCT
ACCTGGGTAAAGGGATCCAGTTTTCTCCCTTTCAAGTAGGTGAAGCTGTTCAGCATTTACAGCTTTCCGGGAGGCA
TTCTGTGCTGCGTGGCCAGGAATGCCGCCTTGCAGACATTATCCCTGAAGGAGAAGAAAAACTGCATTTCTCAGGT
TCTGGTATATGACAGCGGCACAGCAGCAGTTTTGCTGAGTAAAGACTATGTTAGCAGCAAGCGTAAGTTCCCA
>NW004443969.1_7
CTTCACCGCATGAAAGAGAAAGCAAACTACAGCATTCTCCTTAAGCTGCTTAACTTCATCCACTGAAGTTTTCAGTTT
TGTTAGAAGAAACAATCCGCTCAGCACCAGCTGCTTACTCAAAGTTTCTCTCAGTGGTATAAGACTAGAGAGTGCCT
GCCCTGCAATTATTTGAATCCCCTGGTTTTCAAAGGGAACTTTGCCTAACAGGGAATGAAGGTACCTTTGGCATTCA
AAGCAAAACAGGCAGATTGTTTGCCTTCAGCTAGAGGATGGAAGAAAAGGTTCAGTTTTGCTGAGTAAAGCTTGTG
TGCTTAACCTGAGCCTGTAGCCTCTTCCCTGCCAGAAAGAGGAAGCAAAGCCAGTTGCTCTCAGTCTGAGGGATGA
AGTGAGAGGGATTCACTTACTGTGTTTGAGAATCGCTCTGTGTTACATGCCCAGGAATGCTGCTTGTGCATTCTCTT
CCTGGAGAAAGATTTTGAGTTTTCTCCCTTGGACTTCTTGAATCCCTGTGGCTTTTGCAGCTTTCTGGGAGGCGCTC
TTCGCTGCATGGGCAAGAAAGCAGACTTTTCAGGCGGCATCTCTGAAGGAGAAACAAAGTGCCTTTCTCCCGTTC
TGGTATACGACAGCAGTAGAGATATAGTATCTCTAAGCAAACTCTATGTTAGCAGGACTTGTATGTTCCCA
>NW004443969.1_11
ATTCACTGCATAAAAGAGAAAGCAGAATACCGTTTCTCTCACTCGGCTGAAGTCAGTCAGAAGGAATAACATTCTGT
GTTTGAAAATCCCTCTGTGTTACATGCAAGAGAGTGCTGCCTAAACTTTCTCTATGGTTTTAAAGTGTTGAACTTTTC
TCCCATTCAAATATGTGAAGTCATATGGGGTATAGGTGTGTTGAAGCCGTACTCTGTGCTGTCTGTTTGTGAATGTG
CCTTACAGACTGAATCTTTGAAGGAGAAACAATGCTTCTTTTCCCATTCAAGTACAGGACGACAATAGAGATGTAG
TTTTGCTAAGCAAACACTATGTTACCAGAACCTGCATATTGTTCACAATTCACTGCATTAAAGAGAAAGTAAAATAGA
GTTTCCATCACTTGGCTTGAGTAAGTCAGAAGGAATAACTAACTGTGTTTGAATACCCCTCTGCGTTACATGCAAAA
GATGATTGCATATACATTCTCTACGGGGGTAAAGTTTTGAAGTTTTCTTCCATTCGAATACATGAAGCCATATGGAGT
ATAGATGTATTGAAGCAGTACTCTATGATGTTCTTTCCTGAATGTGACTTACACACCGAATCTCTGAAGGAGAAAC
AATGCGTGTTTCCTCGTTCCAGTACAGGACTGCAACAGAGATGTAGATTCGCTAAGGAAACACTGTGTTACAAGAA
CGTGCATGTTC
>NW004443969.1_12
ATTCACTGCTTGAAAGAGAAGCAAAAATACTGTTTCTTTAATAGTGTTTGAGTAAGTCAGAAGGATGAACTTACTGTG
TTTGAAAACCCCTCTGTGTTACACGCACGACAGTGTTGCCTATACGTTCTCTATCAGGGTAAAGTGTTCAAATTTTCT
CCCATTCAAATACATGAAGCTGTAAAGGGTTTAGATGTTCTCAAGCAGTACTCTGTGCTGCCCATTCCTGAATGTGC
CTTAAAGACAGAATCTCTGAAGGAGAAAACAAGGTGTGTTTCTCTGGTTCAAGTACAGGACGGCAATAGAGATGTA
GTTTCCCTAAGTAAACACTGTGTTACCAGAACATGCATGTTCCCAATTAACTGCATGAAAGAGAAAGCAAAATCCAG
TTTCTTTCACTTGGCTTGACTAAGTCAGAAGGATTAACTTACTGTGTTTGAAAACCCCTCTGTGTTACATACACGAGA
GTGTTGCCTACACATTCTGTACCAGGTAAAGTGTTAAACTTTTCTCCCATTCAAATATATGAAAGTCTTTGGGGTTTT
TGATGTTTTCCAGCAGTACTCCATGCTGCCCATTCATTGATGTGCCTTACAGACAGACTCTCTGAAGGGGAACACAA
TGCGTGTTTCTCTTATTCAAGTACAGGACGGTAATAGAGATGTGGTTTTGTTAAGACTGTGTTACCAGAACCTGCAT
GTTCCCA
>NW004443969.1_13
ATTCACCGTATGAAGGAAAACCTAAAATACAGGTTTCTCAGTCTAGTTGTTTAACTTAATCAAGTTGATTAAGGTATC
TGTTTTGGTAAAACACAGTAGACACAGGATTGAAAACTTTTCTTACACATTCTCTGAATGGTAAAAGAATGTCAAGTT
TCTCCCTTTCTTGTTCTTGAAGCCTCAGCAGTTTCAGTTGTTTTCCATAGAAACTCTGTGTTACTGTGACTGAACGTG
CCTTTCTCATTCTAGGAAAAACAAGGCAGGTTGTTTCCCGTTCTACTACAAGATGGCAAAAGTCTTGTAGTTTTGTTG
AGTATACCTTCTGTTGTAAGCACGTGCATGTTTCCAGTTTACTGCACAAAAAAGAAAGCAAAATACAGTTTCTCTCAT
TCTGCTTGACTAAGTCAGAAGGATTAACTTATTGTGTTTGAAAACCATCTGTGTTACATGCAAGAGAGTGTTGCTTAT
ACATTCTCTACCAGGGTAAAGTGTCTAACTTTTCTCCCATTCAAGTAGGTGACGCCATGCGGGGTTTAGATGTCTTC
AAGATGCACTTTGTGCTGCCCGGCCTTGAATATGCTGAATGAGAAAACAATGCAAGTTTCTCCCAGTCAAGCATAG
GACAGCAACAGAGATATACTTTTGCTAAGTTAACACTGTGTTACCACAACACACGTTCCCA
>NW004443969.1_14
TTTCACTGCATTGAAGAGAAACCAAAATACAGGTTTCTTATTCTAGCTGATTAACTTAATCAAGCTGGTTATCTGTTTT
TATAAAACATTGTTACCACACGATGGGAAACCTTTCTTGCACATTCTCTGAATGGGAAAAAAAATATTAAGTTTCTCC
CTTTCTGATACGTGAAGCCCCCGGGCTTTCAGATGTTTTCTATGGGAGCTCTGGTTTACTGGGAAAGAATGTGCCTT
TCACATTCTAGAAAGAACAAGGCTGATTGTTTCCCATTCAACTACAAAACGGCAAAGTGATGTGGTTTTGCTGACT
AAGCCTTGTGTTGTAAGCACGTGCATGTTTCCAATTCACTCCACGAAAGAGAAAGCAAAACACAGCTTCTCTCATTC
TGCTTGAGTAAGTCAGAGGAATTAACTTACTGTGTTTGAAAACCCCTCTATGTTACATGCATGAGAGTGTTGCCTAT
ACATTCTCAGCATAGTAAAGGGTATAACTTTAATCTCATTCAAATACATGAAGTCATACTGGATTTAGATGTTATCAA
GCAGTACTATGTGCTACCCTTTTGTGAATGTGCCTTACAGACAAAATCTCTGAAGGAGAAAACAAAGCGAGTTTCTC
CCCGTGAAGTATAGGACAACAATAGAGATGTAGTTTCCCTAGGTAAACACTGTGTTACCAGAACGTGCATATTCCCA
>NW004443969.1_15
ATTCACTGCATCAAAGAAAACCAAAAGAAAGGTTCTTCGGTCTAGTTGATTAACTTAATCAAGTTGATTAAGTTACC
TGTTTTGATAAAACACTGTTGCCACACGTTTAAAAACACTTCTTACACATTCTCTGAATGGGAAAGAATGTCAAGCT
TCTCCCTTTCCATTACCTGAAGCCCCCAGGCTTTCAGATGTTTTTTATGGGATCTCTGGTTTACCAGTGATGAATGT
GCCTTTCATATTCTAGGAAAAACAAGGCTGATTGTTTCCCATTCAACTACAGAATGGCAAAAGCTATGTGGTTTTGCT
GACTAAGCCTTGTGTTTTATGCACATGCATGTTTCTAATTCATTGCTCAAAAGAGAAACCAAAATACTATTTCTTTCAC
AGTCACTGAGTAAGTCATAAGGATTAAGTTACCGTGTTTGAAAACCCCTCTGTGTTACATGCACAAGAATGTTGCCT
ATACTTTCTGTACCAGAGTAACGGGTATAATTTTAATCCCATTAAATATGTGAAGTCATACGGGATTTAGACATTTTC
AAGCAGTATTCTGTGCTGCCCATTCTTGAATGTGCCTTACAGACAGTATCTCTGAAGGAGAAAACAATGCGAGTTTC
TCCTCGTGAAGTACAGAACAATACAGATGCAGTTTCCCTAAGTAAACACTGTGTTACCAGAATGTGCATATTCCCA
>NW004443980.1_1

ATTCACTGGATGAAAGAGAAAGCAAAATACAGTGTTTTCATTCAAGCTGGTTAAGTTAATCAACTTAATTAAGTTATC
AGTTTGATAAAACACAGTCCTTACACCATCTACAAGATCTCTCAAACATTCTTTCAATGACATAAGACTGGAGAGTTT
CTCCCTTTCAGGTACTTGAATCCGCCGGTTTTCAAATGGAACATTGCACTACAGAGAATGACGGTTCCTTTCACATT
CTAGGAAGAAAAAGGGCAGATTGGTTCCCTTTCAGCTAAAGAGCAGCAGAAGCTGTTCAATTTTGCTGACTACACC
TTGTGTTGCAACCACGTTTATCTTACTTATTCAGTGGAAAAAGGCAATTCAAAAGACACTTGCTCTCAGTCTGACGG
ATGGAGAAGGATTTACTTACTGTGTTTGAAAATCCCTCTGTGTTACATCGCCAGGAAGGTTGACTATGAATTCTCCA
CCTGGGTAAAGAGTTTTTTTTCTCCCTTCCGAGTACATGAAGCCTTGCGCTGTTTACAGCTTTCTGGGAGGCACTCT
GTGCTGCGAGGCTGAGAATGCTGCCTTACAGAAGGTATGACTGAAGACAAAACAAAGCGCGTTTCTTCCGTTTTG
GTACACTACAATAGTGGAAATGTAGTTTTTCTAAGGAAACACTATGTTAGCAGGCCCTGTATGTTCCCA
>NW004443980.1_2
GTTCACTGCATGAAATAGAAAGCAAAATACAGTGTTCTCATTCAAGCTGGTTAACTTGATTAAGTGATCAATTTTGAT
AAAACACAGTCCTCACAGCATCTACAAGCTCTCTCAAACATTCTCTCAATGACATAAGACAGGTGAATTTCTCCCTTT
CAAGTATTTGAATCCCCTGGTTTTCAAATGGAACTTTTCATTACAGAGAAGGAAGGTGTTTTCTGCATTCTAGAAATA
AAAAGGGGCCGATTGCTGCCCCTTAGCTAAAGAGCCGCAAAAGTGGTTCAGATTTGCTGACTAAACCTTGGGTTGTA
ACCATGTGCATGTTCCCTTTTCAATGCGAGAAAGAGAATCCAAAGGACAGTTGCTGTCAGTCTGGTGGATGAAGTG
AGAGTGATTTACTCACTGTTTGAAATCTGTCTGTGTTACATGCCCAGGAACATCATCTATGCATTCTCTACATGGGAA
AAGAGTTTGACTTTTCTCCTTTCCAAGAACGTGAAGCCAAGCGCAGTTTATAACCTTCTGGGAGGCACTCTGTGCTG
TGAAGGCCACAATGCTGCCTTATAGACAGTATGTCTGAAGACGAAAACAGAGCGTGTTTCTCCCATTCTGGTATATG
ACACCACTAGAGAGGTAGTTTTGCTAACGGAACACTATGGTAGCAGGCCCTGCAAGTTCCCA
>NW004443980.1_3
ATTCACTGCATGAAAAAGAAAGCAAAATACAGTGTTCTCATTCAAGCTGGCTAACTTAATCCACTTAAACTATCAATT
TTGATAAAAGATAGACCTCAAACCATCTACAAGCTCTCTCAAACATTCCTCAATGGGATAAGACTGGTGAGTTGCTC
CCTTTCAAATGCTTCAATCTCCCGCTTTTCAAAGGGAACTTTGCATTATATGAACTGAAGGTACTTTTCACATTCAAA
GAAAACAAGGCAGATTGGTTCATCTTCAGATTAAAGACAACAGAAGCGGTTCAGTTTTGCTGATTAAACCGTATGTT
GTAACCATGTGCATATTTCCTATGCACTGAAAGAAAGAGAAACCTAAAGAACCTTGCTCTTCATCTGGTGGATGAAG
TGAAAAGGATTCATTTACTGTGGTTGAAACTCCCTCTGTGTTACATGCCCAAGAATGCATGGTCTACCTGGGTAAGA
GTTTGACTTTTCTCCCCTCAGAGATTGTGAAGCTCTGCAGGGTTTACAGCTTTCCAGGAGTCGTTCTGTGCTGCACA
GCCGAGAATGCAGCCTTACAGACGGTCTCCCTAAAGGAGAAGAGAAAGTGTGTTTCTCCCGTTCTGGTATATGACA
GTGGCAGTGCCGCAGTTTTGCAGAGTAAAGACTATGTTAGCAGCAAGTGTAAGTTCCCA
>NW004443980.1_4
TTCACTGCCTGAAAGAGACAGCAAAAATGCAGCATTCTCCTTCAATCTACTAAACGTAATCCACTGAAGTTTTCTGTT
TTGGTAGAAGAAACAGTCCACACAGCACTAGCTGACTTCTCAAACGTTCTCTCAGTGGGATAAGACTGGAGAGTGC
CTGCCTTGCAAGTATTTGAATCCTCTGCTTTTCAAAGGGAACTTCGGATAACAGGGAATGAATATACCTTTGGCACT
CAAAGCAAACATACAGATTGTTTGCCTACAGGTAGAGGATACCAGAAGAGGTTCAGTTTTGCTGACTAAACCTTGTG
TGCTTAACGCAAGCATGTAGCCTCTTCTCTACCAGAAAGAGGAAGCATAACCCAGTTGCTGTCAGTGTGGGGGATG
AAGTGAGAAGGAGTCAATTACTGTGTTTGAAAATCCCTCTGTGTTATACACCTAAGAAAGTTGCCTGTACATTGTCTA
CCTGGATAAAGTGTTTGACTTTTCTCCCGTTCGAGTAGCTGAAGCCCTGTGGGGTTTACAGCTTTCAGGGAAGATTT
CGGTGCTACGCGGCCGAGAATGCCCACCTTACAGGCGTTATTGCTGATGGAGAAGAGAAACCGCGTTTGTCCCCTT
CTGGTATAGGACCGAGGCAGAGTCGCAGTTTTGCTGAGTAAAGACTATGTTAGCAGCAAGTGTAAGTTCCCA
>NW004444509.1_1
ATTCACCACATGCAAGAGAAAGCAAAATACAGAGTTCTCCTTCAAGCTGCTTAACTCATTCTACTGAAGTTTTTAGTT
CTGTTACCAGAAACAGTCCGCGCACCACAAGCTGCTTTCTCAAAAGTTCTCTCAGAGGGAGAAGACTGGAGAGTGC
GGGCCTAGCAAGTATTTGAATCTCCTCCTTTTCAAAGGAAACTTTGCATAACAGGGAATGAAAGTACTACTGGCTTT
CAAGGCAAAACAGGCAGATTGTTTGCCTTCAGCCGGAGGAGGGCAGAAGAGGTTCAGTCTCGCTGACTAAACCTT
GTGTGCTTAACGCTAGCATGTAGCTTCCTTCTTGCCTGAGGGAGAAAGCAAGATCTTGTTGCTCTCAGGCTGGGGA
TGAAGGGAGAACGATTCACTTACTGTGTTTGAAAATCCCTTGGTGTTACATGCCTAAGAATGCTGCCTGTGCATTTT
CTACCTGGATAAAGTGTATGACTTTTCTCCCTTTCAAGTAGTTGAAGCCGTTCAGGGTTTCCAGGATTTGGTAGGCA
TTCTTTGCTGCACGGCTGAGAATACCGCTTTACAGAAGGTATCACTAAAGAAGAAGAGAAACTGCATTTCTCCCATT
CTGGAACACGAATGAGGCGAAACCGTAGTTGTGCTGAGTAAAAACTTTGTTAGCAGCAAGTGTAAGTTGCCA
>NW004444509.1_2
ATTCACTGCCTGAAAGAGAAAGCAAAATACGGCGTTCTCCTTCAGCTGCTTAACTGAATCTACTGAAGTTTTCAGTT
TTGGTAGAAGAAACACTCTGCCCACCACCAGTTGCTTTCTCAAAGATCTCTCAGTGGGAGAAGACTGGAGAGTGCC
TGCCTTGCTGGTATTTGAATCCCCTGGTTTTCAAAGGGAACTTTGCATTACAGGGAATGAAGGTACTTTTGGCATTC
AAAGCTAAACAGGTACATTGTTTCCCTTCTGCTAGAGGACAGCAGAAGAAGTTCAGTTTTAATGAATAAAATATGTGT
GGTTAACGGGAGCATGTAGTTTCTTCCATATCAGAAAGCTAGAGCCAGTTTCTCTCGGTGTGGGGATGAA
GTGAAAAGGATTCACTTATTGAGTTTGAAAATCCCTCTGTGTTACTTGACTAAGAAAGTTGCCCGTGCATTGTCTAC
CTGGAGAAAGTGTATGACCTTTCTCCCTTTCAAGCAGTTGTAGCCTTTCAGGGTTCATAGCCTTCCGGGAGGCATT
CTGTGCTGCGCGGCTGAGAATACCGCTTTACAGACGTTATCCCTAAAGGAGAAGAGAAACCGAGTTTCTCCCTTTC
AGGAACATGAATGAGGCAGAGCCGCAGTTTTGCTGGGTAAACACTATGTTAGCCGCAAGTGTTTAAGTTCCCA
>NW004444509.1_3
ATTCACTGCATGAAAGAGAAAGCAAAATACAGCGTTCTCCTTCAAGCTGCTTAACTGAATCTACTGAAGTTTTCAGTT
TTGTTGGAAGATAGAGCCCCCCCACCACAAGCTGCTTTCTCAAATGTTCTCTCAGTGGGAGAAGGCAGTAGAGTCA
AGGCCTAGCAAGAATTTGAATCCCCTGGTTTTCAAAGGGAACTTTGCATAACAGGGAATGAAGGTACTCTTGGCATT
CAAAGCAAAACAGGCACATTTTTTCCCTTCTGCTAGAGGATGGAAGAAGCTGTTTAGTTTTCCTGATGAATTTTCAGT
GCTTAACACAAGCATGGAGTTTCTTCCATATCAGAAAGAGAAAGCGAGAGCCAGTTTCTCTTGGTGTGGGGATGAA
GTGAAAAGGATTCACTTATTGAGTTTGAAAATCCTTCTGTGTTAAATGCCTAAGAAAGTTGCCTGTGCATTGTCTACG
TGGACAAAGTGTATGACTTTTCTCCCTTTCAAGTAGTTGAAGCCATTCAGGGTTCATAGCCTTCCGGGAGGCATTCT
GTGCTGCGCAGCTGAGAATGCCCCTTTACAGACGTTATCCCTAAAGGAGAAGAGAAACCGAGTTTCTCCCATTCAG
GAACATGAATGAGGCAGAGCCGCAGTTTTGCTGGGTAAACACTATGTTAGCAGCAGGTGTAAGTTCCCA
>NW004444509.1_4

```
ATTCACTGCATGAAAGAGAAAGCAAAATACAGCGTTCTCCTTCAAGCTGCTTAACTGAATCTACTGAAGTTTTCAGTT
TTGGTAGAAGAAACACTCCGCCCACCACCAGCTGCTTTCTCAAAGGTTCTCTCAGTGGAAGGAGACTGGAGAGTGC
CTGCCTTGCCAGTATTTGAATCCCCTGGTTTTCAAAGGGAAAGTTCAATTACAGGGAATGAAGGTACTCTTGGCATT
CAAAGCAAAACAGGCACATTGTTTCCCTTCTGCTAGATGATGGCAGAAGAGTTTCAGGCGTGCTGAATAAAGTTTGT
GTGCTTAACGCGAGTATGTAGCTTCTTCCATGTCAGAGAGAGAAAGCGGAGCGTTTCTTTCGGTCTTGGGATGAAG
TGAGAAGGTTTCACTTACTGTGTCTGAGAATCCCTCTGTGTTACTTGCCTAAGAAAGTTGCCTGTGCATTGTCTACC
TGGATAAAGTGTATGTCTTTCTTCCTTTCAAGTAGTTGTAGCCATTCAGGGTTCATAGCCTTCCGAGAGGCATTCTAT
GCTGCGCGGATGAGAATACCGCTTTACAGAAGGTATCCCTAAAGGAGAAGAGAAACCACGTTTCTCTTGTTCTGGA
ACACGAATGAGGCAGAGCCGCAGTTTTTCTGAGTAAAGACTATGTTAGCAGCAAGTGCAAGTTCCCA
>NW004444509.1_5
ATTCACTGCATGAAAGAGAAAGCAAAATACAGCGTTCTCCTTCAAGCTGCTTAACTGAATCTACTGACGTTTTCGGT
TTTGGTAGAAGAAACACACCGCCCACTACCAGCTGCTTTCTCAAAGGTTTTCTAAGTGGAAGTAGCTTCTAGAGTGC
CTGCCTTGCCAGTATTTGAATCCCCTGGTTTTCAAAGGGAACGTAGCATAACAGGGAATGAAGGTACTTTTGGCATT
CAAAGAAAAACAAGCAGATTGTTTCCCTTCTGCTAGGGGATGGCAGAAGAGGTTCATTTTTGATGAATAAAGTTTGA
GTGCTTAACGCGAGCATGTAGCTTCTTCCATGTCAGAAAGAGAAAGCGAGAGCCAGCTTGTCTCGGTCTGGGTATG
ATGTGAAAAGGATTAACTTACTGTGTTTGACAATCCCTCTGTGTTACTTGCCTAAGAAAGTTGCCCGTGCATTGTCTA
CCTGGACAATGTGTATGACTTTTCTCCCTTTCAAGTAGTTGTAGCCATTCAGGGTTCATAGCCTTCCGAGAGGCATT
CTATGCTGCGTGGAGGAGAATACTGCTTTACAGAAGGTATCCCTAAAAGAGAAGAGAAACCACGTTTCTCCCGTTC
TGGAACACGAATGAGGCAGAGCCGCAGTTTTCCTGAGTAAAGACTATGTTAGCAGCAAGTGCAAGTTCCCA
>NW004444509.1_6
ATTCACTGCATGAAAGAGAAAGCAAAATACAGCGTTCTCCTTCATGCTGCTTGACTGAATCTACTCAAGTTTTCAGTT
TTGGTAGAAGAAACAGTCCTGACACCACCAGCTACTTTCTGAAACATTCTCTCAATAGGATAAAACTGACGAGTGTC
TGCCTTCCAAGTACTTGAATCTCCTGGTTTTCAAATGGAGATCTAAATTACAGGGAGTGAAGGAACATTTCTCATTCT
AAGGAAAGGAGGCAGATTGTTTGCCTTCACCTACAGGACAGCAGAAGCAGTTCAGTTTTACTCACTGAAGCTTGGG
TGCTTAACATGAGCATGTATCCTATTTCCTTCAAGAAAGAGAAAGCAAAAGACAGTTGCTGTTAGTATGGGAAATGA
AGTTAGAAAGATTCACTTACTGTGTTTCAAAATCCCTCTGTTTTATGTGCCCAAGTAAGTGGCCTGTGCATTGTCTAC
CTGGGTGAAGAGTTTTCCTTTTCTCCCGTTCGAGTATGTGAAGCCCTGCATGGTTTACAGCTTTCCCAGAGGCTCTT
TGTGCTGCGGTGCCGAGAATACCCCCTTTGAAAGAGAAGAGAAAGCGTGTTTCTCCAGTTCTGGTATATGACAGCA
GCTGCACTGTAGTTCTGCTAACCAAACACTATGTTAGCAGCACGTGCATGTTCCCA
>NW004444509.1_7
GTTCACTGCATAAAAGTGAAAGCAAAATACAGTGTTCTCATTCATGCTGGTTAAATTAATCAAGTTGATTAATTTATCT
GTTTTGAAAAACGGAGTTCCTACTTGGTGAAAACCTCTCTTACACGTTTTCTGAATGGAAAACAATGTCAGGCTTCT
CCCTTTCAAGTACTTTAAGTCCCTGTGTTTTCACTTGTTTTGTTTCTAAATTTTAGGTCACTCGCAATGAATGTGCCTT
TCAAATTCTATGAAGAACAAGGCGGTGTGTTTCCCTTTGAACTACAGCAAGGCAAAAATGATGTAGTTTTCCTGAAT
AAACCTTGTGGTGCAAGCACTGGGTAACTTTGCACTTCTCTGAACTAAAGAGAAACCAAAAGACAGTTTCTGTCATT
GTGCTTGAGAAGTCTGAAGGATTTACTTACTGTGTTTGAACACCCCTCTCTGCTACATGCAGGGGAAATTGCCTTT
ACCTTCTCAACCAGAAAAAAGTATTTTGCTTTTCTGCCATCCAAATACTTGAAGCCATACGCGGTTTAGCTGTTTTTA
AGATGCACTTGAGCTGCCCGACCGTGAATGGGCTTTACAGACAGAATCTCTGAAAGAGAATGCACTGCGAGGTTCC
CCCAGTCAAGTATAGGATGGCTATAGAGATGTAGTTTTCGAAGCAAACACTGTGTTAGTGCAACTTGCATGTTCTC
>NW004444509.1_8
ACTCACTGCCTGGAAGAGATGCAAAATACAGTGTAGTCACTCTAGTTGGTTAACTTAATCACATTGATTAAGTTATCT
GTTTTGAAAAACACAGTTCCTACATGGTGAAAACCTTTCTTACACATTGTCTGAATGGGAAAACAATGTCAAGCTTCT
CTTTTTTCAAGTACTTGAAGCCACCGGGTTTTCACATGTTTTCTGTCTAAATTTTAGGTTACTGGCCATGAATTTGCCT
TTCAAATTCTAGAAGAACAAAGCAGTTTGTTTCTCCTTCAACTACAGCACGGCATAAGGGATATAGTTTTGCTGAAT
AAACCTTGTGCTGCAAGCACGCGTTCGTTTGCACTTCACTGCACTAAAGAGAAACAAAAGACAGTTTCTCTCATTG
TGCTTGAGAAAGTCTGAAGGATTTACTTACTGTGTTTGAATACCCTTCTGTTCTACATGCAGGAGAACCTAGTCTTTA
TGTTATCAACCAGAAAAAAGTGTTTTGCTTTTCTGCCATCCAGATACGTGATACCGTAAGAGGTTTAAATGTTTTAAA
ATGCACTTGAGCTGCACAACCTCGACTATGCTTTGCAGACAGAATCTCTGAAGGAGAAGGCATTTCGAGGTTCTCC
CAGTCAAGTATAGGATGGCTATAGAGATGTAGTTTTTGAAGCAAACACTGTTAGCACAACGTGCAACTTCTC
>NW004444509.1_9
ACTCACTGCCTGAAACAGAAGCAAAATAAAGTGTATTCACTCTAGTTGGTTAACTTAATTAAGTTGACGTTATCTGCT
TTGAAAAACACAGTTCCTACATGGTGAAAACCTTTCTTACACGTAGTCTGAATGGGAAAACAATGTCAAACTTCTTCC
TTTCAAGTACTTGAGGCCCCTGGGTTTTCACATGTTTTCTAAATTTTAGGTTACCGGCAATGAATGTGCCTTTCAAAT
TCTAAGAAGAACAAGGCAGTTTGTTTCCCTTTCAACTACAGCATGGCAAAAGGGATGTAGTTTTGCTGAATAAACCT
TGTTCTGCAAGAACGGGCACGTTCACACTTCACTGCACGAAAGAGAAAACCATAAGACAGTTTCTCTCATTGTGCTTA
AGGAAGTCTGCAGGATTTACTTACTGTGTTCGAACACCTGTCTGTGCTACATGTAGGAGAACCTTGCATTTTCTTTC
TATAACAAGGAAAAGTGTCTCATTTTTCTGCCATCCAAATCCGTGAATCCATACGGTGTTTAGATGCTTTTAAGATGC
GCTTGAGCTGCGCAACCGTGAATGTGCTTTGCAGACAGTATTTCTGAAGGAGAATGCAGTGCGAGCTTCTCCCGTT
CAAGTATAGGATGGCTATCGAGATGTAGTTTTGTGAAGTAAACACTGTGTTAGTGCAACGTGCATGTTCTT
>NW004444309.1_1
ATTCACCACATGAAAGAGAAAGCCAAATACAGCGTTCTCCTTCATGCTGCTTACCTGAATCTACTGAAGTTTTCCGT
TTTGTTAGAAGAAAGAGTCTGCCCACCTCAACTGCTTTCTCAAATGCTCTCTCAGTGGGAGAAGGCAGTAGAGTGA
GGGCCTGGCAAGAATTTGAATCCCTGGTTTTGAAAGGGAACTTTGCATAACAGGGAATGAAGGTATTGCTAGCTTAC
AACGCCCAAACCGGCAGATAGTTTGCCTTCAGCTAGAGGACAGCAGAAGAGGTTCAGTTTAGCTAACTAAACCTTGT
GTGCTTAACACGAGCATGTAGCTTCTTCCCTGCCTGAAATAGAAAGCAAGAGCTTGCTGCTCTCAGTCTGGGGATG
AAGGGAGAACGATTCACTTTCTGTGTTTGAAAATCCCTCTTTGTTACATGCCTGCAAATGCTGCCTGTGGATTGTCT
ACCTGGATAAAGTGTATGTCTTTTCTACCTTTCAAGTAGTTGAAGACGTTCAAGGCTTCCAGCATTCCGGGAGACAT
TCTGTGCTGCGCGGCTGAGAATACCGCTTTACAGACGCTATCCCGAAACGAGAAGAGAAACCGCGTTTCTGCCATT
CTGGAACACGAATGAGGCAAAAGCGTAGTTGTGCTGAGTAAGAACTTTGTTAGCAGCAAGTGTAAATTCCCC
>NW004444309.1_2
```

CTTCACCACATGAAAGAGAAAGCCAAATACAGCGTTCTCCTTCATGCTGCTTAACTGAATCTTCTGAAGTTTGGTGT
TTTGTTAGAAGAAAGTGTCCGCCCATCACAAGATGCCTTCTCAATGGTTCTCTCAGTGGGAGAAGACTGGAGAGTG
CCTGCCTTGCAAGTATTTCAACCCCCTGGTTTTCAAAGGGAACTTTGTCTAACAGACAGTGAAGGTACTACTGGCTT
TCAAAACAAAACAGGCAGATTGTTTGCCTTCAGCTAGAAGATGGCAGAAGAGGTTCAGTTTTGCTGACTAAACCTTG
TCTGCTTAACGCGAGCATGTAGCTTCTTCCCAGCCTGAAACAGAAAGCAAGAGCTTGTTGCTCTCGGTCTGGGAAT
GAATTGTGAAGGATTCACTCTCAGTGTTGGAAAATCCCTGTATGTTCCATGCCTAAGAAAGTTGCCTGTGCAGTGTC
TACTTGGCTAAAGTGTATGACTTTTCTCCCTTTTGAGAAACTGAAGCCATTCAGGTTTTACAGCTTTCCGGGAAGCC
TTCTGTGCTGCGTGGCTGCGAATACCGCTTTACAGAAGGTATCCCTAAAGGAGAAGAGAAGGGCGTTTCTCCCGT
ACTAGAACACAGCACGGCAGAGCCGCCGTTTTGCTGAGTAAGAACTATGTTAGCAGCAAGTTAACTTCCCA
>NW004444309.1_3
ATTCGCCGCATGAGAGAGAAAACGAAACACAGCGTTCTCCTTCAAGCTGCTTAACTGAATCTACTGAAGTTTACAGT
TTTGTGAGAGGAAACAATCAGCCCACCACCAGCTGCTTTCTCCAAGGTTCTCTCAGTGGGAGAAGACTGGAGAGTG
CCTGCCTTGCAAGTATTTCAATCCCCTGGTTTTCAAAGGGAACATTACATGACAGGAATGACGGTATTACTAGCTTA
CAAAGCCAAAGAGGCAGATTGTTTGCCTTCAGGTAGAGGACAGCAGAAGAGGTTTAGTTTTGCTGACTAAACCTTG
TGTGCTTAACACGAGCATGTAGCTTCTTCCCTGCCTAAAAGAGAAAGCAAGAGCTTGCTGCTCTCAGTCTGGGGAT
GAAGGGAGAACGATTCACTTCCTGTGTTTGAAAATCCCTCTGTGTTCCATGCCTAAGAAAGTTGCCTGTGCATTGTC
TACCTGGATAAAGTGTATGACTTTTCTCCCTTTCAAGTAGTTGAAGCCGTTCAGTGTTTCCAGCATTCCGGGAGACA
TTCTGTGCTGCGGCTGAGAATACTGCTTTAAAGAACGTATCCCTAAAGGAGAAGAGAAACCGCGTTTCTCCCGT
ACTAGAACACACTGCAGAGCCGCCGTTTTGATGACTAAGAACTATGTTAACAGCAAGTGTAAGTTCCCA
>NW004444309.1_4
ATTCGCCGCATGAGAGAGAAAGCGAAACACAGCTTTCTCGTTCAAGCTGCTTAACTGAATCTACTGAAGTTTTCAGT
TTTGTGAGAGGAGACAATCTGCCCACCACCAGCTGCTTTCTCAAAGGTTCTCTCAGTGGTTTTCAAAGGGAACTTTG
CATAACAGACAGTGAAGGTACTACTGGATTTCAAAGCAAAACAGGCAGATTGTTTGCCTTCAGCTAGAGGATGGCA
GAAGAGGTTCAGTTTTGCTGACTACACCTTATGTGGTTAACGTGAGCATGTAGATTCTTCACGGGCAGAAAGAGAA
AGTAAGAGCCTGTTGTTCTCAGTCTGGGGATGAAGTGAGAAGGATTCAGTCCCAGTGTTGGAAAATCCCTCTATGT
TCCATGCCTAAGAAAGTTGCCTGTGCAGTGTCTACCTGGATAAAATGTATGACTTTTCTCCCTCTCAAGTAGTGGAA
GCCGTGCAGGGTTTACAGCATTCCGGGAGGCCTTCTGTGCTGCGCGGCTGAGAATACCGCTTTACTGACGCTATC
CCTAAAGGAGAAGAGAAACCGCGTTTCTCCCTTTCTGGAACACGGATGAGGCAAAACTGTAGTTGTGCTGAGTAAA
AAGTTTGTTACCAGCAAGTGTAAGTTCCCAT
>NW004444309.1_5
ATTCACCACTTAAAAGAGAAAGCAATATACAGAGTTCTCCTTCAAGCTGCTTAATGGAACATACTGAAGTTTAATCAT
TTCTTAGAAGAAACAGTCCGCCCAGGAGAAGCTGCTCTCTCAAAGGTTGTCTCAGTGGGAGAAGACTGGAGAATGC
CTGCCGTGCAAGTATTTCAATCCCCTGGTTTTCAAAGGGAACTTTGCATAACAGACAGTGAAGGTACTACTGGCTTT
CAAACCAAAACAGGCAGATTGTTTGCCTTCAGCTAGAGGATGGCAGAAGACATTCATTTGCTGACTACACCTTTTGT
GGTTAACGTGAGCATGTAGAATCTTCACGGGCAGAAAGAGAAAGCAAGAGCCTGTTGCTCTCAGTCTGGGGATGA
AGTGAGAAGGATTCACTTACTGTGGTTGAAAATCCCTCTGTGTTCTATGCCTAAGAAAGTTGCCTGTACAGTGTCTA
CCTGGATAAAGTGTATGACTTTTCTCCCTTTCAAGTACTCGAAGCAATTCAGGCTTTACACCATTCCGGAAAGCATC
CCATGCTGCGCGTCTGCGTATACCACTTTACAGAATGTATCCCTAAAGGAGAAGAGAAACCGCGTTTCTCCCGTAG
TAGAACACAGCACGGCCGAGCCGCCGTTTTGTTGAGTAAGAACTATGTTAGCAACAAGTGTAAGTTCGCA
>NW004444309.1_6
ATTGGCTGCATGAGAGAGAAAGCGAAACACAGCATTCTCCTTAAAGCTGCTTTACTGAATCTACTGAAGTTTTCAGT
TTTGTTAGAGGAAACAGTCCGCCCACCACCAGCTGCTTTCTCAAAGGTTCTCTCCGTGGGAGAAGACTGGAGAGTG
CCTGCCTTGCAAGTATTTGAATCCCCTGGTTTTCAAAGGGAACTTTGCATAACAGGGAATGAGGGTATTTCTGTCTC
TCAAAGCAAAACAGGCAAATTGTTTTCCTTCAGCTAGAGGATGGCAGAAGACATTCATTTGCTGACTACACCTTTTG
TGGTTAAAGTGAGCATGTAGAATCTTCACGGGCAGAAAGAGAAAGCAAGAGCCTGTTGCTCTCAGTCTGGGGATGA
AGTGAGAAAGATTCACTTCCAGTGTTGGAAAATCCCTCTATGTTCCATGCCTAAGAAAGTTGCCTGTGCAGTGTCTA
CTTGGATAAAGTGTATGACTTTTCTCCCTTTCGAGTAGTTGAAGCCCTTCAGGGTTTATAGCATTCCGGGAAGCCTT
CTCTGCTGCGCGGCTGCGAATACCGCTTTACAGAAGGTATCCCTAAAGGAAAAGAGAAACTGCGTTTCTCCCTTTC
TGGAACACAAACGAGGCAAAACCGTAGTTGTGCTGGGTAAAAACTTTGTCAGCAGCAAGTATAAGTTCCCA
>NW004444309.1_7
ATACACTGCATGAAAGAGAAAGCAAAGTACAGAGTTCTGTTTGAAGCTGCTTAACTGAATCTACTGAAGTTTTCAGT
CTTGTTAGAAGAAACCTTCCCCCCCACCAGAAGCTGCTTTCTCAAAGTTTGTCTCAGTGGGAGAAGCCTGGAGAGTG
CCTGCCTTGCAAGCTTTCGAATCCCCTGGTTTTCAAAGGAAAGTTTGCATAACAGAGAATGAAGGTACTACTGGCTT
TCAGAGCAAAACAGGCAGATTGTTTGCCTTCAGCTAGAGGATAGCAGAAGAGATTCAGTATTGTTGAGTAAATATTG
TGTGCTTAACTGGACCATGTAGCTTCTTCCCTGCCAGAAAGAGAAAGAAAGAGCGTGCTGCTCTCAGTCTGGGGAT
GAAGTGAGAATGATTCACTTCCAATGTTTGAAAATCCCTCTGTGTTTCATGCCTAAGAAAGTTCCCTGTGTAGTGTCT
ACCTGGATAAAGTGTATGACTTTTCTCCCTTTCAAGTAGTTGAAGTCATTCAGGCTTTACAGCATTTCGGGAAGCATT
CGGTGCTGCGCGGTGAGAATACCTCTTTACACACGGTATCTGTAAATGAGAAGAATAACCGAGTTTCTCCAGTTCT
GGAACACGAATCAGGCAAAAGCGTAGTTGTGCTGAGTAAAAACTTTGCTAGCAGCAAGTGTAAGTTCCCA
>NW004444309.1_8
ATTCACCACTTAAAAGAGAAAGGAAAATACAGCGTTCTCCTTTAAGCTGCTGAACTGAATCTACTGAAATTTTCAGTT
TTGTTAGAAGAAACAGTCAGCCCAGCAGAAGCTGCTTTCTCAAAGGTTCTCTCAATGGGAGAAGCCTGGAGAGGGC
CTGCCTTGCAAGTATTTCAATCCCTTGGTTTTCAAAGGGAAGTTTTCATAACAGGGAATGAAGGTACTTCTGGCTTT
CAAAGCAAAACAAGCAGATTGTTTGCCTTCAGCTAGAGGATGGCAGAAGAGGTTCAATATCGCTGAATAGAACTTG
TGTGCTTAAGGCGAGCATGTAGCTTCTTCCCTGCCTGAAAGAGAAAGCAAGAGCTTTTTGCTCCCTGTGTGGGGAT
GGAGGATTCACTTACTGTGTTTGAAAATCTCTCTGAGTTTACATGCCCAAGAAAGTTGATTGTGTATTGTCTACCTCG
TTAAAATGTATGACTTTTCTTCCTTTCAAGTAGTTGAAGCCGTTCAGGGTTTAGATCATTCCGGGAGGCATTCTTTGC
TGCGCGGCTGAGAATACCTCTTTACACACGGTATCCGTAAAAGAGAAGAATAACCGCGTTTCTCCCGTTCTGGAAC
ATGAATGAGGCAAAACCGTAGTTGTGCTGAGTAAAAACTTTGTTAGTTGCAAGTGTGAGTTCCCA
>NW004444309.1_9

TTTCACCACATGAAATAGAAAGGAAAATACCGCGTTCTCCTTGAACCTGCTTAACTGAATCTTCCGAAGTTTTCAGTT
TTGTTAGAAGAAACATTCCGCACACCTGAAGCTGCTTTCTCAAAGGTTCTCTCATTGGGAAAGACTGGAGAGGGC
TTTCCTTGCAAGTATTTGAATACCCTGGTTTTGAAAGGGAACTTTGCATAACAGGGAATGAAAGTACTTCTGGCTTTC
AAAGCAAAACAGGCAAATTGTTTGCCTTCAGCTAGAGGATGGCACAAGAGGTTCAGTTTGGCTGACTAATCCTTGT
GTGCTTCAAGCGAGCATGTAGCCTCTTTCCTGGAAGAAGGAGAAAGCAAAGCATTTTGCTGTCTGTCCTTGGGAAG
AAGTGAGAAGGATTCACTAACTGTGTTTGAAAATCCCTCTGTGTTCCATGCCTAAGAAAGTTGCCTGTGCATTCTCT
ACCTCCGTAAAGGGATCCACTTTACTCCTTTTCAATTAGGTGAAGTCATTCAGCGTTTACAACTTTCCAGGAGGCAT
TCTGTGCTCTGCGCTGAGAAAGCCGTCTTACAAACGGTATCGCTGAAGTAGAAGAGAAACCGCATTGCCCCCGTTC
TGCTATACGACCACGGCAGAGACGCAGTTTTGCTGAGTAAAGACTATGTTAGCAGCAAGTGTAAGTTCCCA
>NW004444309.1_10
ATTCACCGCATGAAACAGAGAGGAAATTACAGCGTTCTCCTTCAAGCTCCTTAACTGAATCTACTGAAGTTTCCAGT
TTTGTTAGAAGAAAGACTCAGACCAGCTGGAGCTGCTTTGTCAAATGCTCTCTCAGTGGGAGAAGACTGGAGAGTG
CCTGCCTTGCAAGTATTTGAATCCCCTGGTTTTCAAAGGGAATTTTGCATATCAGGGAATGAATTTACCTTTAGCATT
CAAAGCAAAACAGGCAGATTTGCCTTCAGCCAGAGGATGGCAGAAGAGGTTCAGTTTTGCTGACTAAACCTTGTGT
GTTTAAGGCGAGCACGTACCCTCGTCCCTGCCAGAAAGAGGAAGCAAATCCTGTCGCTGTCAGTCTGGGGGATGA
AGTGAGAAAGATTCAGTTACTGTGTTTCAAAATCCGTCTGTGTTCCATGCCTAAGAAAGTTGTCTCTGCATTGTCTAC
CTGGGTAAATGGATCCACTTTTCTCCTTTTCAAGTGGGTGAAGCCCTTCAGCATTTACAGCTTTCCCGGAGGCATTC
TGTGATGCACGCCGGAGAATGCCGCCTTACCGATGTAATCGCTGAAGGAGAAGAGAAAGGGCCTTTCTCCCGTTC
TGGTATACCACCGCGGCACAGCCGCAGTTTTGCTGAGTAAAGACTATGTTAGCAGGAAATGTAAGTTCCA
>NW004444309.1_11
ATTCACCGCACAAAACAGAAAAATACAGCGTTCTCCTTCAAGCTGCTTAACTGAATCTTCAGAAGTTTCCAGTTTTGT
TAGAAGAAACAGTCCGACCAGCACGAGGTGCTTTCTAAAACATTGTCTCAGTGGGAGAAGACTGGAGAGTGCTTGC
CTTGCAAGTATTTGAATCCCCTGGTTTTCAAAGGGAACTTCGCATATCAGGGAATGAAGGTACATTGGGCATTCAAA
GCCAAACAGGCAGCTTGTTTGCCTTCAGCGAGAGGATGACAGAAGAGGTTCAGTTTTGCTAACTAAACCTTGTGTG
CTTAACGCCAGCATGTAGCTTCTTCCCCGATAGAAGGAGAAAGCAAACAAGCTGCTGTCAGTCTTGGGGAAGAAGT
GAGAAGGATTCACTTAGTGTGTTTGAAAATCCGTCTGTATTCCATGCCTAAGAAAGTTGCCTGTGCATTGTCTACCT
GGATGAAGGGACCCACTTTTCTCCCTTTCAAGTAGGTGAGGCCATTCAGCGTTTACAGCTTTCCAGGAGGCATTCT
GTGCTGTGCGCCCAAGAATGCCGCCTTACAGAGGGTATCCCTAAAGGAGAAGAGAAAGCGCGTTTCTCTGGTTCT
GGTAAACAACGGCGTCAGCGCCGCAGTTTTGCTGAGTAAAGAACATTTTAGCAGCAAGTGTAGGTTCCCAA
>NW004444309.1_12
ATTGACTGCATGAAAGAGAAAGCAAAATACAGCTTTTCTCTTTCAAGCTGCTTCACTTAATCCACTGAAAGTTTTCAG
TTTTGGTAGAAGAAACAGTCCGCACACAACCAGCTGCTTTCTCAACCTTCTCTCCTTGGGATAAGAGGGAGAGTAT
CTGCTTTGCAAGTACTTGAATCCCCTGGTTTTCAAACTGAACTTTGCATAACAGGAAAAGCAGGTACCTTTGGCATT
CAAAGGGAAACAGGCAGATTGTTTGCCATCAGCTAGATGGTGGCGGAAGAGGTTCAGTTTTGATCTCTGAACCTTG
TGTGCTTAACGAGAGCATGTAGCCTCTTCCGTGCCAGAAAGACAAAGCAAGAGCCTGTTGCTCTCAGTCTGTTGGT
GAAGTGAGAAGGATTCACTCACTGTGTTTGAAAATCCCTCTGTGTTACATGCCTGATAAGAGTTGCCTGTGCATTGT
CTACCTGGGTAAACGGTTTGACTTTCATCCCGTTAGAGTTGGTGAAGCCCTGTGCGGTTTACAGCTCTCTGGGAAG
CGTTTTGTGCTGCATGGCCAAGAATGCTGCCTTCCAGACGGTATCCAGAAAGAAGAAGAGAAAGCGCCTTTCTCCC
GTTCTGGTATACGAGTGTTTCAGTGCGGCTCGTTTGCTGAGTAAAGACCATGTTAGCAGCAAGTGTAAGTTCCCG
>NW004444309.1_13
ATTCATTACATGAAATAGAAAGCAAACTACATCTTTCTCCCTCAAGCTGCTTCACTTAATCCACTGATTTTTCAGTTTC
GGTAGAAGAAAAATCCGCACGCCACCAGCTGCTTTCTCAAAAGTGCTCTCAGTGGGATTTAGACTGGAGAGTGTCT
GCTTTGCAAGAACTTGAATCCCCTGGTTTTCAAGGGGAACTTTGCATTACAGGGAAGAAGATAACATTAGTATCTGA
AAGATAACAGGGAGATTGATTGCATTCAGCTAGAGGAAGGCAGAAGCCGTTCAGTTTTGCTGACTAAAGCTTGTGT
GTTTCACATGAGCAGGTAACCTATTCTCTGCAAGAGAGAGAAACCCAAAGACCGTTGCTCTCACTCTGGGGGTTGA
AGTGAGAAAGATTGACTTACTGTGTTTGAAAATCTCTCTGTGTTACATACCTAAGAATCTTGCATTTGCATTGTTTAT
CTGAGTAAAGACTTCGACTTATATCCCGTTGGAGTACGTGAAGCCCTGTGGGGTTTAAATCTTTCTGGGAGGCACT
CTTTGCTGTGTGGCTGAGAATGCCTCCTTAAAGACGGTATCCCTGAAGGAGGAGAGAAAGTGCATTTCTCCCCTTC
TGCTCTACGGGCGCCGGCAGAGCTGTAGTTTTTCTAACTAAACTCTATGTTAGCAGTTCGTACATGTTCCCAG
>NW004444309.1_14
ATTCACTGCATGAAGGAGAAAGAAAACTACAATCTTCTCATTCAAGCTGGTTAACTTAACTGAATTTTTCAGTTTTGG
TAGAACAAACTGTCTTCACACCACTTGCTGCTCTCTCAAACATTCTCTCAATAGGATAAGACTAGCAAGTGTCAGCC
TTTCAAGTACTTTACTCCCCTGGTTTTCAAAGGGAACTTTGCATTACAGGGAATGAAGGTACCTTTCGCATTCTATGA
GAAACAAGCAGATTGCTTCCCTTCCAGTGGAAGATGGTAGAAGAGGTTCCGTTTTGCAAACGGAAGCTTGTATGCT
TTACACGATCGTGTTTCCTATTCCCTGCAAGAAAGAGAAACCAAAACACAGTTGCTCTCCGTCTGGGGTATGAAGTG
GGAAGGATTCACTTCCTGTGTTTGAAAATCCATCTGTGTCAAGTGGCTAAGAGAGTTGCCTGTGCATTGTCTACCTG
GGTAAAGAGTTTGAATTCTCTCCAGTTCGAGTAGGTGAAGACCTGTGGGGTTACAGCTTTCTGGGAGGCAATCTGT
GCCGCCCAGCAGAGAATGCCACCTTACAGACACTATCACTGAAGGAGAAGAGAAAGCACTTCTCTCCCTTTCAGGT
ATGTGACAGAGGCAGCGCTGTAGTTTTGTTAACCAAACACTACATTAGCAGCACTTGCAAGTTCCCA
>NW004444309.1_15
ATTCACTGCATGAAAGAGAAAGCCAATTACAGTGTTCTCCTTCAAGTTGGTTAACTTAATCAACTGATGTTTTGAGTT
TTTATAGAACAAAGAGTCCTCACACCACAAGCTGCTTCCCAAACATTCCCTCAATGCGATAATACTGGTAAGCGACT
GCCTTTCAAGTTCTTGAAACCCCTGGTTTTCAAAGGGAACTTTGTATTACAGGGGATGAAGGTACCTTTCACATTTTA
ACAGGCAGATTGATTGCCTTCAGCTGGAGGACTGCGGAAGCTGTTCAGTTTTGCTGACTAAAGCTTGTATGCTTAA
CAGGAGCATTATCTGTTCCCTGCAAGAAAGAGAAACCAAAAGACAGTTGCTCTCAGTCTACGGGATGAAGTGAGAA
AGATTCACTTACTGTGTTTGAAAATCCGTCTGTGTTACATGCCCAAGAGAATTGCCGATGCCTTGTCTAAGTGGATA
AAGAGTTTGACTTTTCTCCCCTTCGAGTATGTGAAGGCCTGTTGGGTTTACAGCTTTCCAGGAGGCACTCTGTGCT
GTGCCGCTGAGAATGCCGCCTTACAGTCGCTATCTCTGAAGGAGAAGGGAAACATCTTTCTCCTATTCTAGTATAC
GACAGCAGCAACGCTGTAGTTTTGCTAACTAAACCCATGGTAGCAGCAAGTGTGTGTTCCCAC
>NW004444309.1_16

```
GTTTACTGCATGAAAGAGAAAGCAACAGACAGTGGTGACACTCAAGCTGGTTAATTTAATCAACTGAAGGTTTCGGT
TTTGGTAGAACAAAAAGTCCTCACACCACCAGCTGCTTTCTCAAACATTATGTCAATGGGATAAGCCTGGCGAGTGT
CTCCCAAGTACTTGAATTCCCTGGTTTTGAAAGGGAAGATTGCATTACAGGGAATGATTACCTTTCACATTCTATGAA
AAACAGGCTGTTTGTTCACCTTCAGCTAGAGGACCGCAGAAGCGGTTCAGTTTTGCTGACTAAAGCTTGTGTGCCT
CACATGAGGATGTATCATATTCCCTGCAAGAAAGCGAAACCAAAGACAGTTGCTGTCATTCTGGCGGATTAAGTG
AGAAGGATTCACTTACTGTGTTTGAAAATCCATCGTGTTACGTGCCCATGACTGTTATCTTTGCATGGTCTGCCTG
GGTGAAGAGTTTGACTTTTCTCCCCTTCGAGTCCGTGAAGCCTTACGGGGTTTACACCTTCCTGGAGGCACTCCGT
GCTGCGCAGCAGAGAATGCTGCCTTACAGACGGTATCCCTGAAGAGGAGAAAGCGCATTTCTCCCATTCTGCT
ACAGGACAGGGGCGGCACTGTAGTTTTATTAACTAAACCCTATGGTAGCAGCACGTGCATATTCCCAC
>NW004444309.1_17
GTTCACTGCATGAAAGAGAAAGCCAACTACAGTCTTCTCATTCAAGCTGGTTAACTTAATCAACTTAAGTTTTCAAAT
TTGGTAGAACAAAGAGTCCTCACACCACCAGCTGCTTTCTCAAACATTCTCTCAATGGGATAAGATTGACGAGTGTC
TGCTTTACAAGTACTTGAATCCCCTCTGTTTCAAAGGGAACTTTGCGTTACAGGGAATGAAGGTACATTTTGCATTCT
AAGAGAAACAGGCAGATTGTTTGCCTCCAGCTAGATGATGGCAGAAGCAGTTCCGTTTCCCTGAGTAAATCTTGTC
TTGTCAGCACGTGAATGTTTCCTATTCACTGCAGAAATGAGAAAACAAAAGATAGTTGCCCTCAGTCCAGTGCATGA
ACTGAGAAGGATTCACTTACTGTCTTTGGAAATCCCTCTGTGTTACTTGCCCAGGACTGTTCCGTGTGCACTCTTTA
CCTGGGTAAAGACTTCCACTTTTCTCCATTTCGACTACATGAAGCTCTGCGGGTTTCCAGCTCTCTGGGAGACACTC
TATGCTGTGCAGCCAAGAATGCAGCCTTACAGCCGGTATCGCTGAGGGAGAAACAAAGCGCATTTCTCCCATTCT
GGCATACGACAGCAGTAGAGATGTAGTTTTGTAAAGTAAACACTATGTTAACAGGATGTGCATGTTCCCA
>NW004444309.1_18
ATTTACCGTATGGAAGAGACAGCAAAATAGAGTGTTCTCATTTAATCTGATTAACTTAATCAAGCTAAGCTATCAGTT
TTGATAAAACACAGTCATCACACTATCTACAAGCTTTCTCAAACATTCTGTCAGTGGAATAGTACTGGCGAGTCTCT
GCCTTTCAAGTACCTGAATCCTGAGGTTTTCAAAGGGAAATTTTCATTACAGTTAATGAAGTAACTTTCTCATTCTAT
GAAGAACAAGCAGATTATTTTCCCTTCAGCTATAGGATGGCGAAAGTGGTTCAGTTTTGCTGACTAAACTTTATGTT
GTAAGCACGTGCATGTTTCCTATTCACTGCAAGAAAGGGAAACCAAAATACAGTTGCTCTCATTCTCGAGGACGAAT
TGAGAAGGATTCACTTACTGTGTTTGAAAATCCCTTTGTGTTACATGTTCAGGAATGTTGCCTGTACATTCTCTACCT
GGGTAAATCGTTAGACCTTTCTTCCCTTCGACTACGTGAAGCCCTGCAGGGTTTATAGCTTTCCACATTGTGCTGTG
TGGTTGAGAATGCAACCTTACAGCCTGTATCCCTGAGAGAGAAAAAAAAAACGAGTTTCTCCCTTTTTGGCATAGGA
TAGTAGTAGGAATGTATTTTTGCTAAGTAAACACTATGTTAGCAGGATGTGTATG
>NW004444309.1_19
GTTCAGTGCATGAAAGAGAAAGCAAGATGCAGTGTTCTCACTCAAATTGGTTAACTTAAACAATTTAAGTTATCAGTT
TTGATAAAATACAGTCCTCACACCATCTATAAGCTTTCTCAAACATTCTGTCAATGGGATAAGACTGGTGAGTCTCTG
CCTGCAAGTACCTGAATCCCTTGGTTTTCTAAGGGAACATTGCTTTACAGGGAATGAAGGTATTTTTCGCAGTATAC
AAAGAAGATTGTTTCATCTTCAGCTAAAGCATGGCAGAAGGGGGTCAGTTTTTCTGACTAAACCTTGTGTTGTAACC
ACGTGCATGTGTTCTATTCACTGGAAGAACGAGAAACCAAAGACAGTTGCTCTCAGTCTGGTGAATGAAGTGAGA
AGGATTCACTTACTGTGTTTGAAAGTGCCTGTGTTACATGCCCAAGAATGTTGCCTATGCATTCTCCTCCTGGGTAA
AGAGTTTGACTTTTCTCCTCTTGGAGTACTTGAAGCCCTGCAGGGTTTACAGCTTTCCGGGAGGCTCTCTGTGCTG
TCTGGCAGAGAATGAGGCCTTACAATCGGTATCCCTGAGGGAGAAAACAAATCGCGGTTTCCCCCGTTGTGGCATA
TGACAGTATTACAGATGTAGTCTCGGTAAATAAACACTATGTTAGCAGGAAGTTCATGCTTCCA
```

Supplementary Sequence File 2. TMAsat monomeric sequences (in fasta) manually isolated from the *Dugong dugon* assembled reference genome (accession GCA_015147995.1).

```
>BMBL01009774.1_1
ATTCACCACATGAAAGAGAAAGCAAAATACAGCGTTCTCCTCAAGCTGCTTAACTGAATCTACTCAAGTTTTCAGTTT
TGTTAGAAGAAAGTGTCTGCCCAGCACAAGCTGCTTTCTCTAATGTTCCCTGCGTGGGAGAAGGCTGGAGAGTGCA
GGCCTAGCAAGTATTTGAATCCCCTGCTTTTGAAAGGGAATTTTTCCTACCAGGAATGAAGGTACTAATGGCTTTCC
AAGCCAAACAGGTAGATTGTCTGCCTTCAGCTAGAGGACAGCAGAAGAGGTTCAGTTTTGCTGACTAAAGCTTGTT
TGCTTAACGGGAGCATGTAGCTTCTTCCCTGCCTGAAAGAGAAAGCAAGAGCTTGTTGCTCTCAGTCTGGGGATGA
ACGGAGCACGATTCACTTTCTGTTTTTCAAAGTCCCTCTGTGTTACATGCCTAAGAAAGTTGCCTGTGCATTGTCTA
CCTGGATAAAGCGTATGACTTTTCTCCCTTTCAACTACTTGAAGCCGTTCAGGCTTTCCAGCATTCCGGGAGGCATT
CTGTGCTGCGTAGCTGAGAATAGGGCTTTACAGACGCTATCGCTAAAGGAGAAGAGAAACCGCATCTCTCCCATTC
TGGAACGCGAAGGAGGCAAAACCGGAGTTCTCCTGAGTAAAAACCTTGTTAGCAGTAAGTTTAAGTTCCCT
>BMBL01009774.1_2
ATTCACCATATGAAAGAGAAAGCAAAAGACAGGCTTCCCTTCAAGCTGCTTAACTGAATCTACTTAAGTTTTCAGTTT
TATTAGAAGAAAGAGTCCGCCCAGCACAAGCTGCTTTCTCCAATGTTCTCTGAGTGGGAGAAGGCTGGAGAGTGCA
GGCCTAGCAAGTATTTGAATCCCCTGGTTTTCAAAGGGAATTTTGCATACCAGGGCATGAAGGTACTATTGGCTTTC
AAAGCCAAACAGGTAGATTGTCTGCCTTCAGCTAGAGGAAGGCAGTAGAGGTTCAGTTTGCTGACTAAAGCTTGT
GTGCTTAACAGGAGCATGTAGCTTCTTCCCTGCCTGAAAGAGAAAGCAAGAGCTTGATGTTCTCAGTCTCGGGACG
AAGGGAGAATTATTCACTTTCTGTGTTTGAAAATCCCTGTGTGTTACATGACTAAGAAAGCTGCCTGGGCATTTTCTA
CCTGGATAAAGCGTATGACTTTTCTCAATTTCAAGCAGTTGAAGCCGTTCAGGTTTTCCAGCATTCCGGGAGTCATT
CTGGTCTGCGCGGTTGAGAATACCGCTTCACAGGCGGTATCCGTAAAGGAGAAGAGAAACAGCGTTTCTCCCATTC
TGGAACGCGAGGGAGTCAAAACTGTAGTTGTGCTGAGTAAAACCTTTGTTAGGAGCAAGTGGAAGTTCCCC
>BMBL01009774.1_3
ATTCACCGCATGAAACAGAAAGCGCACTACAGCGTTCTTCTTCAAGCTGCTTAAGTGAATCGACTGAAGTTTTCAGT
TTTGTTAAAAGGAACCGTCCGAAGAGCACCAGCTTCTTCCTCAAACGTTGCCTCACTGGGAGAAGTCTGGAGAGTG
CCTGCCTTGCAAGTATTTGAATCCGGTGGTTTTCCAAGGGAACTTTGCATATTAAGGAATGAATGGACTACTGGCTC
```

TCAGAGCCAAACAGGCAGATTGCCTGCCTTCAGCTAGAGGAAGGCAGAAGCGGTTCAGTTTTGCTGACTAAACCGT
GTGTGCTTAACACGAGCATGTAGCTTCTTCCCTGCCTGAAAGAGGAAGCAAGAGCTTTTTGCTCTCAGTCTGGGGA
GGAAGGGAGAACAATTCCCTTTCTGTGTTTGAAAATCCCTCTGTGTTACATGCCTAAGAAAGTTGCCTGTGCATTGT
CTACCTGGATGAAGTGTATGACTTTTCTCCCTTTCAAGTAGTTGAAGCTGTTCAGGGTTTCCAGCATTCCGGGAGGC
ATTCTGTGCTGCGCGGCTGAGAATACCGCTTTACAGACGGTATCTCTAAAGGAGAAGAGAAACCGCGTTTCTCCCA
TTCTGGAACCCGAGGGAGGCAAAACCATAGCTGTGCTGAGTAAAAGCTTTGTTAGGAGCAAGTGGAAGTTCCCC
>BMBL01009774.1_4
ATTCACCACATGAAAGAGAAAGCAGAAGACAGCACTGTCCTTCAACCTGGTTAACTAAATCTACTGAAGTTTTCAGG
TTTGTTAGAAGGAAGTGTCCGCCCAGAACAAGCTGCATTCTTCAAATGTTCTCTCAGTGGGAGAAGACTGGAGAGT
GCGGGGCTAGCAATTATTTGAATCCCCTTTTTTTCCAAGGGAACTTTGCATACCAGGGAATGAAGGTACTACTGGCT
TTCAGATCCAAACAGCCAGATTGTTGCCTTCTGCTAGAGGACAGCAGAAGAGGTTCAGTTTTGCTGACTAAACCTTG
TGTGGTTAATAGGATCCTGTATCGTCTTCCCTGCATGAAAGTGAAAGCAAGATCTTGTTGCTCTCAGTCTGGGTAGG
GAGAACAGTTCACTTTCTGGGTTTGAAAAACCCTCTGTGTTACATGCCTAAGAAACTTGCCTGTGCATTGTCTACCT
GGATAAAGTGTATGACTTTTCTCCCTTTCAAGAGGTTGAAGCCATTCAGGGTTTCCAGCATTCCGGGAGGAATTCTT
TGCTGCGCGGCTGAAAATACCGCCATACAGGCAGTATCGCTAAAGGAGAAGAGAAACCCCGTTTCTCCCATTCTGG
AACGCGAGGGAGTCAAAACCGTTGTTGTGCTGAGTAATAACTTTGTTAGCAGGAAGTGGAAGTTCCCC
>BMBL01009774.1_5
ATTTAGCACATGAAAGAGAAAGCAAAATACAGCGTTCTCCTTGAAGCTGCTTAACTGAATTTACTGATGTTTTCAGTT
TTGTTAGAAGGAAGAGTCCATCCAGCACAAGCTGCTTTCTCCAATGTTCTCTGAGAGGGAGAAGGATGGAGAGTGC
GGGCCTAGCAAGTGTTTGAATACCCTGGTTTACAAAGGGAACTTTGCATATCAGGGAATAAATGTACTGCTGGCTTT
CAGAGCTGAACAGGCAGATTGTCTGTCTTCAGCTAGAGGACAGCAGAAGAGACTCAGTTTTGCTGACTAAATCTTG
TGTGCTTAACACTAGCCTGTAGCTTCTTCGCTGCCTGAAAGTGAAAGCAAGAGCTTGTTGCTCTCAGTCTGGGGAC
GAAGGGAGAAGGATTCACTTTCTGTGTTTGAAAATCCCTCTGTGTTACATGCCTAAGAAAGATGCATGTGCTTTGTC
TACCTGGATAAAGTGTATGACTTTTCTCCCTTTCAAGAGGTTGAAGCCGTTCAGGTTTTCCAGCATTCTGGGAGGCA
TTCTGTGCTGTGTAGATGAGAATAGGGCTTTACAAGCGGTATCGCTAAAGGGGAAGAGAAACCGCGTTTCTCCCAT
TCTGGAACGCGAGGGAGTCGAAACTGTATTTGTGCTGAGTAAAAACTTTGTTAGGAGCAAGTGGAAGTTTCCC
>BMBL01009774.1_6
ATTCACCACATGAAAGAGAAAGCAAGATACAGCGTTCTCCTTCAAGCTGCTTAACTGAATCTACTGACGTTTTCGGT
TTTCTTAGAAGAAAGGGTCTGCCCAGCACAAGCTGCTTTGTCAAACGTTCTCTGAGTGGGAGAAGACTGGAGAGTG
AGGGCCTAGCAAGTGTTTGAATCCCCAGGTTTACAAAGGGAACTTTGCATAGCAGGGAATGAATGTACTACTGGTT
TCAGAGCCAAACAGGCAGATTGTCTGCCTTCAGCTGGAGGAAGGCTGAATACGTTCAGTTTTGTTGAGTAAACCTC
TTGTTCTTAACATGCGTCTGTAGCTTCTTCCCTGTTTGACAGAGAAAGCAAGACCTTGTTGCTATGACTCTGGGGAT
GAAGGGAGAAGGATTCACTTTCTGTGTTTGAAAATCCCTCTGTGTTACATGCCTAAGAAAGTTGCCTGTGCATTGTC
TACCTGGATAAAGCGTATGACTTTTCTCCCTTTCAAGTAGTTGAAGCCGTTCAGGGTTTCCAGCATTCCGGGAGGC
ATTCTGTGCTGCGCGGCTGAGAATACCCTTTACAGACGGTATCGCTAAAAGAGAAGAGAAACCGCGTTTCTCCCTT
TCTGGAACGCGAACGAGGCTTAACCGCAGTTGTGCGGAGTAAAGACTGTTAGCAGCAAGTGCCAGTTGCCC
>BMBL01009774.1_7
ATTCACCACATGGAAGAGAAAGAAAACTACAGCGTTCTCCTTCAAGCTGCTTAACTGAATCTACTGCAGTTTTCAGT
TTCTTTGGAAGAAAGTGTCGGCCCAACACAAGCTGCTTTCTCCAATGTTCTCTCAGTGGGATTAGGCTGCAGAGTC
CGGGCCTAGCAAGGATTTAAATCCCTTGGTTTTCAAAAGGAACTTTGCATACCAGGGAATGAGGGGACTACTGGCT
TTCAACGCCAAATTGGCAAATTGTCTGCCTTCAGTTAGAGGACGGCAGAAGAGGTTCAGGTTTGCTGACAAAACCG
TGTCTGCTTAACAGGAGCATGTAGCTTCTTCCCTGCCTGAAAGAGAAAGCGAGAGCTTGCTGCTCTCAGTCTGTGG
ACAAAGGGAGAATGATTCACTTTCTGTATTTGAAAATCCCTCGGTGTTACATGCCTAGGAAAGTTGCCTGTGCATGG
TCTACCTGGATAAAGGGTATGACGTTTCTCCCTTTCAAGTAGTTGAAGCCGTTCAGTGTTCCTACCATTCCTTGAGG
CTTTCTGGGCTGCGCGGCTGAGAATACCGCTTTACAGACGGTATCGCTAAAGGAGAAGAGAAACCACGTTTCTCCC
ACCTGGAACTTGAAGGAGGCAAAATCGTAGTTGTGCTGAGTAGAAACTTTGTTAGCAGCAAGTGTAAGTTCCCC
>BMBL01002074.1_1
ATTCACCACATGAAAGAGAAAGCAAAACACAGCGTCTTCCTTCAAGCTGCTTAAGGGAATGAACTGAAGTTTTCAGT
TTGGATAGAAGAAAGTTTCTGCCCAGCACAAGCTGCTTTCTCAAATGTTCTCTGAGTGGGAGAAGGCTGGAGAGTG
CGGGCCTAGCAAGTATTTGAATCCCCTGGTTTTCAAAGGGAACTTTGTATACCAGGGAATGAATGGAGTACTGGCT
TTCAAAGCCAAACAGGCAGATCGTCTGCCTTCAGCTAGAGCACGGCAGAAGAGGTTCGTTTTTGTTGACTAAACCG
TCTGTGCTGAACACGAGCATGTAACTTCTCCCCTGCCTGAAAGAGAAAGCAAGAGCTTGTTGCTCTCAGGATGGGG
ATCTATGGAGAATGATTCAATTTCTCTGTTTGAAAATCCCTGAGTGTTACATGCCTAAGAAAGTTGCCTGTGCATTGT
CTACCTGGGTAATGTGTATGACGTTTCTCCCTTTCACGTAGTTGAAGCCGTTCAGGGTTTCCAGCATTCCATAAGGC
ATTCTGTGCTGCGTGGCCCAGAATTCGGCTTTACAGACGGTAACGCTAAAGGAGAAGAGAAACCGCGTTTCTCCCA
TTCTGGAACGCGAAGGGGGAAAACCGTAGTTGTGCTGAGTAAAAACTTTGTTAGCAGCAAGTGCAAGTTGCCC
>BMBL01002074.1_2
ATTCACTACATGAAAGGGAAAGCAAAATACGGCGTTCTCCTTCAAGCTTCTTAACAGAATCTACTGAAGTTTTCAGTT
TTGTTGGAAGAAAGAGTCCGCCCAACACAAGCTGCTTTCTCAAGAGTTCTCTGAGTGGGAGAAGGCTGGAGAGTG
CTGGCCTAGCAGGTATTTGAATCCTCTGGTTTTCAAAGCGAACTTTGCATCCAGGGAATTCAGGTACTACTGGCTTT
CAACGCCAAACTGGCAGAGTTTCTGCTTTGAGCTAGAGGACGGCAGAAGAGGTTCAGTTTTGTTGACTAAATCTTG
TGTGCTTAACACGAGCATGTAGCTTCTTCCCTGCCTGATAGAGAATGCAACTGCTTGTTGCCTTCAGTCTGAGGATG
AAGGCTGAAAGATTCACTTTCTGTGTGTGACAATCCCTCTGTGTTACATGCCTAAGAAAGTTGCCTGTGCATTGTCT
ACCTGGATAAAGTGTATGACGTTTCTCCCTTTCAAGTAGTTGAAGCCGTTCAGGGTTTCCAGCATTCCGGGAGGCA
TTCTGTGCTGCGCGGCTGAGACTACCCGTATAACCACGAAAGAACTGGTCTGCCAATCTGCTTCCTTTAAGATTACA
ACCAAGAAAACTCTATGGAGCAC
>BMBL01012666.1_1
ATTCACCGCGTGAAAGAGAAAGAAAAATACAGCATTCTCCTTCAAGCTGCTTAACAGAAGCTACTGAAGTTTTCAGG
TATGTTAGAAGAAACAGTCCTCCCACAACCAGCTGCTTTCTCAAACTTTCTCTCAGTGGGCTAAGACTGTAGAGTGC
TTGTCTTGCAAGTATTTGAATTCCCTGGTTTTCAAAGGGAAATTTGCATCCCAGGGAATGAAGGTGGTTTTGGCATT

CAAAGCAAAGCAGGCAGATTGTTTGCCTTCAGCTAGGGGACGGCCGAATTGTTTCAGTTTGGCGGAATAAACCTTG
TGTGCTTAACGCCAGAAAGTAGCTTCTTCCATGCCTGAAGGAGAAAGCAAGAGCCAGTTGCTGTGAGTCTGGGGAA
GAAGTAAGGAGGAGTCACTTACTGTGTTTGAAAATCCCTCCGTGTTACATGCCTAAGAAAGTTGCCTGTGCATTGTC
TACCTGGATAAAGCGTATGACTTTTCTCCCTTTCAAGTAGTTGAAGCCGTTCAGTCTTTCCAGCATTCCTGGAGGCA
TCCTGTGCTGCGAAGCTGAGAATAGGGCTTTACAGACGGTACCGCTAAAGGAGAAGGAAACCGCGTTTATCCCATT
CTGGAACGCGAAGGAGGCAAAACCGTAGTTGTGCTGAGTAGAAATTTTGTTAGCAGCAATTCTAAGTTCCCC
>BMBL01012666.1_2
ATTCACCACATGAAAGACATGGGAAAATACAGCGTTCTCCTTCAAGCTGCTTAACTGAATCTACTTAAGTTTTCAGTT
TTGTTAGAGGAAAGAGTCCGCCCAGCAAAACTGCTTTCTCCAATGTTCTCTGAGTTGGACAAGGCTGGACGGTGCA
GGCCCAGCAAGTATTTGAATCCCCTGGGTTTCTAAGGGAAGTTTGCATACCAGGGAATGAAGGTACTGTTGGCTTT
CAAAGCCAGACAGGTAGATTGTGTGCCTTCAGCTAGAGGACAATAGAAGAGGTTCAGATTGCTGACTAAAGCTTG
TGTGCGTAACACGAGCATGTAGCTTCTTCCCTGCCTGAAAGAGAAAGCAAGAGCTTGTTGCTCTCAGTCTGTGGAT
GAGAGGAGCACGATTCACTTTCTGTGCTTGAAAATCCCTCTGTTTTACATGTCTAAGAAAGTTCCTGTGCATTGTGT
ACTTTGATAAAGTGTACGACGTTTCTCCCTTTCAAGCAGTTGAAGACGTTCAGGGTTTCCAGCATTCCGGGAGGCAT
TCTGTGCTGCGCTGCTGAGAATACCGCTTTACAGACGGTATCGCTAAAGGAGAAGAGAAACGGCGTTTCTCCCATT
CTGCAAGGCTAAGGAGGCAAAACCGTACTTCTCCTGAGTAAACATTTGTTAGCAGTAAGTTTTAGTTTCCT
>BMBL01099827.1_1
ATTCACTGCATGAAAGAGAAAGCAAAATACAGTGTTATCATTCAACCTGCTTAACTTAATCCACAGAAGTTTTCAGTT
TTGGTAGAACAAACAGTCCTCACACCACCAGCTGCTTTCTCAAACGTTCTATCAGTGGGATAAGACTGTCGAGTGTC
TGCCTTGCAAGTACTTGAATCCTCTGGTTGTCAAAGGGAACTTTGCATTACAGGGACCGAAGGTACTTTTGTCATTC
AAAGTAAAACAGACAGATTGTTTGCCTTCAGCTAGAGGACAGCAGAAGCAGTTCACTTTTCCTGATGAAAGTTTGTG
TGCTCAACAGGAGCATGTACATTATTCCTTGCAAGAAAGAGAAAGCAAAAGACAGTCGCTCTCAGTCTGGGGGATG
AAATGAGAAGGATTTACCTACTGTGTTTGAAACTCCCTGTGTGTTACAAGCCCAAGAAAGTTGCCTGTGCATTGTCT
ACCAGAGTAAAGAGTTTGAGTTTTCTCCCAGTCGAGTTCGTAAAGCTCTGCGGGATTTACAGCTTTCTGGGGGGGCA
CTCTGTGATGCATGGAGAGAATGCCGCCTTACAGACGGAATCCCTGAAGAAGAAAGAAAGTGCGTTTCTCTCATTC
TGGTATACGACAGCAGCAGCGCGGTAGTTTTCCTAAGTAAACACTATGTTAGCAATACTTGCATGTTCCGA
>BMBL01099827.1_2
ATTCCCAGTATGAAAGGGAAGGCAAAATACAGTGTTCACATTCAAGCTGGTTAAGTTAAGCAACTTAAGTTTTCAGT
ATTGGTAGAACAAACGGTCCTCACAGTACCAGCTCATTTCTCAAACATTCTCTCAGTGGAATAAAACTGGCAAGTGT
CTGCCTTTCAAGTACTTGAATCCTCTGCTTTTCAAAGGAACTTTTCATTACAGGGACTAAAGGTAGCTTTTGCATTCT
AAGCAAAACAGGCAGACTGTTTGCCTTCAGCTAGAGGATGTCAGCAGGGGTTCAGTTTTGCTGACTAATCCTTGTG
TGCTTAACACAAGCATTTCTTATTCCTTGCAAGAAAGAGAAAGCAGAAGACAGATGCTCTCAATCTGGGGGATAAAG
TGAGAAGAAGTTACTCACTGTGTTTGAAAATCCCTCTCTGTTACTGGCCCAGGACAGCTGCCTGTGCATTGTATACC
TGGATAAGAGTTTGAATTTTCTCCCATTGGAGTTCGTGAAGCCCTGCGGGGTTTACAGCTTTTCAGGAGGCACTCT
GTGCTGCGAGGCCGAGAATTCCGCCTTACAGAGGGTATCCCTAAAGGAGAAGAGAGCGCTTTTCTCCCGTCCTGG
TATACGACAGTGGTAGCGCTGTAGTTTTGCTGAGTAAAGATTATGTTAGCAGCAAGTGCAAGTTCCCA
>BMBL01099827.1_3
ATTCAATGCACGAAAGAGAAAGCAAACTACAGTGTTCTCCTTCAAGCTGGTTAACTTAACCCACTGAAGCTTTCAGT
TTTGGTAGAATACACAGTCATAACACAACCTCCTGCTTTCACAAAATTTCTCTCAGTGGGATAAGACTGTCGAGCGT
CTTCCTTGCAAGTATTTGAATCCACTGCGTTTCAAAGGGAACATGCAGTACAGGGGATGGAGGTACCTTTGGCATT
CTAAGCAAAACAGGTGGATTGTTTTCCTTCAGCTAGAAGACAGCAGAAGCAGTTCAGTTTTGCTGAGAAAACCTTGT
GTGCTTAACACGAGCAGGTATCCTATTCCCTGCAAGATAGAGAAAGCAAAAGACAGTTGGTCTCAGTCTGGGTGAT
GAAGTGAGGAGGACTCAATTACTGTGTTTGAATCCCTTGGTGTGAAATGCCTAAGAAAGTTGCCTGAGCATTATCTA
CCTGGGTCAAGAGTTTGACTTTTCTCCCGTTTGTGGGGTTTACAGCTTTCCAGGAGGCCTTTTGTGCTGCGCAGAC
GAGAATCTACCCTTACAGACGGTATCACTAAAGGAGAAGAGAAATTGGGTTTCTCCCATTCTGGTATATGAAAGCG
GCAGCATTGTAGTTTTGCTGAGTGAAGACTATGTTAGCAGCAAGTGTAAGTTCTCA
>BMBL01099827.1_4
ATTCACTTCATGAAAGAGAAAGCAAAATACAGCATTCTCCTTCAAGCTGCTTCACTTCATCCACTGAAGTTTTCAGTT
TTGGTAGAAGAAACAGTCTGCACACCACCAGCTGCTTTCTCAAAGTTTCTCTCAGGGGGATGAGACTGGAGACAGC
CTGCCTTGCAAGTGTTTGAATCTCCTGGTTTTCAAACGGAACTTTGCATAACAGGAAATGAAGGTAATTTTAGCACT
CAAAGCCAAATAGGCAGATTGTTTGCCTTCAGCTAGAGGATGGCAGAAGAGGTTCAGTTTTGCTGACTAAACAGTG
TGCTTAACATGACCATGTAGTTTCTGCCATGCCAGAAAGAGGAAGCAAAAGCCAGTTGCTCTCAGTGTGGGGGATG
GAGTGAGAAAGATTCACTTACTGTGTTTGAAATCCCTCTGTGTTACATACCTAATAAAGTTGCCGGGGCATTGTCTA
CCTGGGTAAAGGGTTTGACTTTTCTCCTGTTTGAGTAGGTGAGCCCTTACAGGGTTTACAGCTTTCTGGGAACCATT
CTGTGCTGGGCAGCCGAGAATGCCGCCTTACAGACGGTATCCGAAAGGAGAAGAGATACCGCGTTTCTCCTGTT
CTGTTGTACACCAGTAGCTCTGTAGTTTTGCTGAGTAAAGACTACGTTAGCAGCAAGTGCAAGTTCCCA
>BMBL01099827.1_7
ATTCATTGCCTGAAAGAGAAAGCAAAATACAGTGTTCTCATTCAAGCTGGTTAACTTAATCAACTGACATTTTCAGTT
TTGGTAGAAGAAACTGTCCTTACAGCACCAGTGTGCTTTCTCAAACATTCTCTCAATGGGATAAGAGTGGCAAGTGT
CTGCCTTTCAATTACTTTAATTCCCTGGTTTTCAAAGGGAGCTTTGCATTACAACGAATGAAGGTAACTTTCGCATTC
TAGGAGAACTGATTGTTTAACCTTCAGATATTTGACAGCAGAGGCGGTTCTGTTTTTCTGACTAAACCTTGTGTTGTA
ACCACATGCATGTTTCCTATTCACTGCAACAAAGAGAAACTAAACAACAGTTGCTGTCAGTCTGGTTGATGAACTGA
GAAAGATTCACTTAATATGTTCGATAGTCCCTTTGTCTTACATGTCCAAGAATGTTGCCTGTGAATTCTCTTCCAGGG
TAAACAAAGCATTTCACTTTTCTCTCCTTAGAGTACACGAAGCCTTACGGAGTTTGCAGCTTTCCCCGGGGAACTCT
GTGCTGGGCGCTCAAGAATGTGGCCTTAGAGACCTTATTCCTTAAGGAGAAAACACAAAGCCCATTTCTCCCGTTC
TGATACATGATCGCAATAGAGACATAGTTTTGCTAAGGAAACACTACAGTAGCAGAGCCTGCATGTTTCCT
>BMBL01099827.1_8
TCACTGCATAAAAAAGAAAAATCAGAATTCAGTGTGCTCATTCAAGCTGGTTAACTTAATCAACTTATTAGTTTTGAG
AAACACAGTTCTCCCACCACCTGCAAGATTTCTCAAACATTCTCTTGATGGGATATGACTGGCGAGTCTCTCGTTTT
CCAGTACTTGATTCCCCTGCTTTTCAGAGGGAACTTTGCATTACAGGGAATGAAGTTACCTTTTGCATTCTCAGAAG

AACAAGCAGATTGTTTCACCTTCAGGTAAGGGATAGCAGAGGCGGTTCAGTTTTACTGACTAAACCTTGTGTTTTAT
CTACATGCCTGTTTCTTATTAACTGCAACAAAGAGAAACCAAACGACAGTTGCTCTCATTCTGGTGGATGAAGTGAG
AAGGATTTACTTAATGTGCTTGAAAATCCCTCTGTGTTACATGCCTAAGAATGTCGCCTATGCATTGTCTGCCTGGG
TAAAGCATTTCACTTTTCTACCCTTGAAGTACATGAAGCCCTGCGGGGTTTACAGCTTTCCGGAAGGCACTCTGTTG
CGTGGCTGAAAATGTGGTCTTATAGACAGTATGTCTGAATGAGAAACAAAATGTGTTTCTCCTGTTCTGGTATACG
ACTTCAGCAGAGATGTAGTTTTGCTAAGGAAACACCACGTTAGCAGGACCTGCGTGTTCCAA
>BMBL01099827.1_9
TTTCACTTCATGAGAGAGAAAGCAGAGTACAGTGTGTTCATTCAAGCTGGTTAACTTAATCAACTTATCAGTTTTGAG
AAAACACAGTCCTCCCACCATCTGCAAGCTCTCTCATTCTCTCAGTGGAATAAGACTGGCTTGTGTCTGGCTTTCCA
GTATTGAACTCCCTGCTTTTCAAAGGGAAGTTTGCATTACAGGGAAAGAAGGTAGCTTTCACATTCTAGGAAGAACA
AGCAGACTGTTTGGCCTTCAGCTAAAGACTGCAGAGGGGGTTCAGTTTTTTGCATATTTCCTATTCACTGCCGCAAA
GAGAACCAAATGACACTTGCTCTCAGTCTGGTGGATTAAGTGAGAAGGATTCATTTACTGTGTCTGAAAATCTCTCT
GTATCACATGCCAAGGAATGTTGCCTATGTATGCTCTACCAGGGTAAAGAGATTGACTTTTCTCCCCTTCGAGTACG
TGAAGCCCTGCAGGATTTACAGTTTTCAGGGAGGCACTCTGTGCTGCCCGTTCAGAAAGCGGCCTGACAGAAGGT
ATCTCTGAATTAGAAAAAAAGTGCCTTTCTCCTGTTCTTATATACGACAGTAGTAGAGATGTAGTTTTAGTATGTAAA
GACTATGTTAGCAGAACTTGCATGTTCCCA
>BMBL01099827.1_10
TTTACAGCATGAAAGAGAAACCAAAAATATATTGTTCTCATTCAAGTTTGTTAACTTAATCAATTGAATAAGCTATCAG
TTTTGATAAAACACTGTTCCCACACAATCGAGAAAATTTCTCATACATTCTCTGAATTGGAAATGAATGTCCAGCTTC
TCTCTTTCAAGAACTTGTATCCCCCGCATTTTCAGATGTTTTCTAAAGGAGTTTTGAGTTATAGGGAAATGTGTGTGC
ATTTCACATTCTAGGAAGAACAATGCAGACTGTTTCCTGTTCAACTAAAAGATGGCAAAACCGATATAGTTTTGCTGA
CTGAATCATGTGTGGTAACCACGTGTGTGTTTCCAATTCACTACAAGAAAGAGAAACCAAATTCAGTTTTCTCATTC
AGATTGGTTAATTGAATCTAGTTGATTAAGTTGTCAGTTTTGATAAAACACACTTCCCACAGTATCCACAACCTTCCT
CATACGTTCTCTGAATGGGAAATGAATGTCATGTTTCTCCCTTTCCAGTACTTGAATCCTCCAGGTTTTCAGACTTTT
CTTATGGGAACTTTGCATTACTGGGAATGGATTTGCCTTTCACAATCTAGGAAGAACAAGGCAGATTGTTTCCCATT
CATCTAAAGATGGCAAAAACATTGTACTTTGCTGACTGAACCCTGTGTGGTAAGCACTTGTATGTTTCCA
>BMBL01099827.1_14
ATTCACTGCATGAAAGCGAAACCAAAACTGTTTTCTCATTCTAGTTGATTAACTTAATCTGGTTGATTAAGTTATCTGT
TTTGATAAAACACTCTTCCCATACGTTCTATGAATGGGAAAAGAATGTCACGTTTCTCTCTTTCAAGTATATGAAGCC
TTTGCGTATTTAAGTGTTTTCTATGAGAACTTCTAGTTTCCGTGAATGAATGTGTCTTTCACATTCTAGGAAGACCAA
AGCGGACTGTTTCCCATTCAACTGCAATAAGCCAAAGGAGATGCAGTTTTGCTGACTAAACCTTGTAACCACGTGAA
TATTTCAAATTCACTGCATGAAAGAGAAACCAAATGCAGTTTTCTCATTCTAGTTGATTAACTTACTCTAGTTGGTTA
AGTGGTCTGTTTTGATAAAACACTCTTCCCACACATTCTGTGAATGGGAAAAGAATGTACAGTTTCTCCCTTTCAAGT
AAATGAGGACTTCAGGTATTCTGATGTTTTCTATGAGAACTTTGGGTTACCAGGAATGAATGTGTCTTTCACTTTCTG
GGAAGAACAAGGCAGATTGTTTCCCGTTCAATTACAAGACAGCAAAGGTATGTAGTTTTGCTGACTAAACCTTATGT
TGTAAGCATGTACATGTTTCCA
>BMBL01099827.1_15
ATTCACTGCATGAAAGAGAAATCACAATACAGTTTCTTTCATTCTAGTTGATTATGTCAGAGGGTATTATCTGTTTTTG
ATGAAACACTGTTCCCACACGTTCTCTCAATGAAAAAAGAATGTCAACTTTCACACTATCAAGTACGGGAGTCTTCT
GGGTTTTCAGATGTTTTCAATGGGAACTTTGCGTTACCGGAAATGAACGTGACGTTCACATTCTAGGAAGACAAAGG
CAGATTGTTTCCAGTTCAATTAAAGGACAGCAAAGACAATGTTGTTTTGCTGACTAAACCTTGTGTTGTAACTGCATG
CATATTTCCAATTCGCTGCACGTAGGAGAAACCAAAATATTTTCTCTCATTCTGATTGATTAAGTCAGAAGGATTAAC
TTATTGTGTTTGAAAATAACTCTGTGTTACTTGCTGGAGAATGTTGCCTATACATTCTCTACCAGGGTGAAGAATTTA
ACTTTTCTCCCATTCAAATACATGGAGCCATATGAGGTTTAGATGTTTCCAAGAAGCACTCTGTGCTGCCTGGCCAT
AAATGTGCTTTACAGACAAAATATCTGAAGGAGAAACAATTCGAGTTTCTCCCGTTCAAGTATAGGACGGTAATGG
AGAAGTAATTTTGCTAACTAAACACTGTCACCGGAAGTGCACATTCCCA
>BMBL01099827.1_16
ATTCACTTCATGAAAGTGAAAAAAAATACAGTTTTCTCATTTTAGTTGATTTAACTTAATCTAGTTGATTAAGTTGTCT
GTTTTGTTAAAACACTGTTCCCACAGGATCAAAAACATTTCTTATATATTCTCTGACTGGGAAAAGAATGTCAAGTTT
CTTCCTTTCAACTACGTGATGCTTCTGGGTTTTCAGATGTTTTCTATGGGAACTCTGTGTTACCAGAAATTAATGTGC
CTTTAAAACTCTAGGAAGAAGAAGGTATGTTGTTTCCTGTTCAACTACGGGAAGGCAAAAGCGATGTAGTTTTGCTG
AGTAAACCTTGTGTTGTAAGCACGTGCATGTTTCCAATTCAGTGCATGACAGAGAAACCAAAATACAGTTTTCTCATT
CTAGTTGATTAATTTAATCTTGTTTTGAGAAAACACTGTTCTCGTTCTCTGGATGGGAAAGAATACCAAATTTCTCC
CTTTCAAGTACATGAAGCCCCAGTGTTTTCAGATATTTTCTATGGGAACTTTGGGTTATGGGGAATAACTTGGCCTTT
CACATTCTAAGAAGGACAAGGCAGATTGTTTCCTGTTCAACTACAGGTCAGCGAAAGCCATGTAGTTTTGCTAAGCA
AACACTGTGTTACCAGAAAGTGCATGTTCCCA
>BMBL01099827.1_17
ATTCACCACATGGAAGGGAAACCAAAATACAGTTTCTCTTGTTCTAGTTGATTAAGTCAGAAGGTGTTATCAGTTTGA
TAAAACACTGTTCCCACATGACCGAGAACCACTTTTATACCTTCTCTGAATGGGAAAAGAATGTCAAGTTTCTCCTTT
TCAAGTATGTGAAGCCTCTGAGTTTTCAAATTTTTTCAATGGAAACTTTGGGTTATGGAGAATGAACATGCTTTTCAC
ATTCTAGGAAGAACAAGGCAGATTGTTTCCTGCTCACCTACAAAACTGCAAAATGATATACTTTTGCTGACTGAAA
CTTTTGTTGTAAACACGTGCATGTTTCCAATTCACTGCAAGAAAGAGAAACCAAATACAGTTTCTCATTCTGGTTGA
TTAAGTCAGAAAGATTAAATTACTGTGTTTGAAAACTTCTGTGTTACTTGCACGAGAAGGTTGCCTATACTTTCTCTA
CCAAGGTAAAGAATTTAACTTTTTTTCCCAATCTAACTGGAGAAGCCACACGGGATTTAGATGTTTTGAAGCAGTATTC
TGTGTTTCCCAGGCTTGTATGTGCCTTAGGGACAAAGTCTCTGAACAAGAAAACAATGCGAGTTTCTTCCATTCAAG
TATAGGATGGCATTAATGATGTAGTTTTGCTCAGCAAACATTATGTTACCAGACTGAGCACTTTATCA
>BMBL01099827.1_18
ATTCACTGTATGAAAGAGAAACCAAAATACAGTTTCTCTCATCTGCTTGATTAAGTCACATGGTATTATCTGTATTGA
TAAAACTCTGTTCCTACATGATGGAGGATCACTCTTATACATTTTCTGAATTTGAAAATAATGTCAAGTTTCTGACTTT
CTAGTATGTGAGTGTTTTGGTTTTCAGGAATGAATGTGCCTTTCATAGTCTATGAAGAACAAGGGAGATTGTTTCCC

GTTCAACTACAGGACAGCAAAAACTATGTAGTTTTACTAATTAAACCTTGTGTTGTAAGCACGTGCATGTTTCTAATT
CATCACATGAAAGATAAACCAAAACACAGTTTCTCTCATTCTGGTTGATTATGTCAGAGCGCTGACTTACTATGTTTG
AAAACACTCTGTGTTTCATGCATGAGAATGGTTTGTATGCATTCTCCACCTGGGTAAATAATTTAAATTTTGTCCCTT
TCAAATAGATGAAACCATGCAGTGTTTAGATGTTTTCAAGAAGTACTCTGTGCTGCCTGACCATGAATGGGCCTTAC
AGATAGAATCTCTGAACCAGAAACAATGCAAGTTTCTCCTGTTCAAGTATAGGACAACAATGTAGTTTTGCTAAGTA
AATACTCTGTTACCAGAATGTGCAAGTTCCCA
>BMBL01105564.1_1
ATTCAGTGCATGATATAGAAAGCAAAATACAGCGTTCTCCTTCAAGCTGGTTAACTTAATCCACTGAAGTTTTCAGTT
TTGTTAGAAGAAACTGTCTGCCCACCACCAGCTGCTTTCTCAACGGTTCTCTCAGTGGGATAAGACTGGAGATTGC
CTGCCTTGCAAGCATTTGAATCCCCTGGTTATCAAAGCTATCTTTGCCTAACAGGGAACGAAGGAACTTTTGGGATT
CAAAGTCAAAGAGGTAGATTGTTTGCTTTTAGCGAGAGGATGTTAGAAGAAGTTCAGTTTGGCTGAGCAAAGCTTGA
GTGCTTCACGTGAGCATGTAGCCTCTTCCCTGCCAGAGAGAGGAAGCAAAAGCCAGTTGCTCTCAGTCTGGGGGA
TGAAGTGAGAAGGATCCACATACTGTGTTTGAAAATCCCTCTGTGTTACATGCCCAAGAATGTTGCCTGTGCATTGT
TGACCTGGATAAAGAGTTTGACTTTTCTCCCTTTCGAGTACATGAAGCCCTGCTATGTTACAGCTTTCTGGGAGGC
ACTCTGTGCTGCCCAGATGAGAATGCCGCCTTGCCGGAGGTATCCCTGAAGGAGGGGAGAAAGCGCGTTCTCCCC
TTCTGCTATACAACAGTGGCAGCACTGTAGTTTTGCTAATTAAGCACTATATTAACAGCACGTACATGTTCATA
>BMBL01105564.1_2
ATACAGTGCCTGAAGGAGAAAGCAAAATACACTGCTCTCGTGGAAGCTGGTTAACGTTTCTGTTTTGGTAGAAGAAA
CAGTCCTCACACCACCAGCTGCTTCCTCAAACATTCTCTCAATGGGATAAGACTGGCGAGTGTCTGCCTTTCAGGT
ACTTGAATCCCCTGCTTTTCAAAGGCAACTTTGCATTACAGGGAAGGAAGGTAACATTCGCATTCTAAGAAAAACAG
GCAGATTGTTTGCCTTCAGCTAGAGGACCGCAGAAGCAGTTCAGTTTTGCTGACTAAAGCTTGCTTGGTTAACACG
AGCATGTATCCTATTCCCCGTAAGAAATAGAAACGGAATGGCAGTTGCTCTCATTATGGGGGAATGAAGTGAGAAA
GATTGACTTACTGTGTTAGAAAATCCCTCTTTATTATATGCTCAGAAATGTTGTCTGTGCATTGTTTACCTGGGTAAA
GACTTTGACTTTTCTCCCTTTCCAGTACATGAAGCCCTGCGGGGTTTACAGCTTTCCGGGAGGCACTCTGTGCTGC
GCCACCGAAAATGTAGCTACTGAGTCTTTAAGAAAGAACTGTTGAGTATTATCGTATTATAAAATGATGTATATAAT
TCACACCATGATTTT
>BMBL01043776.1_1
ATTCACTGCATGAAAGAGAAAGTAAAATAGTGTTCTCATTCAAGTTGGTTAACTTAATCAAGTGGATTGTTATGAGTT
TTGATAAAACACTCTCCTCACCCCATCCACAAGCTTTCTCAAACATTCTCTCAGTTGGAATGATGAGACTCTCTCTTT
CAAGTACTTGAATCCCTCAAGTTTTCAGAGGGAGCTCTGTGTTACCCGGAATGAAGGTACCTTTCACATTTTAGGAA
GAACAAGGCAGATTGTTTCACTTTCAGCTAAAAGACAGCAGAAACAGTGCAGGTATGCTGACTAAACATTGTGCTGT
AAGCACTTTCATGTTTCCAATTCACCACAAGAAAGCAAACCAAAAGACAGTTGCTCTCAGACTGGTAGATGAAGTC
AAAAGGATTCACTTACTATGTTTCAAAATCCCTCTGTGTTATGAGCCCAAGAATGTTGCCTATGCATTCTTCACCTGG
ATAAAGAGTTCGGATTTTCTCTCCTTCGAGTACGTGAAGACCTGTGGGCTTAAGAGGTTTCTGGGAGGCCCTCTGT
GCTGCCCAGCCAAGAACGCAGCCTTATAGATGGTATCTCTAAAAGAGAAACAATGTGCGTTACACTTATACAGGTA
TAAGAGAGCAGTAGACATGTGATTTGCTAAGTAAACACTATGTTAACAGGGCATGCATGTTCCCA
>BMBL01043776.1_2
ACTCACTGCAAGAAAGAGAAAGCAAAATACAGTGTTCCCATTCAAGTTGGTTAACATAATCAAGTAGATTAACTTATG
AGGTTTGATAAAACACACTCCTCACACCATCCACAAGCTTTCTCAAACATTCATTCAGTGGGATAAGAATGGTGAAT
GTTGCCCTTTCAAGTACTTGAATCCGCCAGGTTTTCAGATGTTCTGTAACGGAGCTGTGTGTTAGCCTGAATGCAG
GTACTTTCCGCATTCCAGGAAGGACAAGGCAGATTGTTTCCAGTTCAGCTAAGGGACGATAAAAGCAGTGCAGGTT
TGCTGAGTAAACCTTCTGTTGTGAGGACTTGCTTGTTTCCAATACACTGAAAGAAAGAGAAATCAAAAGATGGCTGC
TCTCAGTCTGGTGGATGAAGTCAGAAGGATTCACTTACTGTGTTTGAAAATCTCTCTCTGTTCCATGCCCAAGACTG
TTGCCTGTGCATTCTCCACCTGGGTAAAGAGCTCGACTTTTCTCCCCTTTGAGTACGTGAAGCCCTGCAGGGTTCA
AGGCTTTCTGGTGGGCACTCTATGCTGCACAGCCAAGAAACCAGCCTTACACAAAATGTCTCTGAAGGGGAAAAGA
AAGCAAATTTCTCCCATTCTGGTATACCGCAGCAGTAGAGATGAAGTTTTGCTAAATAAACACTATGTTAGCAGGAC
GTGCATGTTCTCA
>BMBL01043776.1_3
ATTTCCTGCAGGAAAGAGAAAGCAAAATACAATGTTCTCATTCAAGTTGGCTAAGTTCAGTTCATCAAGTTATCAGTT
TTGATAAAACACAGTCCTCACACCATCGAGAAGATTTCTCAAACATTGTTTCAGTGGGATAATTCTGGTGAGTCTCTT
GCTTTCAAGTGATTGAATCCCCTGGTTTTCAAGGGGGACTTAACAGAGAATGAAAGTACCCTTTGCATTTAGGAAAA
ACAAGTTAGATTCTTTTGCCTTCAGCTAAGGCATGGCAGAAGCAGTTCAGTTTGCCTGACTAAACCTAGTGTTGTAA
GCATGTGTAAGTTTCCTATTAACTGCAAGAAAGACAAAGCCAAAGACAGTTGCTCTCAGTCTGCTAGATGAAGTGG
GGATTCGCTTACTTTGTGTTGGAAAGTCCCTCTGTGTTACACACCCATGAATGTTGTCCAGGAATTCTCTACCTGGG
TAAAGACTTTGATTTTTCTCCCCTTCCAATACGTGAAGCACCGCAGAGTTTACAGCTTTCCGGGAGGAACTCTATGCT
GCCCGGCTGAGAGAGTGGCCTTACACAGCGAGTCTCTGAAGGGGAAAAGAAAGCGCGTTTCTCCCATTCTGGTAT
ACCACAGCAGTGGAGATATAGTTTGGAGAAGTAAACACTACGTTAGGGGGAAGTGCATGTTCCCA
>BMBL01043776.1_4
ATTCACCCCAGGAAAGAGGAAGCAAAATACGGCGTTCTCCGTCAAGCTGCTTAACTGAATCTACGGAAGTTTTCAG
TTTTGTTAGAAGAAACAGTGTGCCGACCAGTAGCTGCTTTCGCAAAGGCTCGCTCAGTGGGAGAATACAAGAGAGT
GCCTGCCTTGCAAGTATTTGAATCCCCTGGTTATCAAAGCTATCTTTGCCTAACAGGGAACGAAGGAACTTTTGGGA
TTCAAAGTCAAAGAGGTAGATTGTTTGCTTTTAGCGAGAGGATGTTAGAAGAAGTTCAGTTTGGCTGAGCAAAGCTT
GAGTGCTTCACAGGAGCGTGTAGCCTCTTCCCTGACAGAAAGAGGAAGCAAAGCCAGTTGCTGTCAGTCTGAGGG
ATGAAGTGAGAAGGATTCACTACTGCGTTTGAAAATCCCTCTGTGTTCCATGCCTAAGGAAGGAAGTTGCCTGTGC
ATTGTGTACCTGGGTAATGGGATCCACTTTTCTCCCTTTCAAGCAGGTGAAGCCATTCTGGGTTTAAAGCATTCCTG
ATGACATTCTGTGCTGCTGGGCCGTGAATGCAGCTTTACAGACGCTATCCCTAAAGGAGAAGAGAAACCTTTCTCC
AGTTCTGGTATACCACCGTGGCAGAGCCGCCGTTTTGCTGTGTCAAAAGTATGTTAGCAGCACTTATAACTTCCCA
>BMBL01006510.1_1
ATTCACCGTATAAAAGAGAAAGCAAAATACAGCGTTCTCCTTCAAGCTTCTTTACTGAATCTACTGAAGAAAGAGTC
CACCCACCACCAGCTCTTTGCTCAAATGTTCTCTCTCAGTGGGAGAAGACTGGAGAGTGTCTGCCTTGGAAGCATT

TGAATCTCCTGATTTTCAAAGGGAACTTTGCATAACAGGGAATGAGTTACTTCTGGCATTCAAAGCGAAACAGGCAG
ATTGTTGGCCTTCAGCTAGAGGAAATGCAGAAGAGGTTCAGTGTTGCTGACTAAACCTTATTTGCTTAATGTGACCA
CGTAGCTTATCCTTGCCAGAAGGAGAAAGCAAAGCAAGTTGCTGTGAGTCTGGAGGATGAAGTGAGAAGGAGTCA
CCTACTGTGTTTGAAAATTCCTATGTGTTACATGCCTAAGAAAGTTGCCTCTGCATTGTCTCCCTCGGTAAAGGGAT
CCACGTTTCTCCCTTTCAAGTATGTGAAGCCATTCAGAGTTTACAGCTTTCGAGGAGGCATTCTGTCCTGCGCGGC
CGAGAATGCCACCTTTCAGACGGTATCACAGAAGGGAGTGAAACCGTGTTTCTCTCGTTCTGGTATAGGACAGCGG
TGGAGCCACAGTTTTGCTGAGTAAAGACTATGGTACCAGCAAGTGTAAGTTCCCA
>BMBL01061436.1_1
ATTCAGTCTATTAAAAATGAAGGCAAAATGCTGTGTTCTCATTCAAGCCTGTTTCTTTAATGAACTGAAATTTTCACTT
TTGTTAGAACAAGCAGTCCTGTCACCATCAGCTTCTTTGTGAAACATTCTCTGAATGGGTGTAACTGGCCTTTCAAG
TACTTGGATCCTGTGGTTTTCGAATGGAAAGTTTAATTATAGAGATTGAAGGAACACTTTGCATTGAAAGGAGAACA
GTCAGATTGTTTGCCTTCACCGAGAGGACAGCACAAGCAGCTCAGCTTGTCTGACTAAAGCTTGTGCGGTTAATAT
CAGCATGTATCCTATTCTGTTCAAGAAAGAGAAACCAAAAGACAGTTGCTCTCAGGCTGTGGGAATGAAGTGAGAA
GGATTCACTTACTGTGTTAGAAATTTCCTCTGTGTTACATGCCCAAGAACTTTCCCTGTGCTTTTTTGCCCTGGGTAA
AGTGTTTGACTTTTCTCCCTTTCGAATACGTTTATCCCTGCGGGGTTTACCGCTTTCCGTGTGGCTCTCTTTGCTGC
GCTGCGGATAATGCCGCCTTGCATTCCGTATCGCTGAAGGGAAGAAAAAGCGGATTTCTCCCATTCCCATATACGG
CAGCAGCAGTGCTGTACTTTTGCTAACTAAACACTATGTTAGCAACATTTGCATGCTCCCA
>BMBL01061436.1_2
ATTCACTGTATGAGAAAGAAAAATACCGTGTTCTCATTCGACCCTGTATACTTAATGAACTGAAGTTTTCCGTTTTGG
TAGAACAAGCAGCCCTGACACCACCAGCTACTTTCTGAAACACTTTCTCAATGGGATAAGACTGACCAGTTTCTACC
TTTCAAGTACTTGAGTCCTCTGGTTTTCAAATGGAACTTTGCCTTACAAGCAAAATGATAACTTTCGCATTGTAAGG
AAAATAGGCAGATTGTTTACCTTCAGCTAGAGGATAGCAGAAGCAGTTCAGGTTTGCTAACTAATGCTTGTGTGGTT
AACATGAGCATGTTTCCTATTCCCTGCAAGTAACAGAAACCTAAAGACTCTTGCTGTCAGTCTGGGTAATGAAGTGA
GATGGAGTCACTTATTGTGTTTGAAAATCCCTCTCTGTTACAGGCCCAAAAATGTAGCATGTACATTGTCTACCTGT
GTAAAGAGTTTGCCTTTTCTCCCCTTCGAGTACGTGAAGCCCTGCGGGGTTTACAGCTTTCAGGGAGGCCCTCTCT
GCTGCACGGCGGTGAATGCCTTCTTACTGACATCTTCACTGAAGAAGAGAAACTCGTTTCTCCAGTTCTGGTATAT
GAGAGCGGCAGCAGTGTAGTTTTGCTAACTAAACACTGTGGAAGAAACACGTGCATGTTTCAC
>BMBL01061436.1_3
TTCTCTGCATGAAGAAGAAAGCAAAATACCGTGTTCTCATTCAAGCCTGTTTACTTAATCAACTGAAGATTTTCAGTT
TCGGTAGAGCCAACGGTCGTGACATAACCAGACACTTTCTGAAATATTTTCTCAATGGGATAAGACTGGTGATTGTC
TGCTTTTCCAGTACTTGAATCCCCTGGTTTTCAAATGGAACTTTGCCTCACAGGCAAAGAAGGTAACTCTCATTCTAA
GGAAAATAGGCAGATTGTTTGCCTTCAGCTAGAGGACAGTAGAAGCAATTCAGTTTTGCTTACTAAAGCTTGTGTGG
TTAACTTGAGCATGTTTCCTATTCCCTGCAAGAAAAAGAAACCAAAAGACAGTTGCTCTCAGTCTGAGCAATGAAGT
GACAAGGTTTCACTTACTTTGTTTGAAAAATCCCTCTGTTTTACATGCCCAAGAAAGTTGCCCCTGCGTTGTTTACTG
GGGTAAAGAGTTTGACTTTTCTCCCATTCGACCACGTGAAGACCTGTGGGGTTTACTGCTTTACAGGAGGCACTCT
GTGCTGCTAGGGAGAAAATGCCACTTTACAGACTCTATGTGTGTGGGAGAAGAGAAACGCATTTCTCCCTTTCAG
GTATATGACAGCAACAGCTGTGTTGTTTTGCTAACTAAAGACTATGTTAGTAGCACATGTATCTTCCCA
>BMBL01061436.1_4
AATCAATGTATGAAAAAGGAAGCTAAGTACAGTGTTCTCACTGAAGCCTGTTTACTTAATCAACTGAAGTTCTCGGTT
TTGGTAGAGCAAACAGTCCTGACACCACCAGCTACTTTCTGAAACATTCTCTCATTGGGATAAGACTGACCAGTGTC
TACCTTTCTAGTACTTGAATATCCTGGTTTTCAATTGAACTTTGCACTGTAGGGAATGAACATTTTGCATTAAAAAAAA
ACAAGCAGATTGTTTGCCTTCTGCTAGAGAACGACAGAAACAATTAATTGTTGCTTACTAAAGCTCATGTGGTTAAC
ATGAGCATGTTTCCTATTCCCTGCAAGAAAGACAAACCAAAAGACAGGTGCTCTCAGTCTGGGGAATGAAGTGAGA
AAGATTCACTTTGTTTGAAAATCCCTCTATGTTACATGCCCAAGAAAGTTGCCTGTGCGTTGTCTACCTGGGTAAAG
AGTTTGACTTTTCTCCCCTTTGAGTACTTGAAACCCTGAGGGGTTTACAGCTTTCTGGGAGGCACTCCGTGCTGCA
CGGCCGAGAATGCTGCCTTACAGATGTTATCGGTTGAAAGGGAGAAGCTTCATGTTCTTTTCACACTCAGATAATGT
ATAAGAAAAGTTTTTCACCG
>BMBL01061436.1_5
ATTTCCTGCATGAAAGAGAAACCAAAATACAGGCTTTTCATTTTAGTAGTTTAGCTTAATCAAGTTGATTAATTTATAT
GTTTTGATAAAACACTGTTGCCCACGCCATCGAAACCATTCCTACACATTCTCTGGGAAAACAATGTCAAGTTTCTC
CCTTTCTCGTACTTGAAGCCCCTGGGCTTTCAGATGTTTCTATGGGATCTCTGGTTTACCGAGAACCAATATGCCTT
TCACATTCTAGGAGAAAGAGGCTAATTGTTTCCCATTCAACTACAGAATGACAAAAGCGATGTGGTTTTGCAAACT
AAAGCTTGTGTTGTCTACAAGCACGTTTGTAATTCACTGCTCGAAAGAGAAATCAAATATTGTTTCTTTTATGCTGCT
TTAGTAAGTCTGAAAGATTAACTTCCTGTGTTTAAAAAACCCTCTGTGTTACATGCATGAGAGTGTTGTCTTTACTTT
CTCTACCAGGGTAAAGGGTTGAACTTTTCTCCCTTTCAAATACATGAAACCCTACGGGGTTTAGATGTTTTCAAGCA
GTACTTGGTGCTTTCCGTTAGTGAATGTGCCTTACATACAGAATCTCTGAAGGAGAAAACAATGCTTGTTTCACCTG
TTCAAGTACAGGACCACAGCAGACATGCAGTTTCTCTAAGTAAACACTATGTTACCAGAACGTGCGTGTTCCCA
>BMBL01061436.1_6
ATTCACTGCATGAAAGAGAAACCAAAATACAGGTTTTTCATTTTAGTAGTTTAGCCTAATCAAGTGATTAAGTTATCT
GTTTTGATAAAACGCAGTTGCCACAAGATAGAATACCATTCTTACACATTTTCTGAATGGGAAAGCAATGACAGTTTT
CTCCCTTTTTGGCATTTGATGCCTCTGGGCTTTCAGATGTTTTCTATGGGCTCTCAGGTTTACCAGAAATGAATGTG
CCTTTCAAATTCTAGGAAGAACAAAGCTAATTGTCTCCCATTCAACTACATAACGGCAAAAGCGATCTGATTTTGGT
GACTAAATCTTGTTTCATTCATCCACATGTTTCTAATTCACTGCTCAAAAGAGAAACCAAAATACTGTTTCTTTCATAC
TGCTTGAGTAAGTCAGAAGGATTAACTTACTGCGTTTGAAAACTCCTCTGTGTTAAATGCACAAGAGTGTTGCCTAT
ACATTCTCTACCAAGGTAAAGTGCTGAACTTTTCTCCCATTCAAATACGTGAAGCCATATGGGGTTTAGATGTTTTCC
AGCAGTACCCTGTGCTTTCTATTCGTGAATATGCCTTATAGACAGAATCTTTGAAGGAGAAAACAATGTGTGTTTCTC
CCATTCAAGTACAGGATGGCGACACAGATGTAGTTTCATTAAGTAAACATGGTGTTACCAGAAAGTGCATGTTCCCA
>BMBL01061436.1_7
ATTCACAACATGAAAGAGAAAGCAAAATGCAGTTTCTCTCATTCAGCTTGAGTATCTCAGAAGGATTAACTCACTGT
GTTTGAAAAGCCCTCTGTGTTACATGCATGAGAATGTTGCCTATACATTCTCTACCAGAGAACTTTTCTCCCATTCAA

ACACGTGAAGCCATATGGGGTTTAGATGCTTTCAAGCAGTACTCTGTGCTTTCTGTTCATGAACGTGCCTTACAGAC
AGAATCTCTGAAGGGGAAAACAATGCATGTTTCTCCGGTTCAAGTACAAGACGACAATAGAGATGCAGTTTCACTAA
GTAAACACTGTGTTATCAGAATGTGCATGTTCCCAATTCACTACATGAAAGAGAAAGCAAAATACAATTTCTCACATT
CAGCTAGAGTAAGTCAGAAGAATTAAATTACTGTGTTTGAGAAACACTGTGTTACAGGCATGAGAACATTGCCTATA
TTTTCTATATGTGGGTAAAGAATTGAACTTTTCTCTCACTCAAATACGTGAAGCTATGCAGACTTCATATGTTTTCAA
GTTGCACTCTGTGCTGCTCGGCTGTGAATGTGCCTTACACACAGAATCTCTGACGGAAAAATAATGGGAGTTTCTC
CCGTTCAAGTATAAGACAGTAATAGCGACGTCGTTCTGCTAAGCACTGTGTTACCAGAAAGGGCATGTTCCCA
>BMBL01082959.1_1
ATTCACTGCCTGAAAGAGAAAGCAAAACAGTGTTCTCTTTCAAGTTGGTTAACTTAATCAATTGATTAAGTTATCAGT
GTTTATAAAAACACAGTTCCTGCACAATCCACAAACTTTCCCATGCATTCTCTAAATGGAAAAGAATATCCAGCTTC
TACCTTTCAAGTACTCAAATCCCCCGAGTTTTCAGATGTTTCCTAAGGGAACTTTGCATTACCGGGAATAAAGGTGG
GTTTCACACTGTAGGAAGAACAAGGCAGATTGTTTCCAGTTCACCTAAAGGATGGAAAAAGTGATGTAGTTTTACTG
ACTAAAGCTTGTGTTGTAAGCATGGGCATGTTTCCCATTCACTGCAAGAAAGAGAAACCCAAATACAGTTCCTCTCA
TTGTGGTGGAATAAGTCAGAAGGATTAAGTTATTGTGCTTGAAAATCCCTGTGCTACATGCACAAGAATGTTGCTTA
TGCATTCTCCACCTGGGTAGAGTTTAACTTTTCTCACATTCGAGTACGTGAAGCCATGTGGCATTTAGATGTTTCCA
GTAGGCGCTCTGTGCTGCCCGGCCTCAAATGGGCCTTACAGACAGAATTTCCAAAGGCGAAAACAATGCGAGTTTC
TGCCGTTCAAGTATAACGCAGCAACAAAGATGTCATTTCGCTAAGGAAACACTATGTTTAGCAGAACTTGCACGTAC
CCA
>BMBL01082959.1_2
ATTCACTGCATGAAAGAGGAAGCAAAATACAGTGTTCTCATTCAAGTTGGTTAACTTAATCAATTGATTAAGTTATCA
GTTTTGATAAAACACTGTTCCCACATGATCAATAAACTTTCTCATAATTCTCTGAATTGGAAAAGAATGTCAAGGTTC
TCCCTTTCAAGTACTTGAATCCCCCAGGTTTTCAGATGTTTTCTAAGGGAACTTTGCGATACCAGGAATGAATGTGT
TTCACATTCTAGGCAGAACAAGACAGATTGTCTCCCATTCAACTAAAGGATGGCAAAAGCAAGGTAATTTTGCTGAC
TAAACCTTGCGGTTTATGCTCAGGCATGTTTCCCATGCACTGAAAGAAAGAGAAACCAAATTACAGATTCTCTCATT
CTGCTGGAATAAGTCAGAAGGATTAACTTACTGTGGTTGAAAATCTCTTTGTGTTACTTGCAAAAAATGTTGCCTATA
CATTCTCCACCTGGGTAAAGAGATAAGCTTTTCTCCCATTTGAGTACGTGAAGGCATGTGGGGTTTAGGTGTTTTCA
GTAGGCACTCCGTGCTTCCCAGCCATGAATGTGCCTTACAGACAGTATCTCTAAAGCAGAAACAATGCGAGTTTC
TCCCATTCAAGTATAGGACAGCAATAGAGATGTAGTTTTGCTAAGGAAACACTATGTTACCAGCATGTGCACGTTCC
CA
>BMBL01041892.1_1
ATTCACTGCATGAAAGAGAAAGTAAAATAGTGTTCTCATTCAAGTTGGTTAACTTAATCAAGTGGATTGTTATGAGTT
TTGATAAAACACTCTCCTCACCCCATCCACAAGCTTTCTCAAACATTCTCTCAGTTGGAATGATGAGACTCTCTCTTT
CAAGTACTTGAATCCCTCAAGTTTTCAGAGGGAGCTCTGTGTTACCCGGAATGAAGGTACCTTTCACATTTTAGGAA
GAACAAGGCAGATTGTTTCACTTTCAGCTAAAAGACAGCAGAAACAGTGCAGGTATGCTGACTAAACATTGTGCTGT
AAGCACTTTCATGTTTCCAATTCACCACAAGAAAGCAAAACCAAAAGACAGTTGCTCTCAGACTGGTAGATGAAGTC
AAAAGGATTCACTTACTATGTTTCAAAATCCCTCTGTGTTATGAGCCCAAGAATGTTGCCTATGCATTCTTCACCTGG
ATAAAGAGTTCGGATTTTCTCTCCTTCGAGTACGTGAAGCCATGTGGGCGCTTAAGAGGTTTCTGGGAGGCCCTCTGT
GCTGCCCAGCCAAGAACGCAGCCTTATAGATGGTATCTCTAAAAGAGAAACAATGTGCGTTACACTTATACAGGTA
TAAGAGAGCAGTAGACATGTGATTTGCTAAGTAAACACTATGTTAACAGGGCATGCATGTTCCCA
>BMBL01041892.1_2
ACTCACTGCAAGAAAGAGAAAGCAAAATACAGTGTTCCCATTCAAGTTGGTTAACATAATCAAGTAGATTAACTTATG
AGGTTTGATAAAACACACTCCTCACACCATCCACAAGCTTTCTCAAACATTCATTCAGTGGGATAAGAATGGTGAAT
GTTGCCCTTTCAAGTACTTGAATCCGCCAGGTTTTCAGATGTTCTGTAACGGAGCTGTGTGTTAGCCTGAATGCAG
GTACTTTCCGCATTCCAGGAAGGACAAGGCAGATTGTTTCCAGTTCAGCTAAGGGACGATAAAAGCAGTGCAGGTT
TGCTGAGTAAACCTTCTGTTGTGAGGACTTGCTTGTTTCCAATACACTGAAAGAAAGAGAAATCAAAGATGGCTGC
TCTCAGTCTGGTGGATGAAGTCAGAAGGATTCACTTACTGTGTTTGAAAATCTCTCTCTGTTCCATGCCCAAGACTG
TTGCCTGTGCATTCTCCACCTGGGTAAAGAGCTCGACTTTTCTCCCCTTTGAGTACGTGAAGCCCTGCAGGGTTCA
AGGCTTTCTGGTGGGCACTCTATGCTGCACAGCCAAGAAACCAGCCTTACACAAAATGTCTCTGAAGGGGAAAAGA
AAGCAAATTTCTCCCATTCTGGTATACCGCAGCAGTAGAGATGAAGTTTTGCTAAATAAACACTATGTTAGCAGGAC
GTGCATGTTCTCA
>BMBL01041892.1_3
ATTTCCTGCAGGAAAGAGAAAGCAAAATACAATGTTCTCATTCAAGTTGGCTAAGTTCAGTTCATCAAGTTATCAGTT
TTGATAAAACACAGTCCTCACACCATCGAGAAGATTTCTCAAACATTGTTTCAGTGGGATAATTCTGGTGAGTCTCTT
GCTTTCAAGTGATTGAATCCCCTGGTTTTCAAGGGGGACTTAACAGAGAATGAAAGTACCCTTTGCATTTAGGAAAA
ACAAGTTAGATTCTTTTGCCTTCAGCTAAGGCATGGCAGAAGCAGTTCAGTTTGCCTGACTAAACCTAGTGTTGTAA
GCATGTGTAAGTTTCCTATTAACTGCAAGAAAGACAAAGCCAAAGACAGTTGCTCTCAGTCTGCTAGATGAAGTGG
GGATTCGCTTACTTTGTGTTGGAAAGTCCCTCTGTGTTACACACCATGAATGTTGTCCAGGAATTCTCTACCTGGG
TAAAGACTTTGATTTTTCTCCCCTTCCAATACGTGAAGCACCGCAGAGTTTACAGCTTTCCGGAGGAACTCTATGCT
GCCCGGCTGAGAGAGTGGCCTTACACAGCGAGTCTCTGAAGGGGAAAAGAAAGCGCGTTTCTCCCATTCTGGTAT
ACCACAGCAGTGGAGATATAGTTTGGAGAAGTAAACACTACGTTAGGGGGAAGTGCATGTTCCCA
>BMBL01041892.1_4
ATTCACCCCAGGAAAGAGGAAGCAAAATACGGCGTTCTCCGTCAAGCTGCTTAACTGAATCTACGGAAGTTTTCAG
TTTTGTTAGAAGAAACAGTGTGCCGACCAGTAGCTGCTTTCGCAAAGGCTCGCTCAGTGGGAGAATACAAGAGAGT
GCCTGCCTTGCAAGTATTTGAATCCCCTGGTTATCAAAGCTATCTTTGCCTAACAGGGAACGAAGGAACTTTTGGGA
TTCAAAGTCAAAGAGGTAGATTGTTTGCTTTTAGCGAGAGGATGTTAGAAGAAGTTCAGTTTGGCTGAGCAAAGCTT
GAGTGCTTCACAGGAGCGTGTAGCCTCTTCCCTGACAGAAAGAGGAAGCAAAGCCAGTTGCTGTCAGTCTGAGGG
ATGAAGTGAGAAGGATTCACTACTGCGTTTGAAAATCCCTCTGTGTTCCATGCCTAAGGAAGGAAGTTGCCTGTGC
ATTGTGTACCTGGGTAATGGGATCCACTTTTCTCCCTTTCAAGCAGGTGAAGCCATTCTGGGTTTAAAGCATTCCTG
ATGACATTCTGTGCTGCTGGGCCGTGAATGCAGCTTTACAGACGCTATCCCTAAAGGAGAAGAGAAACCTTTCTCC
AGTTCTGGTATACCACCGTGGCAGAGCCGCCGTTTTGCTGTGTCAAAAGTATGTTAGCAGCACTTATAACTTCCCA

>BMBL01022471.1_1
ATTCAGTGCATGAAAAAGAAAGCAATATGCAGTGTTCTCTTTCAAGCTGCTTCACTTAAACAACTTAAGTTTTCAGTG
TTGGTAGAACAAAGCGTCCTCATACCACCGGCTGCTTTCTCCACCTTTCTTGCAATGGAAGAATACTGGCGAATGTC
TGCCTTTCAAGAGATTGCATCCCCTGGTTTTCAAAGGGAACTTTGCAGTACATGGAATGAAGGTAAGATTTGCATTC
TGTGAAAAACATGCAGATTTTTTGCCTTCAGCTAGAGAAGGGCGGAAGTGGATCAGTTTTGTTGACTATATCCTGTG
TGGTTAACACGAGCAGGTATCTTATTCCTTCCAAGACAGAGAAACCAAAAGACAGTTGCTCTCATTCTGGGGGATG
AAGTGAGATGGTTTCACTTACTGTGCTTGAATATCCCTCTATGTTACATGCCCAGGAAATTTTCCTGTCGATTGTCTC
CCTGAGTAAAGTGTTTGAATTTTCTCCCCTTCGAGTACAAGAGGCCCTGGGGGTTCACAGCTTTCCCGAACCACTA
GTTGCTGCGCGGCCTAGAACACCGCCTTATAGACTGTATGCCTGAAGGAGAAGAGAAAGCGTGTTTTTCCTGTTCT
GGTAAGACAGTGGCAATGCTGTAGTTTTGCTAACTAAAGACTATATTAGCAGCACTTGCAAGTTCCTA
>BMBL01022471.1_2
ATTGAAAGCATGAAAGATAAAGCAAAACACAGTGTTTTCATTCAAGCTGCTTCACTTAAGCAACTTAAGTTTTCCGTT
TTGGTAGAAAAAACTGTCCTCACACCACCAGCTGCTTTCTCCAACATTCTTTCACTGGGATAAGACAGGCGAGTGTC
TGCCTTTCATGTAATTGAATCCCCTGGTTTTCAAAGGGAACTTTGCTGTACATGGAATGAAGGTAACATTTGCATTCT
GTGAAAAACACGCAGATGATTTGTCTTAAACTAGAGGAGGGAAGAAGCGGTTCAGTTTTGCTGACTAAACCTTGTGT
GGTTAACACCAGCAGGTAGGTATCTTATTCCTTGCAAGAAAGAGAAACCAAAAGACAGTTGCTCTCATTCTGGGGG
AATGAAGTGATCGGGATCCCACTTACTGTGCATGAAAGTCCCTCTGTGTTACATACTCAGCAATGTTGCATGTACATT
GTCTCCCTGAGTAAAGAGTTGTAATTTTCTCCCGTTCGATACGTGAAGCTCTGTGGGGTTTACAGCTTTCTGGGA
GGCATTCTGTGCTGTGTGGCCGAGAACGCCCCTCTTACAGATGGTATCCCTGAAGCAGAAGAGAAAGCGTGTTTCTC
CCTTTCTGGTATAAGACAGTGGCAATGCTCTAGTTTTGCTAACTAAACACTATGTTGGCAGCACATGCATGTTCCCA
>BMBL01022471.1_3
GTTGAAAACATGAAAGAAAGAGCAAAATACAGTGTTCTCATTCAAGCTGCTTCACTTTAGCAACTTAAGTTTTCAGTT
TTGGTGAACATACCGTCCTCACACCACCAGCTGCTTTCTCCACCTTTCTTTCAAAGGAATAAGACAGGCGAGTGTCT
GCCTTTCAAGTACTTGAATACACTGGTTTTCAAAGGGAGCACTGCAGTATGGGGAATGAAGGTAACGTTAGCATTCT
ATGAAAAAAAACGCAGATTATTTGCCTTCAGCTAGAGGAGGGCAGAAGTGGTTCAGTTTTGCTGACTAAAGCTTGTG
TGGTTAACGCGAGCAGGTATCCTATTCCTTGCAAGAAAGAGAAACCAAAAGACAGTTGCTGTCATTCTGGGGGATG
AATTGAAAAGGATTCACTTACTGTGTTTGAAAATCCCTCTATGCTACATGTCCAAGAAAGTTGCCTGTGCATTGTATC
CCTGGGTAAAGGCTTTGAATTATATCCAGTTCAAGTACGTGAAGTCCTGCGTGGTTTACAGCTTTCCGGGAGGCAC
TCTGTGCTGCGCGGGTGAGAAGGCCTCCTTACAGACCTTATGCCTGAAGAAGAAGAGAAAGCGTGTTCCTCCCATT
CTGGTATAGCCCAGTGGCAATGCGGTTTTGCCAAGTAACCACTATTTTAGCAGCACTTGCAAGTTCCCA
>BMBL01022471.1_4
ATTCAAAGCATGACAGATAAAGCAAAATAAGTGTTCTCATTCAAACTGCTTCACTTAAGCAACTTAAGTTTTCAGTTTT
GGTAGAAAAAACTATCCTCACACCACCAGCTGCTTTCTCCCATGTTCTTTCAATGGGATAAGACTGGCGAGTGTCTG
CCTTTCTAGTACTTGAATCCCTTGGTTTTCAAAAGGAACTTTGCAGTACAGGGAAGGAAGGTAACATTTGCATTCTAT
GAAAAACATGCAGATTGTTTGCCTTCCGCTAGAGGAGGGCAGAAGTGGTTCAGTTTTGCTGACTAAAGCTTGTGTG
GTTAACACGAGCAGGTATCCTATTCCTTGCAAGAAAGAGAAACAAAAGACAGTTGCTCTAGTTCTGGGGGATGAA
GTGAGAAGGTTTCACCTACTGTGCTTTAATATCCCCTCTGTGTTACATGCCCAAGAAAGTTGCATGTGCATTGTCTT
CCTGGGTTAAGAGTTTGAACTTTCTCCTGTTCGATTACGTCAAGCCCTGTGTGGTTTACAGCTTTCCTGTAACCACT
GGTTGCTGCGCGGCCAAGAACGCCCACCTTATAGAAGGTATCCCTGAAGGAGAAGGGTAAGCGTGTTTTTCCCATTC
TGGTAAAGACAGCGGCAATGCTGTAGTTTTGCTAACTAAACAGTATTTTAGCAGCACGTGCAAGTTTCCA
>BMBL01022471.1_5
ATTGAAAGCATGAAAGATAAAGCAAAATACAGTGTTCTCATTCAAGCTGCTTCACTTAAGCAGTTTTGGTGGAATATA
CCGTCCTCACACCACCAGCTGCTTTCTCCACCTTTCTTTCAACGGAATAAGACTGGCGAGTGTCTGCCTTTCAAGTA
ATTGAATACCTTGGTTTTCAAAGGGAACTTTTCAGTACATGGAATGAAGGTAACATTTGCCTTCTATGAAAAAAAACGC
AGATTGTTTGCCTTCCGCTAGAGGAGGGCAGAAGTGGTTCAGTTTTGCTGACTAAAGCTTGTGTGGTTAACACGAG
CAGATATCCTATTCCTTGCAAGAAAGAGAAACCGAAAGACAGTTGCTCTCATTCTGGGGGATGAAGTGAGAAGGAT
GCACTTACTGTGCTTGAAAATCTCTCTGTGCTACATATCCAAGAAAGTTGCCTGTGCATTGTATCCCTGGGTAAAGG
CTTTGAATTATCTCCCGTTCAAATACTTGATGCCCTGCGGGGTTTACAGCTTTCAGGGAGGCACTCTGTGCTGTGC
GGCCGAGATCGCCGCCTTACAGACGGTATCCCTGAAGCAGAAGAGAAAGCGTGTTTCTCCCGTTCTGGTATACAA
CAGCAGCAACGCTGTAGTTGTGCTAACTAAACACTATGTTAGCAGCACGTGCAAGTTCCCA
>BMBL01022471.1_6
ATTCAATGAATGAAAAGAAATCAAAATACAGTGTTCACATTCAAGCTGGTTAACTTAATCTACTTAATTTATTGTTTT
GCATAAAACACACTGCTCACACCATTTATAAGCTTCGTCCAACATTCTCTCACTGGGATAAGACTGGGAGTCAATTC
CTATGAAGTATTTGAATCCCCAGCTTTAAAAGGGAACTTTGCGTTACAGGGAATGAAGGTAGCTTTCACATTCCAGG
AAGAGCAAGCAGATTCTTTCGCCTTCAGCTAATGGCTCTCAGAGGCAGTTCAGTTTTGATGACTAAAGCTTGTGTTG
TAACCGCGTGCATGTTTCTTATTCAGTGCAACAAAGAGAAAGCAAACGACAGTTGCTCTCAGTCTGGGTGATGCAG
TTAGAAGGATTCACTGAGTTTGAAAATCCCTCTGTGTGACATGCCCAACAATTCTGCCTGTGCGTTGTCTACCTGGG
TAAAGAGTTTGACTTTACTCCCCTTCCAGTAAGTGAAGTCCTTCGGGTTTACAGCTTACCAGGTGTCACTCAGTGCT
GTGCGGCCCAGAATGCGGCCTTACGGACAGTATGTCTGAAGGAGAAACAAAGCGCGTTTCTCCCGTTCTGGTCT
ATGACCGCTTTAGAGATGTAGTTTTGCTAAGTAAACGCTACTTTAGCAGGACCTGCATGTTCCCA

Supplementary Sequence File 3. TMAsat monomeric sequences (in fasta) manually isolated from the *Hydrodamalis gigas* assembled reference genome (accession GCA_013391785.1).

>JACANZ010423367.1
GTTCACTGTATATAAGAGAGCAAAATACAGTGTTTTTCATTCAAGCTGGTTAACTTAATCAACTGAAGTTTTCAGTTTT
GGTAGAACAAACAGTTCTGACACCACCAGCTACTTTCTGAAACATTTTCTCAATGGGATAAGTCTAGCGAGTGTCTG

CCTTTCAAGTACTTCAATCCCCTGGTTTTCAAAGGAAATTTTGCATTACAGGGAACGAAGGAACTTTCGCATTCTAA
GGGAAACAGGCAGATTGCCCTTTAGCTAGAGGACGGCAGAAACAGTTCAGTTTTGGATACGAAATCTTGTGTGGTTA
ACATGAGTAAGTATCCTATTTCCTGCAACAAAGAGAAACCAAAAGTTGCTCTCTGTTTCAGGAATGAAGTGAGAAG
GATTCATTTACTGTGTTTGAAAATCCCTTTATGTTACTTGCCCAAGAAAGTGGCTGGTGCATTGTCTACCTGGGTAAA
AAGTTTGACCTTTCTCCCATTCCGGTACTTCAAGCCCTGCGTGGTTTACAGCTTTCCGGGAGGCATTCTGTGCTGC
ACGGCTGAGAATGCGGCCTTACGGACGGTATCCCTGAAGGAGAAGACAAAGCGTGTTTCCCCCATTCTGGTATAT
GACAGTGGCAGGGCTGTAGTTTTGCTAAGTAAACACTGTTAGCAGCAAGTACATGTTCCCA
>JACANZ010389660.1
ATTCACTGCATGAAAAAGAAAGCAAAATGCAGTGTTCTCATTCCAGCCAGTTTACCTAATCAACTGAAGTTTTCAGTT
TTGGTACAACAAACAGTCAGGACAGCACCAGCTACTTTGTGAAACATTCTCTCAATGGGATTAACTGTCAAGGGTCT
GCCTTTCTAGTACTTGAATCACCTGGTGTTCAAATGGAAATCTGAATTACAGGGAATGAAGGAAGCTTTCACATTCA
AAGAAACAGGCAGATTGTTTGCCTTCAGCTAGAGGACAGCAGAAGCTGTTCAGTTTTGCTGACTAAAGCTTGTGTG
GTTAACATGAGCAGGTATCCTATTCCCTGCAAGAAAGAGAAACCAAAAGATAGTTGCTCTCACTCCGGGGAATGAA
GAGAGAAGGTTTCACTTATTGTGTTTGAAAATCCCTCCGTTTTATATGCCCAAGAAACTTGCCTGTGCATTGTCTACC
TGGGTCAAGAGTTTGACTTTTCTCCCCTTAGAATACGTGAAGCCCTGCAGAGTTTATAGCTTTCCGGAGGCACTCTG
TGCTGCGCGGCAGAAATGCTGCCTTACAGACGGTATCGCTGAAGGAGAATAGAAAGCTCGTTTCTCCGTTTCTGG
TATAGGACAGCGGCAGCATTGTATTTTTGCTAACTAAACATTTTGTTAGCAGCTCATGCATGTTCCAA
>JACANZ010311924.1
ATTCCCAGTATGAAAGGGAAGGCAAAATACAGTGTTCACATTCAAGCTGGTTAAGTTAAGCAACTTAAGTTTTCAGN
NTGGTAGAACAAAAGGTCCTCACAGTACCAGCTCATTTCTCAAACATTCTCTAAGTGGAATAAAACTGGCAAGTGTC
TGCCTTTCAAGTACTTGAATCCTCTGCTTTTCAAAGGAACTTTGCATTACAGGGACTAAAGGTAGCTTTTGCATTCTA
AGCAAAACAGGCAGACTGTTTGCCTTCAGCTGGAGGATGTCAGCAGGGGTTCAGTTCTGCTGGCTAATCCTTGTGT
GTTTAACACAAGCATTTCTTATTCCTTGCAAGAAAGAGAAAGCAGAAGACAGATGCTCTCAATCTGGGGGATAAAGT
GAGAAGAAGTTACTCACTGTGTTTGAAAATCCCTCTCTGTTACTGGCCCAGGACAGCTGCCTGTGCATTGTATACCT
GGATAAGAGTTTGAATTTTCTCCCATTGGAGTTCGTGAAGCCCTGCGGGGTTTACAGCTTTTCAGGAGGCACTCTG
TGCTGTGAGGCCAAGAATTCTGCCTTACAGAGGGTATCCCTAAAGGAGAAGAGAGCACATTTCTCCCATCCTGGTA
TACAACAGTGGTAGCGCTGTAGTTTTGCTGAGTAAAGATTATGTTAGCAGCAAATGCAAGTTCCCAC
>JACANZ010229584.1
ATTCACTGTATGAAAAAGAAAGCAAAATGCAGCGTTCTCATTCAAGCCTGTTTACTTAATGTAGTGAAGTTTTCCGTT
TTGGTAGAACAAGCAGTCCTGCACCACCAGCTACTCTCTGAAACATTGTCTGAATGGGATATAACTGGTCTTTCAAG
TACTTTGATCCTCTGGTTTTCAAAGGGAAGTTTGTATTACAGGGAATGAAGGTAACTTTCGCATTCTAAAGAAAAGA
GGCAGATTGTTTGCCTTTAGCTAGAGGACAGCAGCAGCAGTTCAGTTTTGCTGACTAAAGCTTGTGTGGTTAATATG
AGCATGTTTCTTATTCCCTGCAAGAAAGAGAAACCAAAAGACTGTTGCTCTCAGTCTGAGAAGGATTCACTTACCGT
GTTTGATAATATCTCGGTGTTACATGCCCAAGAAAGTTTCCTGTGCATTGTCTACCTGGGTAAAGAGCTTGACTTTT
CTCTCCTTCTAGTACGTGAAGCCCTGCGGGGTTTACAGCTTTCTAGGAGGCACTCTGAGCTGCACGGCGGAGAAT
GCCTCCTTACAGACCTTATCGCTGAAGGGAAGAAAAAGCACATTTCTCCCCTTCTCGTATACGACAGCATCAGCGC
TTTAGTTTTTCTAACTAAACACTATGTTAGCAGCACGTGCATGTTCCCA
>JACANZ010159291.1_1
AGCCAACTACAGTCTTCTCATTCAAGCTGGTTAACTTAATCAACTTAAGTTTTCAAATGTGGTAGAACAAAGAGTCCT
CACACCACCAGCTGCTTTCTCAAACATTCTCTCAATGGGATAAGATTGGCGAGTGTCTGCCTTTCAAGTACTTGAAT
CCCCTCGGTTTCAAAGGGAACTTTGCATTACATGGAATGAAGGTACATTTTGCATTCTAAGAGAAACAGGCAGATTG
TTTGCCTCCAGCTAGATGATGGCAGAAGCGGTTCAGTTTCCCTGAGTAAATCTTGTCTTGTCAGCACATGAATGTTT
CCTATTCACTGCAGAAATGAGAAACCAAAAGATAGTTGCTCTCAGTCCGGTGCATGAACTGAGAAGGATTCACTTAC
TGTCTTTGGAAATCCCTCTGTGTTACTTGCCCAGGACTGTTCCGTGTGCATTCTTTACCTGGGTAAACACTTTCACTT
TTCTCCATTTCGACTACATGAGCTCTGCGGGTTTCCAACTCTCTGGGAGGCACTCTGTGCTGTGCAGCGAAGAATG
CAGCCTTACAGCCGATATCGCTGAGGGAGAAAACAAAGCGCATTTCTCCCATTCTGGCATACGACAGCAGTAGAGA
TGTAGTTTTGTAAAGTAAACACTACGTTAACAGGATGTGCATGTTCCCA
>JACANZ010159291.1_2
ATTTACCGCATGGAAGAGACAGCAAAATAGAGTGTTCTCATTTAATCTGATTAACTTAATCAAGCTAAGCTATCAGTT
TTGATAAAACACAGTCGTCACACTATCTACAAGCTTTCTCAAACATTCTGTCAGTGGAATAGTACTGGTGAGTCTCT
GCCTTTCAAGTACCTGAATCCTGAGGTTTTCAAAGGGAAATTTTCATTACAGTTAATGAAGTAACTTTTGCATTCTAG
GAAGAACAAGCAGATTATTTTCCCTTCAGCTATAAGATGGCAGAAGTGGTTCAGTTTTGCTGACTAAACTTTGTGTT
GCAAGCACGTGCATGTTTCCTATTCACTGCAAGAAAGGGAAACCAAAATACAGTTGCTCTCATTCTAGAGTACGAAT
TGAGAAGGATTCACTTACTGTGTTTGAAAATCCCTTTGTGTTACATGCCCAAGAATGTTGCTTATGCAGTCTCCACCT
GGGTAAAAGTTTGACTTTTCTCCCCTTCGAGCACATGAATGCCTGCCAAATTTACAGCTTTCCAGGAGGCTATCTGT
TCTGAACGGCTGAGAATGCAGCCTTACAGATGGTATCTGTGAAGCAGAAACAAAACGTGTTTCTGCCTTTCAAGTA
TGCTGCAGCAACAGAAATGTAGTTTTGCAAAGAAACACTATGTTACCACAACATACATGTTCCC
>JACANZ010122094.1_1
ATTCACTACACGAAACATAAACCAAAATACAGTATTCTCATTCAAGTTGGTTAACTTAATCTAGTTGATTAACTTATCT
GTTTTGATAAAACACAGTTCCCACATAATTGAGAACCCTTCTTATACATTCTCTGAATGGCAATAGAATATCAAGTTTT
TCCCTTTGTAAACTGTGAAGTCCACGAGTTTTCAGATATTTTCTATGCGAATTTTGTGTTACCAGAAATGAATGAGAC
TTTCACAATCCAGGCATAACAAGACAGACTCTTTCCCATTCAACTACAGGACCACAAAAACAATGTACTTTTGCTGA
CTAAACCTTGTGTTGTAAGCACGTGCGTGTTTCCAATCCACTGCACAAAAGAGAAACCAAAATACAGTTTCTCATATT
CTGGTTGATTTAAGTCAGAAGGATTAACTTATTGTGTTTAAAAATCCCTCTGTGTTACATGCCCGAGAATGTTGCCTA
TACATTCTCCACCTGGATAAAGAATTTAACTTTTCTCCCATTAGACTACATGAAGCCATGCGGAGTTTAGATGTTTTC
AGTATGCATTCTGTGTTACCTAGCCAAAAATGTGCCTTACAGACCGAATCTTTGGAGGAGAAACAATGTGAGCTTC
TTCCTTTCTAGTACTTGAATCCTCCAGGTTTTCAAATGTTTTCTATGGGAACTTTGCAATACCGGGAATGAATTGCCT
T
>JACANZ010122094.1_2

ATTCACTGCATGAAAAATAAAGCAAAACACAGTATTCTCATACAATCCTGTTTACTTAATCATCTGAAGTTTTCAGTAT
TGGTTAGAACAAGCAGTCTTAGGACCACCAGCTATTTTCTGAAACATTCTCTCAATAGGATAAAGTACTTGAATCTG
CTGGTTTTCAAATGGAAATTTGAATTATGGGGAATGATGGAATGTTTCATATTCAAAAAAAACAAAAAACAAAACAG
GCAGATTGTCTTCTGCTAGAGGACTGAAGAAGTAGTTCAGTTTTGCTGACTAAAGCTTGCATGGTTAACATGAGCAT
GTATGCTATTCCCTGCAAGGAAGAGAAACCAAAAGACTGTTGCCCTCAGTCTGGGGAACGAGGTGAGAAGGATTCA
CTTATTGTTTTGAAAATCCCTCCCAAGAAAGTTGCCTGTGCATTGTCTACCTGGGTAAAGGGTTTGCTTTTTCTCCTG
TTTTTGAAAGTGAAGCCTTGTGGGGTTTACAGCTTTCTGTGAGGCATTCTGTGCTGTGCGGCGGAGAATGCTGCAT
GACAGACTGTATCCCTAAAGGAGAAGACAAACCGCATTTCTCCCATTCTGGTATATGATAGCGGCAGCGTTGTAGT
TTTCCTGAGTAAAGACAATCTTGGCAGCAAGTGTAAGTACCCAT
>JACANZ010120854.1_1
ATTCACCACATGAAAGAGAAAACACATTACAACGTTCGCCTTCCAGCTGCTTAACTGAATCTACTGAAGTTTCCAGT
TTTGTTAGAAGAAAGTGTCCGCCCAGCACAAGCTACTTTCTCCAATGTTATCAGAGTGGGAGAAGGCTGGAGAGTG
CGGGCCAAGCACGTATTTGAATCCCCTGATTTTCAAAGGGATCTTTGCGTACCAGGGAATGAAGGGATTACTGGCT
TTCAAAGCCACACCTGCCCATTGTCTGCCTTCAGCTAGAGGATGGTAGAAGAGGGTCAGGTTTCCTAAGTAAACCT
TGTGTGCTTAACACGATCCTGTAGCTTCTTCCCTGCCTGAAAGACAAAGCAAGAGCCTGTTGCTCTCAGTCCAGGG
TTGAAGGGTGAACGACTCACTTTCTGTGTTTCAAAATCCCTCTGTGTTACATGCCTAAGGAAGTTGCCTGTGCATTG
TCTACCCGGATACAGTGTATGACGTTTCTCCCTTTCCAGCAGTTGAAGCCATTCAGAGTTTAGAGCATTCTGGAAGG
GATTCTGTGCTGCGTGGCTGGCTGAGAATACCGCTTTACAGACTATTCTGAAAGGAGAAGAGAAACCGCGTACCTCCCGT
TCTGGAACACGAACGAGGCAAAACTGTAGTGGTGGAGAGTAAACACTTTGTTAGCAGCAAGTGTAAGTTCCCA
>JACANZ010120854.1_2
ATTCCCCACGTGAAAGAGAAAGCAAAATGCAGCGTTCTCCTTCAAGTTGCTCAACTGAAGCTACTGAAATTTCCCAT
TTGGTTAGAAGAAAATGTCCGCCCACCAGAAGCTGGTTTCTCAAACGTTCTCTCAGTGGGCAAAGGCTGGAGAGTG
CCTGCCTTGGAAGTATTTGAATCCTCTGGTTTTCAACGGAAATTTGGAATAAGAGGGAATAGAGGTACTACTGGCTT
TCAAAGCAAACGCAGAGATTGTTTGACTTCACCTAGAAGATGAAAGAAGAGGTTCAGTTTTACTGACTAAAACTTG
AGTGCTTACGCGAGTGTGTAGACTCTTCCCTGCCTGAAAGAGAAAGCAAGAGCTTGTTGCTGTCAGTCTGGGGATG
AAGGGAGAACGATTCACTTTCTGTGTTTGAAAATCCCTCTGTGTTCCATGCCTAAGAAAGTTGCCTGTGCATTGTCT
ACCTGGAAAAAGTGTATGACTTTTCTCCCTTTCAAGTAGTTGAAGCCGCTCAGGGTTTCCAACATTTGGTAGGCATT
GTTTGCTGCACGGCTGAGAATACAGCTTTACAGACGGTAGCGCTAAAGGAGAAGAGAAACTGCCTTTCTCCCATTC
TGGAACACGAATGATACAAAACCGTAGTGGTGCTGATTAAAAACTTTGTTAGCAGCAATTCTTAGTTCCCA
>JACANZ010078570.1_1
ATTCACCACATGAAAGAGAAAGCAAAATACACCGTTCTCCTTCAAGCTGCTTAACAGAATCTACGGAAGTTTTCAGT
TTTGTTAGAGGAAAGAGTCCGCCCAGGACAAGCTGCTTTCTCAAACGTTCCTTGAGCGGGAGAAGGCTGGAGTGT
GGAGGCCTAGCAAAGATTTGAATGCCCTGGTTTTCAAAGGGAACTTTGCATTCCAGGGAATGAAGGTACTACTGGA
TTTCAAAGCAAAACAGTCAGATTGCGTTCAGCAAGAGGACGGCAGACGAAGTTCAGTTTTGCTGACTAAACCTTGT
GTGCTTAACACACACATGTAGCTTTTTCCCTGCCTGAAAGAGAAAGCAAGAGTTTGTTGCTCTCAGATTGGGGATGA
AGGGAGAACGATTCCCGTTCTGTGTTTGAAAATCCCTCTGTGTTACATGCCTAGGAAAGTTGCCTGTGCATTGTCTA
CCTGGATAAAGTGTATGACGTTTCTCCCTTTCAAGTAATTGAAGCCTTTTAGGGTTTCCAGCATTCCGGGAGGCATT
CTGTGCTGCTCGGCTGAGCATACCGCTTTACACACGGTATCGCTAAAGGAGAAAAGAAACCGCGTTTCTCCCATTC
TGGAACGTGAAGTAGGCAAAACTGTTGTTCTGCTTAGTAAACGTTTGTTAGCAGCAAGGGCAAGATCCCC
>JACANZ010078570.1_2
ATTCACCTCATGAAAGAGAAAGCAAAATACAGTGTACTCTTTCTTGCTGCTTTTCTCAATCTACTGAAGTTTCATTTT
TGTTAGAAGAAAGAGTCCGCCCAGCACAAGCTTCTTTCTCAAATGTTCTCTGAGTGGGAGAAGGCTGGAGAGAGCA
GGCCTAGCAAGGATTTGAATCCCGTGGTTTTCAAAGGGAACTTTGCATACTAGGGAATGAATATACTACCGGCCATT
AAACCCAAGCAGGCAGATTGTCTACCTTCAGCTAGAGGATGGCCGAAGAGGTTCAGTTTGGCTGACTAAACCTTGG
GTGCTTAACAGGAGCATGTAGCTTCTTCCCTGCCTGAAAGAGAAAGCAAGAGTTTGTTGCTCTCAGTCTGGGGATG
AACGGAGCACGATTCACTTTCTGTGTTTGAAAATCCCTCTGTGTTACATGCCTAAGAAAGTTGCCTGTGCATTGTCT
ACCTGGATAAAGTGTATGATGTTTCTCCCTTTGAAGTAGTTGAAGCCTTTCGGGGTTTCCAGCATTCCGCGAGGCAT
TCTGTGCTGCGCGGCTGAGAATCTCGCTTTACAGATGGTATCACTAAAGGAGAAGAGGAACCGTGTTTCTCCCATT
CTGGAATTCGAAGGAGGAAAAACCGTAGTTGTGCTGAGTAAAAAGTTTGTTAGCAGCAAGGGCAAGTTCCCC
>JACANZ010060488.1
AGTTCACTGTCTGAAAGCAAAAGCAAAATACAGCATTCTCCTTCAAGCTGCTTTACTTAATACACTGAAGTTTTCAGT
ATTGGTAGAAGAAACAGTCCGCACACCACCAGCTGCTTTGGGAAATGTTCTCCCAGTGGGATAAGACTGGAGAGTG
CCTGCCTTACAAGTACTTGAATCCCTTGGTTTTCAAAGGGAACTTTGCATAACAGGGAATGAATGTACCTTTGGCAT
TCAAAGCAAAACAGGCAGATTGTTTGCCTTCAGCTAGAGGATGGCAGAAGAGGTTCAGTTTTGCTGACTAAACTGT
GTGCTTAACACGAGCATGTAGTTTCTTCTGTGCCAGAAGAGGAAGCAAAAGCCAGTTGCTCTCAGTGTGGGGGAT
GGAGTGAGAAGGATTCACTTACTGTGTTTGAAATCTCTCCGTGTTACATACCTAATAAAGTTGCCGGGGCCTTGTCT
ACCTGGGTAGAGGGTTTGACTTTTCTCTTGTTTGAGTAGGTGAGCCCTTGCAGGGTTTACAGCTTTCTGGGAACCA
TTCTGTGCTGGGCAGCTGAGAATGCCGCCTTACAGACGGTATCCCTGAAGGAGGAGATAAAGAGGGTTTCCCCAT
TCTGGTATACAACAGAGGCAGCGCTGTACTTTGGCTAACTTCTATTTTACCAGCATGTTCATGTTCCCA
>JACANZ010031279.1
ATTCACCACCTGGAAGAAAAAGCAAAACACACTGTTTTCCTTCAAGCTGCTTAACTGAATCTACTGAAGGTTTCAGTT
TTCTTAGAAGAAACAGTCCGCCCACCAGAAGCTGCATTTACAAATGTTCACTCTGTGGGAGAAGAGTGGAGAGTGT
CTCCCTTGGGAGAATTTGAATCCCCTGGTTTTCAAAGGGAAATTTGCATCACAGAGTAGGAAGGTACATTTGGCTTT
CGATGCAAAACAGGCAGATTGTTTGCCTTCAGCTAGAGGATGGCAGAAGAGGTTCAGTTTTGCTGACTAAACCTTG
TGTGCTTAAGGCGAGCGTTAACTCCTTCCCTGCCAGCTAGAGAAAGCAAGAGACTGTTGCTCTCAGTCTGGGGATC
GAGAACGATTCACTTACTGTGTTTGAAAATCCCTCTATGTTCCATGCCTAAGAAAGTTGCTTGTGTATTGTCTACCTG
GAAAAAGTGTATGACTATTATCCGTTTTAAGTAGTTGAAGCCATTCTGGGTTTACAGCATTCCGGGAGGGATTCTGT
GCTGTGCGGTTGAGAATATCTCTCTACAGATGGTATCCCTAAAGGAGAAGAGAAACGGCATTTCTCCCCTTCTGAA
ACACGACTGAGGCAGAGCCGCTGTTTTGCCGATTAAAAACTACGTTAGCAACAAGGAGACGTTCCCT
>JACANZ010507883.1

ATTCACCGTATGAAATAGAAAGCAAAATACAGCGTTCCCTTCAAGCTTCTTTACTGAATCTACTGAAGAAAGAGTTC
GCCCACCACCAGCTCTTTGCTCAAATGTTCTCTCTCAGAGGGAGAAGACTGGAGAGTGTCTGCCTTGGAAGCATTT
GAATCTCCTGATTTTCAAAGGGAACTTTGCATAACAGGGAATGAGTTACTTCTGGCATTCAAAGCAAAACAGGCAGA
TTGTTGGCCTTCAGCTAGAGGAAATGCAGAAGACGTTCAGTGTTGCTGACTAAACCTTATTTGCTTAATGTGAGCAT
GTAGCTTATACCTGCCAGAAAGAGAAAGCAAAGCAAGTTGCTGTGAGTCTGGAGGATGAAGTGAGAAGGAGTCAC
CTACTGTGTTTGAAAATTCCTCTGCGTTACATGTCTAAGAAAGTTGCCTCTACATTGTCTCCCTCGGTAAAGGGATC
CACGTTTCTCCCTTTCAAGTATGTGAAGCCATTCAGAGTTTACAGCTTTCGAGGAGGCATTCTGTCCTGCGCGGCC
AAGAATGCCGCCTTTCAGACGGTATCACAGAAGGGAGTGAAACCGTGTTTCTCCCGTTATGGTATAGGACAGCGGT
GGAGCCACAGTTTTGCTGAGTAAAGACTATGGTACCAGCAAGTGTAAGTTCCCA
>JACANZ010255388.1
ATTCAGTGCATGAAAAAGAAAGCAATATGCAGTGTTCTCTTTCAAGCTGCTTCATTTAAACAACTTAAGTTTTCAGTG
TTGGTAGAACAAAGCGTCCTCATACCACCGGCTGCTTTCTCCACCTTTCTTGCAATGGTGGGAGAAGGCTGGAGAG
TGCAGGCCTAGCAAGTATTTGCATCCCCTGGTTTTCAAAGGGAACTTTGCAGTACATGGAATGAAGGTAAGATTTGC
ATTCTGTGAAAAACATGCAGGTTTTTTGCCTTCAGCTAGAGAAGGGTGGAAGCGGATCAGTTTTGCTGACTATATCC
TGTGTGGTTAACACGAGCAGGTATCTTATTCCTTCTAAGACAGAGAAACCAAAAGACAGTTGCTCTCATTCTGGGGG
ATGAAGTGAGATGGTTTCACTTACTGTGCTTGAATATCCCTCTATGTTACATGCCCAGGAAATTTTTCTGTCGATTTT
CTCCCTGAGTAAAGTGTTTGAATTTTCTCCCCTTCGAGTACAAGAAGCCCTGGGGGTTCACAGCTTTCCCGAGCCA
CTAGTTGCTGCGTGGCCTAGAACACCGCCTTATAGACCGTATGCCTGAAGGAGAAGAGAAAGCGTGTTTTTCCTGT
TCTGGTATAAGACAGCGGCAATGTTGTAGTTTTGCTAACTAAAGACTATATTAGCAGCACTTGCAAGTTCCCAN
>JACANZ010245123.1
ATTCACTGCAAGAAAGAGAAAGCAAAATACAGTGTTCTCTTTCAAGCTGGTTAACTTAATCAGTTAAGTTTTCAGTTT
TGATAGAACAAACAGTCCTCACACCACGTACTAGATTTCTCAAACTTTCTCTCAATGGGATAAGACTGGCGAGTTTC
TGTTTTTTATCTACTTGAATACCCTGGTTTTCAAAGGGAACTTTACATTACAGGGAATGAAGTTACCTTTTGCATTCTA
GGAAAAAGAGGGAGATTATTTCGCCTTCAGCTAGAGGACCGCAGAAGTGGTTCAGTTTTGCTGATTAAACCTTATGT
GGTTAGCACGTGTATGTTTCCTACTCCCTACAATAATGAGAAAGCAAAGACAGTTGCTCTTAGTCTGGGGGATGAA
CTAAGAAGGATTCCCTTAGTGTGCTTGAAGATCCCCCTGTGTTACATAACCAAGAAACTTGCCTGTGCATTGTCTAT
CTGGGTAAAGACTTTGACTTTTCTCCCGTTGGGTATGTGAAGCCCTGTGGGCTTACAGCTTTTGGGAGGCACTCT
GTGCTGTGCGGCCAAGAATGCTGCCTTACAGACCTATCTCCATAGTAGAGAAAGTGCATTTCTCCCCTTCTGCCAT
ATGATAGTGGCAGTGCTGTAGTTTTGCTAACTAAACACTTGTTAGCAGCACATACATTTTCCCAG
>JACANZ010129270.1_1
ATTCAATGCTTGAATGAGAGAGAAAAATACAGTGTTGTCCTTCAAGCTGCTTAACTTAAGCTCTCACTTTTGATCAAA
CACGGTCCTCACACCATCTACAAACTTTCTCAAGGGGTTCTCTCAATGGGAAAAGGCAGGCGAGCGTCTCCCTTTCA
AGTGCCCGAATCCCTTGGGTTGCAAATGGAACTTTGCTTTACAGGGAATGAAGGTACCTTTCACACTCTAAGAAGAA
CAAGCAGACTGTTTCGGCTTCAGCTAGAGGAGAGCAGAAGCGGTTCATTTTTGCTGGATAAACCTTGTGTGGTAAA
CGCGTTCATGTTTCCTATTCACTGCAAGAAAGGGAAACCAAAAGACCCTTGCTCTCCGTCTGGTGGATGAAGTGAG
AAGGATTCACTTACTGTGGTTGAAAACCCCTCTGGGTTATATGCACAAGAATGTTGCCTGTGCATTATCCATCTTTG
TAAAGAGTTCCCCTTGTCTCTGCTGCGAGTACGTGAAACCCTGCGGGGTTTAACGGCTTTCCAGGAGGCACTCTGTG
CTGCGCAGCTGAGAATGCGGCCTTACAGATGGTATGTCTGAAGTCAGAAAGAAAGCGCTTTCCTCCCGTTCTAAGA
CAGCAGTAGGGATGTAGTTTTGCTAAGCAAACACTATGTGCACCTTCCCA
>JACANZ010129270.1_2
ATTCACTGCATGAAAGAGAGAGCAAACTGCACTGTTCTCATTCATGCTAGTTAACTTTCATCAGTTTTGATAAAATAC
AGTTCTCGCTCCATCTACAAGCTTTCTCAAGCATTCTCTCAGTTGGATTAGAACAGCGAGTGTCTCCTTTTGAAGTA
CGTGAATCCCCTGGCTTTCCAAGTGAACTTTACCTTCTAGAGAATGAAAGTACCTTTCGCACTCTAGGAAGAACAAG
CAGATTGTTTGCCTTCAGCTAGAAGACAGCAGAAGCGGTTCAGTTTTACTGAGTAAACTTGTGTGGTAACCGCGCA
CGTATTTCCTATTCACTGCAAGACAGACAAACCAAAAGAAACTTGTTCTCCGTCTAGTGGATGAAGTAGGAAGGATT
CACTTACTGTGATTGAAAATCCCTCTGTGTTACATGCCCAAGAATGTTGCCTGTGCATTGTCTACCTGAGTAAAGGG
TTTGCCTTTTCTCCCTTCAGAGTACGTGAAGCCTGTGGGGTTTACAGCTTTCAGGGAGACAGTTTGTGCTGCGTAG
TCAAGAGGGCGGCCTGAGAGCTGGCATGTCTGAAGGAGGAAAGAAAGCACGTTTCTCCCGTTCTGGTATAGGTGG
GTGGTAGAGATGAAGTTGTGCTAAGTAAACACTATGTTAGCAGGACTTGCAAGTTCCCA
>JACANZ010754534.1
ATTCACTGCATGAAATGAGAGCAAAATACAGTGTTCTTATTCAAGCTGGTTAAGTTCTCAGTTTTGATAAAACTTGGT
CCTCACAGCATCTATAAGCTTCCTCAAGCATTCTCTCAATGGGACAAGACTGGTGAGTGTTTCCCTTTCAAGTTCTT
GAATTCCCTGGTTTTCAAAGGGAACTTTACATTATAGGGAATGACAGTACCTTTCCCATTCTAGGAAGAACAAGCCA
ATTGTTTCGCCTTCAGCTAGAGGATTGCAGAAGTGGTTTAGTTTTGCTGAGTGAACCTTGTGTGTTAATGGCCTGCG
TGTTTCCTATTCTCAGCAAGAAGGAGAAACCAAAAGACAGTTGCTCTCCGTGCTGGATGAAGGAGAAGGATTC
ATTTACTGTGGTTGAAAATCCCTCTGTGTTACATGCCCAAGAATGTTTCCTGTGCATTGTCTACCTGGGTGAAGAGT
TTGACTTTTCTCCACTTCCAGTACATGAGGCCCTGCAGGCTTTCCAGCTTTCCAAGAGGCACTGTCTGCTGCACGG
TTGAAAATGCGGCCTTATAGACGGTGTGTCTGAAGGAAGAAAGAAAGCATGTTCCTCCAGTTCTGGTATACAATGG
CAGTAGAGATGTATTTTTGCTAAGGAAACACTATGTTAGCAGGA
>JACANZ010175637.1
ACTCACTGCACAGAAGAGAAGGCAAAGTAGAGTGTTCTCATTCAATCTGGTTATCTTAATCAACTTGATTAAGATATC
AATTTTGATAAAACAGAGTCCTCACACCAGCTACAAGCTTTCTCAAACATTCACTTGGTGGAATAAGACTGGAGACT
CCCTCCCTTTGAAGTACTAGAATCCTACGGTTTTCAAAGGGAACTTTGCATTACAGGTATGGAAGTTACCTCTTGCA
TTGTAGGAAGAACAAGGCACATTGTTTTGCCATCAGCTAAAGGACGGCAGAACCGGTTCAGTTTTGCTGATTAAAC
CTTGGGTTGTAACCACCTGCATGTTTGCTATTCACTGCAAGAAAGAGAACCCAAAAGACAGTTGCTCTAAGTCTGGT
GAATGAAGTGGGTAGGATTTACTTACTGTGTTTGAAAATTCCTCTGTGTTACATGCCCATATACATTCTACACCTGG
GTAAGGAGTTTGACTTTTCTCCCCTTCCAGTACGTGAAGCCCTGCGGGGTTTACAGCTTTGCGGGAGGCACTCTAG
GCTGTTCGGCTGAGAATGCCACCTTACAGACGGTATGTCTGAAGGAGAAAAAAGGCACATTTCTCCCATTCTGGTT
ATACGACAGCAGTAGAGACTTAGTTTTGCCAAGGAAACACTATGTTAGCAGGACCTGCATGTTCCCA
>JACANZ010271070.1

ATTCACTGCATGAAAGAGAAATCAGAATACAATCTTCTCATTCAAGTTGATTAACTTAATCATGTTGATTATGTTGTCA
ATCGTGACACAGCACAGTCCTCACACCATCCACAAGCTTTCTCAAATATTATCTCCATGGGATAAGAAAGGAGAGTC
TCTCCTTTTCAAGTACTTTTATCCCCTGAGTTTTAAGATGTTTTCTAAGGGATCTTTGTGTTCCTGGAATGAAGGTAT
CTTTCACAGTCTAGGAAGAACAAGGCAGATTGTCTCGCCTTCAGCTGAAAGACGGCAGAAGCTCTGCAGTTTTGCT
GACTGAAACTTGTGTTGTAATCACATGCATTTTTCTATCCCCTGCAAGAAAGAGAAACCAAAAGACAGTTGCTCTCA
GTCTGGGGAATGAAGTGAGAAGAATTCCCTTACTGTGTTTGAAATTCCCTCTGAGTGAAATGCCCTCGAACGTTGC
CTGGGCATTTTCGCCCTGGCTAAAGATTTTGAGTTTTCTCCCGTTTGAGTAGGTGAAGCCCTGCGGGGTTTACAGC
TTTCTGGGAGGCACTCTGTGCTTCGCAACGGACAAGGCCAGCTTACAGACGGTATCGCTGAAACTGAGGAGAAAG
CCCATTTCCCTCTTTATGGTATAGGACAGCAGCAGTGCTGTACTTTTGCTAAGGAAACACTATGTTAGCAGCACCTT
CGTGTTCCCA
>JACANZ010035624.1_1
ATTCAGTGCAGAAAGAGAAACAAAATACGGTGTTCCTTTGAAGAAGGTTAACTTAAGTTGATTAGGTCATTAGTAT
TGATAAAACACAGTAGTCACACCATCCACAAGCTTTGTCAAACATTCTCTCAGTGGGATAAGAATGGAAAGTCTCTC
CCTTTCAAGTACTTGAATTCCCCGGATGTTCAGAGGTTTTCTAGGAGAACTTTGCATTACCCGGTATGAAAGTACCT
TTTTCTCTCCAGGAAGAACAAGGCAGATTGCTTAGCATTCATGTGAAGGATGGAAGAAGTGGTGCAGTTTTGCTGA
CTAAGCCTTGTGTTGTAAGCACGTGCATGTTTCCAGTTCACTGCAAGAAACAGAAACCAAAAGGCAGTTGCTCTCA
GTCTGGTGGATGAAATCAGAAGGATTCACTAACTGTTTGGAAATCCCTGTGTTACATGCCCAAGAAAGTAGCCATG
CTTTCTCCATGTGGGTAAAGAGTTTGACTTTTCTCCCCTTTGAGCAAGTGAAGCCCTGCGAGGTTTAGAGGTTTCCG
AGAGCCACTGTGTGCTGCCTGGCTGAGAATGTGGCCTTCTGGAAGGTATCTCTGAAGGAGAAACAATGCACGTTTC
TCCCATTCAGGTATACGACAGCAGTAGAGATGTGCTTCTGCTAAGTAAACACTGTATTAGCAGAACGTGCATGTTCC
CAG
>JACANZ010035624.1_2
ATTCCCTGAATGAAAGAGAAAGCAAAACACAATGTTCTCATTCAAGTTGGTTAACTTAATCAAGTTGAATAAGTTATG
AGTTTTGATAAACCACAGTCCTCAAACCATCCAAAAGCTTTCTCAGACACTCTCTCAGAGGGATAAGAATGGCGAGT
CTCTCCCTTTCTAGTACTTGAATCCCCCAGGTTTTCTGTTGTTTTCTAAGAGAACTTTGCATTACCCAGAATGAGGGT
AACTTTTGCATTCTAGGACGAACAAGGCAGATTGCTTTGCAATCAAGAGACGTACAGCAGAAGCAGTACGGTTTTTT
TTTTTTTTAACTAACCCTTGTGTTGTAAACACTTGCATGTTTCCAGTTCACTGCAAGAAAGAGAAACCAAAACAGAGT
TGCTCTCAGTCTGATGGATGAAGTCAGAAGGATTCACTGACTGTGTTTGGAAATCCCTCTGTGTTACATGCCCAATA
ATGTTGCCCATACCTTCTCCACCTGAATAAAGACTTTGACTTTTTTCCCCTTCGAGTACGTGAAGCTCTGCAGGGTT
TAGAAGTTTCCAGGAGGCACTCTGTGCTGCCCGGCCGACAATGCGGCTTTACATATGGTATCTCTGAGGAAGAAAA
CAATGCACTTTTCTCCCGTTCAAGTATACAACAGTAGAGATGTGATTTTGCTAAGTAAAGACTATGTTAGCAAGATAT
GCGTATTCCCA
>JACANZ010035624.1_3
ATTCACTGTATTAAAGAGAAAGCAAAACACAGTGTTCTCATTCAGGACGGTTAACTTAATCAAATCGAATAAGTTATC
AGTTTTGATAAACCACAGTCCTCACACCATCCACAAGCTTTCTAAACATTTTCTCAGTGGGATAAGAATGGCGAGT
CTCTCCCTTTCAAGTCCTTGAATCCTGCGGGTTTTCTGTTGTTTTCTAAAAGAACTTTGCGTTACCCAGAATGGAGG
ACCTTTTGCATTCTAGGAAGAACAAGGCAGATTGTTTGCCACTCAGCTAAAGGACGGCGAAGCTGTGCAGTTTGG
TGACTAACCCTTGAGTTGTAAGCACATGCATGTTTCCAATTAACTGAAAGAAAGGGAAACCAAAAGGCAGTTGCTCT
CAGTCTGCTGGATGAAGTGAGGATGCACTTAACGTGTTTGAAAATCCCTCTGTGTTACATGCTCAAGAACGTTGCCT
GTGCATTTTTGCCCTGGGTAAAGTGTTTGACTTTTCTCCCATTCGAATACGTGTATCCCTGCGGAGTTTACCGCTTTT
CCGTGTGGCTCTCTTTGCTGCGCTGCAGATAATGCCGCCTTGCATACCTTATCGCTGAAGGGAAGAAAAAGCGCAT
TTCTCCCATTCTCATATACGACGGCAGCAGTGCTGTACTTTTGCTAACTAAACACTATGTTAGCAACATTTGCATGCT
CCCA
>JACANZ010035624.1_4
ATTCATTCACATGGGCAAGAAAAAGACCGTGTTCTCATTCAACCCTGTATACTTAATCAACTGAGGTTTTCCGTTTTG
GTAGAGCAAGCAGCCCTGACACCACCAGCTACTTTCTGAAACATTCTCTCAAGGGGTTAAGATTGACCAGTTTCTG
CCTTTGAAGTACTTGAGTCCCCTGGTTTTCAAATGGAACTTTGCCTTACAAGCAAAATGATAACTTTCACATTGTAA
GGAAAATAGACAGATTGTTTACCTTCAGCTAGAGGATAGGAGAAGCAGTTCAGGTTTGCTAACTAATGCTTGTGTGG
TTAACATGAGCGTGTTTCCTATTCCCTGCAAGTAACAGAAACCTAAAGGCTCTTGCTGTCAGTCTGGGTAATGAAGT
GAGATGGAGTCACTTATTGTGTTTGAAAATCCCTCTCTGTTACAGGCCCAAAAATGTAGCATGTACATTGTCTATCT
GTGTAAAGAGTTTGCCTTTTCTCCCCTTCGAGTACCTGAAGCCCTGCGGGGTTTACAGCTTTCCGGGAGGCCCTCC
GTTTCTCCAGTTCTGGTATACGAGAGCGGCAGAGCTGTAGTTTTGCTAACTAATCACTGTGGAAGAAGCACGTGCA
TGTTCCCAT
>JACANZ010107986.1_1
ACTCACTGCAAGAAAGAGAAAGCAAAATACAGTGTTCCCATTCAAGTTGGTTAACATAATCAAGTAGATTAACTTATG
AGGTTTGATAAAACACACTCCTCACACCATCCACAAGCTTTCTCAAACATTCATTCAGTGGGATAAGAATGGTGAAT
GTTGCCCTTTCAAGTACTTGAATCCGCCAGGTTTTCAGATGTTCTGTAACGGAGCTCTGTGTTAGCCTGAATGCAGG
TACCTTCCGCATTCCAGAAAGGACAAGGCAGATTGTTTCCAGTTCAGCTAAGGGACGATAAAAGCAGTGCAGGTTT
GCTGAGTAAACCTTCTGTTGTGAGGACTTGCATGTTTCCAATACACTGAAAGAAAAAGAAATCAAAGATGGCTGCT
CTCAGTCTGGTGGATGAAGTCAGAAGGATTCACTTACTGTGTTTGAAAATCTCTCTGTTCCATGCCCAAGAATGT
TGCCTGTGCATTCTCCACCTGGGTAAAGAGCTCGACTTTTCTCCCCTTTCAGTACGTGAAGCCCTGCAGGGTTTAA
GGCTTTCTGGTGGGCACTCTATGCTGCACAGCCAAGAAACCAGCCTTACACAAAGTGTCTCTGAAGGGGAAAATAA
AGCGCATTTCTCCCATTCTGGTATACCGCAGCAGTAGAGATGAAGTTTTGCTAAATAAACACTATGTTAGCGGGACG
TGCATGTTCTCA
>JACANZ010107986.1_2
ATCTCCTGCAGGAAAGAGAAAGCAAAATACAATGTCCTCATTCAAGTTGGCTAAGTTCAGTTCATCAAGTTATCAGT
TTTGATAAAACACAGTCCTCACACCATCTACAAGCTCTTTAAGACATTCTCTCAATGGGATAAGATGAGTGAGTGTTT
CCCTTTCAAGAACTTGAATCTCCTGGTTTTCAAAGGAAGTTTGCATTACAGGGAAGGAAGGGGCCTTTGGCATTGTA
GAAAATCAAGGCAGATTGTTTCAGCTTCATCTGAAGCATGGCAGAAGCAGTTCAGTTTTGCTGAATAAATCTCGTGT
TGTAAGCACGTGCATGTCCGTTATTCACTGAAAGAGAAACTAAAAGTCAATTTTGTCACTCTTCTGGATGAAATGAG

AATGAGTCACTTAACCGGTTTGAAAATCTGTCTGTGTTACATGCCCAAGAATGTTGCCTATGAATTGTCCATCTGGA
TAAAAGCTTGGATATTTCTCCCTTTTAAGTACTTGATTTAAGTACTGCGGGATTTACAGCTTTCCAGGAGGCATTCTA
TGCTGCCTGGCGAGAATGCAGCCTTACAGACAGTATATCTGAGGGGAAAAAGAAACCACGTTTTCTTCCATTCTGG
TATACCACAGCGGTACAGATGGAGTGTTGCTAAGTAAACCCTATGTTAGGCAGACGTGCATGTTCCCT
>JACANZ010061387.1_1
AATCACTGTAAGAATGAGAAAGCAATATAGAGTGTTCTCCTTCAAGCTGGTTAACATACGTTCATCAGATTCTCAGTT
TTGATCAAACACAGTCCTCACATCACCTACAAACTATCTCAAACATTGTCTCACTGAAATAAGATTGGTGTCTCCCTT
TTAAGTACTTCAATCTCCCAGTTCTGAAAAGGAACTTTGCAGTACAGTGAATGAAGGTACGTTTTGTATTCTAGGAA
GAACGCAAACTGTCTCGCCTTCCGTTTAAGCACGGCAGAAGCGGTTCAGTTCTGCTGACAAAACATTGTGTTGTTA
CCATGTGCATGTTTTCTCTTCACTGCAAAAAAGGGAAACCAAAAGACAGTTGCTCTCACTCTGGGGGATGAGGTCA
GAAGGACTCACTTAAGGAGTCTGAAAATCTCTCTGTGTAACATGCCAAAGAAGGTTGCCTATGCATTCTCCACCTGG
GTAAAGTGTTTCAGTTTTGTCCCTTTGGCGTATGTGAAGCCCTGCAGGGTTTACAGCTTTCTGCCAGGCAGTCTGTA
CTGCCTGTGGAGAACGCAGCTGTACAAACGGTATCTCTGAAGGAGAAAAGAAAGCGCCTTTCCCCTGTTCTGGTAC
AGCACACCAACAGAGTTGTAGTTTTGCTAAGTGAACACTATCCTAACATTACCTGCAAGTTCCCCT
>JACANZ010061387.1_2
ATTCACCGTATGAATGAGAAAGCAATATACAATGTTCTCCTTCAAGCTGGTTAAGTTCATCAAGTTCACCGTTTTGAA
CAAACACAGTCCTCACACCATCTACAAACTTTCTGAAACATTCTCTCACAGAGATAAGAGTGGCAAACGTCTGCCTT
TCAAGTACTTCAAGTTCCCAGTTCTGAAAAGGAACTTTTCAATACAGTGAATGAAGGTACGTTTTGTGTTCTAGGAA
GAACAATGCAAACTGTCTCGCCTTAGCTAAAGCACGCTAGAAGCAGTTCAGTTATGCTGACAGAACATTGTGTTGTT
AACATGTGCATGTTTTATCTTCACTGCAAGAAAGAGAAACCGAAAGACAGGGGCTGTCACTCGGGGGGATGAAGTA
AGAAGGACTCACTTAAGGTGTCTGAAAATCCCACTGTGTTACATGCCAAAGAAGGTTGCCTATGCATTCTCCACCTG
GGGAAAATGTTTCAGTGTTCTCCCTTGGCCTTACCCGAAGCCCTGTGGGGTTTACACCTCTCCGTGAGGCAGTCCG
TGCTGCCAGGGGAGAATGTGGCCTTACAGATGGTATCTCTGAAAGTGAGAAGAAAGCGCCTTTCCCCTGTTCTGGT
AAAGCACAGCAACAGAGTTGTTTTGCTAAGCCAACACTATGCTAACACTACCTGCAAGTTCCCCC
>JACANZ010515044.1
ATCCCCTGCAAGGAATAGAAAGCAAAACAATGATCTCATTCAAGCTGGTTAACGTAATCGACTTAAGATATGAGTTT
TGATAACACACAGTCCTCACACTGTCTAGAAGCTTCCTCCAACATTCTCTCAAAGGGAGAAGGCTGGGGAGTGTCT
CCCTTTCAAGGACTTGAATGCCCTTTCTTTCCAAGTGAACTTTGCATTACAGGGAAACAGGGTACCTTTTGCCTACT
TGGAAGAAAAAGTCAGATTTTTCTGCCTTCAGCTAAAGCACGGCAGAAGGTGTTCAGTTTTGATGACTAAACCTTGT
GTTGTAAGCACGTGTATGTTTCCTGTTAAGTGCAAGAAAGAGAAACTAAAAGACAGTTGCTCTCAGTCTGGTGGATA
AGGTCAGAAAGATTCACTTACTGTGTTTCAAAATCTGTGTTACATGCTCAGGAATGCTGCTTGTGCATGCTCTCCCT
GGAGAAAGATTTTGGGTTTTCTCCCCCTGGACTTCTTGAATCCCTGTGGCTTTTGCACCTTTCTGGGAGGCGCTCTT
CGCTGCATGGCCAAGAAAGCAGATTTTTCAGGCGGCATCTCTGAAGGAGAAAACAAAGTGCTTTTCTCCCGTTCTG
GTATACGACAGCAGTAGAGATATAGTATCGCTAAGCAAACTCTATGTTAGCAGGACATGTATGTTCCCAC
>JACANZ010190329.1_1
ACTTACTGTATGAAAGAGAAACCAGAATACAGGGGTTCTCATTCAAATTGGTTAAGCTAATCAAGTTTATTAAGTTATC
AGTCTTTATACAACACTGTACTCACACCATCCACAAGCTTTCCAAACATTATCTCAATGGGATAAGAAAGGAGAGTC
TTTCCATTTCAAATGCTTGAATCCCCCGGATTTTCAGACGTCTTCTAAGGGATCTTTGTGTTACCAAGAATGAAGGTT
ACCTTTCCCAGTCTAGGAAGAGCAAGGCAGATTGTCCCATCTCAAGCTAAAGGACTGCAGAAGCGGTGCAGTTTTG
AGGAGTGAACCTTCTGTTGTAAGCACAGGCATGTTTCCTATTCACTGTAAGAAAGAGATACCAAAAGACAGTTCATG
CCAATCTGGTGGATGTAGTCTGAAGGATTCAGTTACTGTGTTTCAAAATCCCTCTGTGTTACATGCCCCAGAATGTT
GCCTATGCATTGTCCACCTAAGTAAAGTGTTTGACTTTTCTCCTCTTCGAGTACATGAAGCCCTACAGCATTTACAG
CTTTCGCGAGCACTCTGTGCTGCACTGCTGAGAATGCGGTCTTACAGGCAGTATCTCTGAAGGAGAAAGAAAGTG
CCTTTCTCCCGTTCTGCTGTAGGACAGCAGTAGAGACGTAGTTTTGCTAAGTAAACACTATGTTAGCATTCCATGCA
TGTTCCCA
>JACANZ010190329.1_2
ATTCACTGTATCAATGAGAAAGCAATATACAGTGTTCTCATTCAAGCTGGTTAACTTAATCAAGTACATTAAGTTCTC
AGTTTTGATAAAACACACTCCTAACACCATCTATAGGCTTTCTCAAACATTCTCTCAATGGGATAAGAGTGGCAAGT
GTCTGCCTTTCAATTACTTTAATTCCCTGGTTTTCAAAGGGAGCTTTGCATTACAATGAATGAAGGTAACTTTCGCAT
TCTAGGAGGAACTGATTGTTTAACCTTCAGATATTCTGTTTTTCTGACTAAACCTTCTGTTGTAACCACGTGCATGTT
TCCTATTCACTGCAACAGAGAGAAACTAAACAACAGTTGCTGTCAGTCTGGTTGATGAACTGAGAAGGATTCACTTA
ATATGTTCGATTGTCCCTCTGTCTTACATGTCCAAGAATGTTGCCTGTGAATTCTCTTCCAGGGTAAAGCATTTCACT
TTTCTCTCCTTAGAGTACGTGAAGCCTTACGGAGTTTACAGCTTTCCCTGGGGAACTCTGTGCTGGGCGCTCAAGA
ATGTGGCCTTAGAGATCTTATTCCTTAAGGAGAAAACAAAGCCCATTTCTCCCGTTCTGATACATGATGGCAATAGA
GACATAGTTTTTGCTAAGGAAACACTACGGTAGCAGAGCCTGCATGTTTCCTA
>JACANZ010402190.1
ATTCACTGCATGAAAGAGAAATCAAACTACAGTTTTCTCCTTCAAGCTGGTTAACTTAATACAGTTGATTAAGTTATA
AGTTTTCATCAAACACAGTTCTCACACGATCCACAACACTTCTCATACATTCTCTGAATGGGAGAAGAATGTCAAGC
TTCTCCCTTTCAAATATTGAATCCCCGTGGTGTTCAGATGTTTTCTGCGGGAGCTTTGCGTTACCAGGAATGAATGT
CCTTTCACATTCTAGGAAGAACAAGGCAGACTGTTTCCTGTTCACGTAAAGGACAGCAATAGTGATGTAGTTTAGCT
GATTAAACCCTGTGTTGCAAGCACGTGTATGTTTCCAATTCACTGCAAGAAAGAGAAACAAAGTACTGTTTCTCTC
ATTAGGGTGGATGAAGTCAGAAGTATTAACTTACTGTGTTTGAAAATCTCTCTGTGTTACATGCACAAAATGTTGCC
TATACATGCTCCACCTTGGTAAAGAGTTGAACTTTTCTCCCATTCGAGTACATGAAGCCAAGTAGGGTTTAGATGTT
TTCACTAGGCACTCTGTGCTGCCCGGCTGAGAATGTGCCTTACAGACAGTATCTCTAAAGCAGAAACAATGTGAG
TTTCTCCCATTCAGGTGTAGGACGGCGATAGTGATGTCGTTTTGCTAAGTTGAAACTGTGTTACCAGAAAGTGCATG
TTCCCA
>JACANZ010202422.1
AATTAGTGCCTTAAAGAGAAACCAGAATACAGTTNNNNNNNNNNTTGGGTAACTTAATCTACTTGAGTAAGTTATCA
GTTTGCATAAAACACACTGCCCACAGGATCCACAACGTTCCTCCTACATTCTCTGTAAGGAAGAAGAATGTCAAGTG
TCTGCCTTTAAAGTACTTGAATTCCCCGGGCTTTCAGAGGTGTTCTGGGGAAAATGTGTGCTACCAGCAGTGAATG

TGCATTTCGCATTCTAAGAACAAGAAGGTAGATTCTTTCCCGTTCAGCTACAGGACGGCAGAAGCGTGGTTGTTTTG
CTGACTAAACCTTGTGTTGTAATCACGTGCGTGTTTCTTACCACAAGAAAGAGAAACCTAAAGACAGTTGTTCTCAG
TCTGCTGAAAGAAATGAGAAGGATTCACTTACCATGTATGAAAATCCCTCTGTGTTAAGTGTCCAAGAATGCGGCCT
ATGTATTCTCCTTCTAGGTAAAGAGTTTAACTTTTCTCTCCTTCGAATAGGTGAAAACCTGCAGGGTTTACAACTTTC
CGGGAGGCACTCTATGCTGCCTGGAGGAGAATGTGGCCTTACAGACAGTATCTCTGAAGGGGAAAAGAAAGCAGG
TTTCTTCTGTTCTGGTATACCACAGCAGAAGAGATGTAGTGTTGCTAAGTAATCCCTATGTTACCAGGACATGCATG
TTTCCAC
>JACANZ010316682.1
ATTCACTGCATGAAAGAAAAAGCAAAATACAGAGTTCTCATTCAAGTCGGTTAACTTAATCAATTAATTAAGCTATCA
GTTTTGAGAAAACAGTTTTCTCACACGAGCCACAAACTGTCTCATACATTCTCTGAATGGGAAAAGAACGTCAAGCT
TCTCCCTTTCCAGAACTTGAGTCCCCTGAGTTTTCAAATGTTTCCTAAGGGAACACACGTTTTGCATTCTACGAAGA
AGGCAGATTGTTTCCCGTTCAATGAAAGGACTGCAAAAGCGATCTAATTTTGCTGACTAGAACTTGTGCTTTAAGGA
CGTGCGCGTTTCCTATTCACTGGAAGAAAGGGAAACCGAAATACAGTTTCTCTCCTTCGGGCGGAATCAGTAAGAA
AGATTAACTTACTGTGCTTAAAAATCCCTTTGTGTTACACGCACCAGAATGTTGCATATACATTCTTTGCCTGGGTAA
AGAGTTTATTTTCTCCCCTTGGATTACATGAAGCCATGCGGGGTTTAAGTGTTTTCAGTAGGCACTCTGTGCTGTCC
GGCTAACACTATGCCTCATGGACCGTGCTTTAAAGGGAGAAACAATTCGAGTTTCAGCTGTTCAAGTATACGGGA
GCAACAGAGATGGCGTTTTGCTAAGTGAACATTGCTTTAGCAGAAAGTGCATACTCCCAT
>JACANZ010418236.1
TTTACTGCATGAAAGAGAAAGCAAAATACAGCATTCTCATTCAAGTTGGTTAACTTAATCAAGTGAATGAAGTAACGA
GTTTTGATAAAACTCCTTCCTCACACCGTCCACAAGTTTTCTCAATCTCTTAGTGGGATAAGAATGACAAGTCTCTCC
CTTTCAAGTACTTGAATCCCTCGGATTTTGAGATGTTTTCTAAGGGCACTCTGCAGTACCCAGAATGAAGGTACCTT
TCCCATTCTAGGAAGAACAAGGCAGATTGCTTCGCATTCAGCTAAAGGAAGGCAGAAGCGGTGCAGGTTTGCTGAC
TAAACTTTGTATTGTAGATGTTTCTAATTCACTTCGAGAAAGAGAAACCTAAAGACAGTTACTCTCAGTCTGGTGGAT
CAAGACAGAAGGATTCACTTACCGTGTTTGAAAAATCCTCTGTGTTACATGCCCAAGAATATTGCTATGCATTCTCC
ACCTGGGTAAATATTTTGACTTTTCTCCCGTTAGAGACGTGAAGCCCTGCAGAGTTTAGAGGTTTCCAGGAGGCAC
TTTGTGCTGCCCAGATGAAAATGCAGGCCTACAAACGGTATCTCTGAAGGAGAAACAATGTGCATTTCTCACTTTG
ACATACAAGACAGAAGTAGGGGTTGTACTTTTGCGAAGCAAACACTATATTAGCAGGACGTGCATGTTCCCAN
>JACANZ010596973.1
ATTCACCACACGCAAGAGAAAGCAAAATACAGCGTTCTCTTTCAGGCGCTTAACTGTATGTAGTGAAGTTTTCAGTT
TTGTTAGAAGAAACAGTCCGCACACCACAAGCTGCTTTCACAAATGTTCTCAGTGGCATGAGAGTGCAGAGTGCTG
GCCTAGCAAGTATTTGAATCTCCTGCTTTTCAAAGGAATCTTTGCATACCAGGGAATGAAGGTACTACTGGCTTTCA
AAGCAAAACAGGCAGATTTTTTGCCTTTAGCCAGAGGACGGCAGAAGAAGTTCAGTTTTGCTGACCAGATCTTGTG
TGCTTAACACGAGCATGTAGCGTCTTCCCTGCTTGAAAAGGTAAGCAAGAGCTTGTTGCTCTCAGTCTGGGGATGA
AGGGAGAATAATTGACTTTTCCTGTTTGAAAATTCCTCTGTGTTACATACCAAAGAAAGTTGCCTGTGGATTGTCTAC
ATGGATAAAGTGTATGACGTTTCTCTCTTTCAAGTAGTTGAAGCCCTTCAGGGTTTCCCGCATTCCGTGAGGCATTC
TGGGTTGCGCGGCTGAGAATACTGCTTTACAGACGGTATCGCTAAAGGAGAAGAGAAACCGCGTTTCTCCCTTTCT
GGAACGCGAAA
>JACANZ010334892.1
CCCCTTCTGCTATAAGACAGCGGCAGGGCTGTAGTTCTGCTAACTAAACAGTATGTTCATCTAGGTAAATTTTCAGT
TTTGGTAGAACAAACAGTCCTCACACCATCAGCTGCTTTCTCAAACATTCTCTCTATGGGAAAAGACTGGGGAGTGT
CTGAATTTCAAGTTCTTGAATTACCTAGTTTTCAAAGGGAACTTTGCCTTAAAGGGAATGAAGGTGACATTTGCATTT
TAAGAAAAAAGGGCAGATTGTTTGCCTTCAGCTAGTGGATGGCAGAAGCGATTCAGTTTTGCTGACTAAAGCTTGT
GTGTTTGACCCGAACATGTATCCGATTTCCTGCAGGAAAGAGAAACCCAATGACAGTTGCACTCCCTCTGGGGGAT
GAAGTGAGTAGGATAGACTTACTCTGTTTGAAAATCCCCTGTATTACATGCTCAAGAATGTTGCCTGTGCTTTGTTTA
CCCGGGTAAAGACTTTGACTTTTCTCTCTTTAGAGTGCGTGAGGCCCTGCTATGTTTACAGCTTTCCGGGAGGTACT
CTGTGCTGTGTGGCCGAGAATGCCGCGTTACAGACGCTATCCCTGAAGGAGGAGAGAAAGGGCCTTTCTCCCCTT
CTGTTACATGGCAGCGGCAGTGCTGTCGTTTTGCTAACTAAACACTATGTTAGCAACACATATATGTTCCCAA
>JACANZ010280849.1
ATTCACTGAATGAAAGAGAAAGCAAAATNNNNNNNNNNTCACACCACCAGCTGCTTTCTCAAACTACATCTCAATGG
GATAATACTGGGGAGTGTATGCCTTTCAAGTACTTTAATCCCCTGGTTTTCAAACTTAACTTTGCTTTACAGGGAATG
AAGGTAACATTTGCATTCTATGAAAACAGGCAGATTGCCTTCAGCTAGAGGACGGCAAAAACGATTCAGTTTTACT
GACTAAACCTTGTGTGGTTGACAAGACCATGTATCCTATTCCTTGCAAGAAAGAGAAACCAAGTTGCTCTCATACTG
GGGAATGAAGTGAGAAGGATTCACTTACTGTGCTTGAAAACCTCTCTGTTTTCCATGCTGAAGAGAGTTGCGTGTAC
ATTGTCTACCTGGGTAAAGAGTCTGAATTATTTCTCTTCCGAGTACCTGAAGGCCTGTGGGGTTTACAGCTTTCCGG
GAATGATTCTGTGCTGTGCGGCTGAGAATGCCGCCTTACAAACGCTATCTCTCAAGGAGAAGACAGAGTGTTTCTC
CCGTTCTGGTATACAACAGCAGCAGCGCTGTAGTTTTCCAGACTAAACACTATGTTAGCAGCAGGGGCAAGATCCA
A

5. DISCUSSÃO

Em ambos os capítulos usamos *reads* curtas (110-150 pb) e agrupamento por similaridade como ponto de partida para a identificação de satDNAs nos genomas estudados. Tendo o genoma de interesse sequenciado e disponível online, essa é a estratégia mais barata para acessar os elementos repetitivos presentes no genoma, dado que os softwares RepeatExplorer e TAREAN são gratuitos e com possibilidade de rodar as análises na plataforma Galaxy, sem dispor de recursos de hardware ou instalação local. Outra grande vantagem desses programas é que é possível fazer identificação *de novo*, sem a necessidade de uma base de dados com elementos previamente descritos, como foi o caso do TMAsat apresentado no capítulo 2. Tanto o RepeatExplorer quanto o TAREAN têm sido amplamente usados para a identificação e caracterização de satDNAs em várias espécies, de plantas a vertebrados (Pagán *et al*., 2012; Araújo *et al*., 2017; Kirov *et al*., 2017; Palácios-Gimenez *et al*., 2018; Pamponét *et al*., 2019; Ulio *et al*., 2019; Utsunomia *et al*., 2019; Boštjančić *et al*., 2020; Sena *et al*., 2020).

Como o RepeatExplorer e o TAREAN apresentam sequências consenso para os possíveis satDNAs, nós conseguimos validar os resultados usando material biológico das espécies de interesse. Elaboramos iniciadores para cada satDNA hipotético que, ao serem usados em PCR com o DNA genômico da espécie-alvo, apresentaram resultados que estavam de acordo com o esperado para satDNAs, ou seja, amplificação em escada, com bandas correspondentes a monômeros, dímeros, trímeros e etc. Após a amplificação, conseguimos clonar as sequências dos satDNAs, garantindo possibilidades de análise das sequências monoméricas de espécies sem genomas sequenciados. Esse foi o caso das sequências do alfa obtidas em *S. ustus* e *S. vanzolinii* e do TMAsat oriundo de *T. inunguis*. A clonagem também foi fundamental para produzirmos sondas usadas nos experimentos de FISH e assim mapear os satDNAs nos cromossomos das espécies-alvo.

Também fizemos uma análise comparativa das sequências de monômeros do mesmo satDNA em espécies diferentes. No capítulo 1 analisamos sequências do alfa em quatro espécies de *Saimiri* e no capítulo 2, sequências do TMAsat em quatro espécies de Sirenia. Nos dois casos não houve agrupamento espécie-específico dos monômeros, o que foge do previsto pela Evolução Combinada, segundo a qual é esperada maior semelhança intraespecífica dos monômeros. Nas duas situações levantamos possibilidades que explicariam a falta de sequências espécie-específicas, desde a amostra usada nas análises, o tempo de divergência entre as espécies, a sequência ser conservada por talvez exercer alguma função biológica e também pela Evolução Combinada não ser o modelo que melhor explique o cenário desses satDNAs.

Para aprofundar os estudos dos satDNAs tanto do capítulo 1 quanto do capítulo 2, será necessário o uso de *reads* longas geradas pelo sequenciamento de terceira geração,

como pelas plataformas Pacific Biosciences (PacBio) e Oxford Nanopore Technologies (Nanopore). Essa abordagem facilitará verificar tanto a sequência completa dos monômeros, quanto caracterizar as cadeias longas típicas dos satélites, principalmente para satDNAs que apresentam HOR. A desvantagem das *reads* curtas é que muitas vezes são menores que os monômeros. Desta forma, a montagem pode não representar bem os monômeros ou as cadeias que formam, já que a alta identidade entre os monômeros acaba sendo um desafio no processo de montagem. Esse foi o caso do alfa e CapA em *Saimiri* e do TMAsat nos peixes-boi. Usando *reads* longas, a montagem gera menos ambiguidade, principalmente em regiões altamente repetitivas. A montagem pode ser até desnecessária para determinadas regiões, já que as *reads* podem chegar a até um Megabase (van Dijk *et al.*, 2018).

6. CONCLUSÕES

Nós investigamos os satDNAs presentes em *S. boliviensis* usando seu genoma sequenciado e o software RepeatExplorer2. Alfa e CapA são os dois satDNAs mais abundantes e estão presentes em *S. sciureus*, *S. vanzolinii* e *S. ustus*. O satDNA Alfa apresentou localização centromérica e homogeneidade nas sequências das quatro espécies analisadas. Essa homogeneidade interespecífica pode estar relacionada à divergência recente das espécies dentro do gênero, mas outro fator a ser considerado é o possível papel desse satDNA na função e manutenção centromérica. Já o CapA está associado à heterocromatina não pericentromérica e com localização variável entre as espécies, o que evidencia seu potencial como marcador citogenético para *Saimiri,* principalmente quando associado ao NF. Como o CapA também está presente em outros gêneros de Platyrrhini é um potencial marcador que pode contribuir para o entendimento da evolução cromossômica e genômica de outros primatas neotropicais.

No capítulo 2 identificamos caracterizamos pela primeira vez um satDNA no genoma do peixe-boi marinho, o TMAsat, com aproximadamente 687 pb e localização centromérica em *T. manatus* e *T. inunguis*. O TMAsat também está presente em duas espécies de Sirenia fora do gênero *Trichechus*: *D. dugon* e *H. gigas* (espécie já extinta). Monômeros de TMAsat dessas quatro espécies não apresentam sequências espécie-específicas, apesar dos ~43 Ma de divergência entre Trichechidae e Dugongidae, contrariando a previsão da Evolução Combinada. Também encontramos poucas sequências do TMAsat em elefantes e hiraxes, mas sem o padrão típico de satDNA nesses genomas, indicando que a sequência do TMAsat se expandiu como um satDNA em Sirenia há menos de 69 Ma, data da divergência entre Sirenia e os outros Paenungulata.

7. REFERÊNCIAS BIBLIOGRÁFICAS

Alfaro, J. W. L. (2018). "Primate phylogeny", in *The International Encyclopedia of Biological Anthropology* (Hoboken, NJ, USA: John Wiley & Sons, Inc.), 1–4. doi:10.1002/9781118584538.ieba0404.

Alfaro, J. W. L., Boubli, J. P., Paim, F. P., Ribas, C. C., Silva, M. N. F. d., Messias, M. R., *et al.* (2015). Biogeography of squirrel monkeys (genus *Saimiri*): South-central Amazon origin and rapid pan-Amazonian diversification of a lowland primate. *Mol. Phylogenet. Evol.* 82, 436–454. doi:10.1016/j.ympev.2014.09.004.

Alkan, C., Ventura, M., Archidiacono, N., Rocchi, M., Sahinalp, S. C., Eichler, E. E. (2007). Organization and evolution of primate centromeric DNA from whole-genome shotgun sequence data. *PLoS Comput. Biol.* 3, 1807–1818. doi:10.1371/journal.pcbi.0030181.

Alves, G., Canavez, F., Seuánez, H., Fanning, T. (1995). Recently amplified satellite DNA in *Callithrix argentata* (Primates, Platyrrhini). *Chromosom. Res.* 3, 207–213. doi:10.1007/BF00713044.

Alves, G., Seuánez, H. N., Fanning, T. (1994). Alpha satellite DNA in neotropical primates (Platyrrhini). *Chromosoma* 103, 262–267. doi:10.1007/BF00352250.

Alves, G., Seuánez, H. N., Fanning, T. (1998). A Clade of New World Primates with Distinctive Alphoid Satellite DNAs. *Mol. Phylogenet. Evol.* 9, 220–224. doi:10.1006/mpev.1997.0462.

Alves, S. ., Silva Júnior, J. S., Ravetta, A. L., Messias, M. R., Lynch Alfaro, J. W. (2021). *Saimiri ustus* (amended version of 2019 assessment). *IUCN Red List Threat. Species 2021*, e.T41538A192584351. doi:https://dx.doi.org/10.2305/IUCN.UK.2021-1.RLTS.T41538A192584351.en.

Anderson, P. K., Domning, D. P. (2009). "Steller's Sea Cow", in *Encyclopedia of Marine Mammals*, orgs. W. F. Perrin, J. G. M. Thewissen, B. Würsig (Elsevier), 1103–1106. doi:10.1016/B978-0-12-373553-9.00252-2.

Araujo, N. P., De Lima, L. G., Dias, G. B., Kuhn, G. C. S., De Melo, A. L., Yonenaga-Yassuda, Y., *et al.* (2017). Identification and characterization of a subtelomeric satellite DNA in Callitrichini monkeys. *DNA Res.* 24, 377–385. doi:10.1093/dnares/dsx010.

Assis, M. F. L., Best, R. C., Barros, R. M. S., Yonenaga-Yassuda, Y. (1998). Cytogenetic study of *Trichechus inunguis* (Amazonian manatee). *Brazilian J. Genet.* 11, 41–50.

Barros, H. M. D. d. R., Meirelles, A. C. O., Luna, F. O., Marmontel, M., Cordeiro-Estrela, P., Santos, N., *et al.* (2017). Cranial and chromosomal geographic variation in manatees (Mammalia: Sirenia: Trichechidae) with the description of the Antillean manatee karyotype in Brazil. *J. Zool. Syst. Evol. Res.* 55, 73–87. doi:10.1111/jzs.12153.

Bender, M. A., Mettler, L. E. (1958). Chromosome Studies of Primates. *Science (80-.).* 128, 186–190. doi:10.2307/1754678.

Bertram, G. C. L., Bertram, C. K. (1973). The modern Sirenia: their distribution and status. *Biol. J. Linn. Soc.* 5, 297–338. doi:10.1111/j.1095-8312.1973.tb00708.x.

Biscotti, M. A., Olmo, E., Heslop-Harrison, J. S. (2015). Repetitive DNA in eukaryotic genomes. *Chromosom. Res.* 23, 415–420. doi:10.1007/s10577-015-9499-z.

Boinski, S., Cropp, S. J. (1999). Disparate data sets resolve squirrel monkey (*Saimiri*)

taxonomy: Implications for behavioral ecology and biomedical usage. *Int. J. Primatol.* 20, 237–256. doi:10.1023/A:1020522519946.

Boštjančić, L. L., Bonassin, L., Anušić, L., Lovrenčić, L., Besendorfer, V., Maguire, I., *et al*. (2020). The *Pontastacus leptodactylus* (Astacidae) Repeatome Provides Insight Into Genome Evolution and Reveals Remarkable Diversity of Satellite DNA. *Front. Genet.* 11. doi:10.3389/FGENE.2020.611745.

Byrne, H., Rylands, A. B., Carneiro, J. C., Alfaro, J. W. L., Bertuol, F., da Silva, M. N. F., *et al*. (2016). Phylogenetic relationships of the New World titi monkeys (*Callicebus*): First appraisal of taxonomy based on molecular evidence. *Front. Zool.* 13, 1–26. doi:10.1186/s12983-016-0142-4.

Cellamare, A., Catacchio, C. R., Alkan, C., Giannuzzi, G., Antonacci, F., Cardone, M. F., *et al*. (2009). New insights into centromere organization and evolution from the white-cheeked Gibbon and marmoset. *Mol. Biol. Evol.* 26, 1889–1900. doi:10.1093/molbev/msp101.

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

Chiatante, G., Capozzi, O., Svartman, M., Perelman, P., Centrone, L., Romanenko, S. S., *et al*. (2017). Centromere repositioning explains fundamental number variability in the New World monkey genus *Saimiri*. *Chromosoma* 126, 519–529. doi:10.1007/s00412-016-0619-0.

Cohen, S., Segal, D. (2009). Extrachromosomal Circular DNA in Eukaryotes: Possible Involvement in the Plasticity of Tandem Repeats. *Cytogenet. Genome Res.* 124, 327–338. doi:10.1159/000218136.

Costello, R. K., Dickinson, C., Rosenberger, A. L., Boinski, S., Szalay, F. S. (1993). "Squirrel Monkey (Genus *Saimiri*) Taxonomy", in *Species, Species Concepts and Primate Evolution* (Springer US), 177–210. doi:10.1007/978-1-4899-3745-2_8.

Cropp, S., Boinski, S. (2000). The Central American squirrel monkey (*Saimiri oerstedii*): Introduced hybrid or endemic species? *Mol. Phylogenet. Evol.* 16, 350–365. doi:10.1006/mpev.2000.0814.

de Souza, É. M. S., Freitas, L., da Silva Ramos, E. K., Selleghin-Veiga, G., Rachid-Ribeiro, M. C., Silva, F. A., *et al*. (2021). The evolutionary history of manatees told by their mitogenomes. *Sci. Rep.* 11, 3564. doi:10.1038/s41598-021-82390-2.

Deutsch, C. J., Self-Sulivan, C., Mignucci-Giannoni, A. (2008). *Trichechus manatus*. *IUCN Red List Threat. Species* e.T22103A9, 36. doi:10.2305/IUCN.UK.2008.RLTS.T22103A9356917.en.

Domning, D. P. (2018). "Sirenian Evolution", in *Encyclopedia of Marine Mammals*, orgs. B. Würsig, J. G. M. Thewissen, e K. M. Kovacs (London, United Kingdom: Academic Press), 856–859. doi:10.1016/b978-0-12-804327-1.00229-6.

Domning, D. P., Hayek, L. C. (1986). Interspecific and Intraspecific Morphological Variation in Manatees (Sirenia: *Trichechus*). *Mar. Mammal Sci.* 2, 87–144. doi:10.1111/j.1748-7692.1986.tb00034.x.

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

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

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

Feliciello, I., Pezer, Ž., Sermek, A., Bruvo Mađarić, B., Ljubić, S., Ugarković, Đ. (2021). Satellite DNA-Mediated Gene Expression Regulation: Physiological and Evolutionary Implication. 145–167. doi:10.1007/978-3-030-74889-0_6.

Fry, K., Salser, W. (1977). Nucleotide sequences of HS-α satellite DNA from kangaroo rat *Dipodomys ordii* and characterization of similar sequences in other rodents. *Cell* 12, 1069–1084. doi:10.1016/0092-8674(77)90170-2.

Goecks, J., Nekrutenko, A., Taylor, J., Afgan, E., Ananda, G., Baker, D., *et al*. (2010). Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* 11, 1–13. doi:10.1186/gb-2010-11-8-r86.

Groves, C., Wilson, D.E., Reeder, D.A.M. (2005). "Order Primates", in *Mammal species of the world: a taxonomic and geographic reference*, orgs. D. Wilson e D. Reeder (Baltimore: The Johns Hopkins University Press), 111–184.

Hartley, G., O'Neill, R. J. (2019). Centromere Repeats: Hidden Gems of the Genome. *Genes 2019, Vol. 10, Page 223* 10, 223. doi:10.3390/GENES10030223.

Hershkovitz, P. (1984). Taxonomy of squirrel monkeys genus *Saimiri* (Cebidae, platyrrhini): A preliminary report with description of a hitherto unnamed form. *Am. J. Primatol.* 7, 155–210. doi:10.1002/ajp.1350070212.

Heymann, E. W., Calouro, A. M., Vermeer, J., Mollinedo, J. M., Silva Júnior, J. S., Shanee, S., *et al*. (2021). *Saimiri boliviensis* (amended version of 2018 assessment). *IUCN Red List Threat. Species 2021*, e.T41536A192584127. doi:https://dx.doi.org/10.2305/IUCN.UK.2021-1.RLTS.T41536A192584127.en.

Jones, T. C., Ma, N. S. F. (1975). Cytogenetics of the squirrel monkey (*Saimiri sciureus*). *Fed. Proc.* 34, 1646–1650. doi:10.1007/978-1-4684-2640-3_3.

Jones, T. C., Thorington, R. W., Hu, M. M., Adams, E., Cooper, R. W. (1973). Karyotypes of squirrel monkeys (*Saimiri sciureus*) from different geographic regions. *Am. J. Phys. Anthropol.* 38, 269–277. doi:10.1002/ajpa.1330380222.

Keith Diagne, L. (2015). *Trichechus senegalensis* (errata version published in 2016). *IUCN Red List Threat. Species*, e.T22104A97168578. doi:https://dx.doi.org/10.2305/IUCN.UK.2015-4.RLTS.T22104A81904980.en.

Kellogg, M. E., Burkett, S., Dennis, T. R., Stone, G., Gray, B. A., McGuire, P. M., *et al*. (2007). Chromosome painting in the manatee supports Afrotheria and Paenungulata. *BMC Evol. Biol.* 7, 6. doi:10.1186/1471-2148-7-6.

Kirov, I. V., Kiseleva, A. V., Van Laere, K., Van Roy, N., Khrustaleva, L. I. (2017). Tandem repeats of *Allium fistulosum* associated with major chromosomal landmarks. *Mol. Genet. Genomics* 292, 453–464. doi:10.1007/s00438-016-1286-9.

Kuhn, G. C. S., Sene, F. M., Moreira-Filho, O., Schwarzacher, T., Heslop-Harrison, J. S. (2008). Sequence analysis, chromosomal distribution and long-range organization show that rapid turnover of new and old pBuM satellite DNA repeats leads to different patterns of variation in seven species of the *Drosophila buzzatii* cluster. *Chromosom. Res.* 16, 307–324. doi:10.1007/s10577-007-1195-1.

Lau, Y. F., Arrighi, F. E. (1976). Studies of the squirrel monkey, *Saimiri sciureus*,

genome i. Cytological characterizations of chromosomal heterozygosity. *Cytogenet. Genome Res.* 17, 51–60. doi:10.1159/000130687.

Lavergne, A., Ruiz-García, M., Catzeflis, F., Lacote, S., Contamin, H., Mercereau-Puijalon, O., *et al*. (2010). Phylogeny and phylogeography of squirrel monkeys (genus *Saimiri*) based on cytochrome b genetic analysis. *Am. J. Primatol.* 72, 242–253. doi:10.1002/ajp.20773.

Lima, C. S., Magalhães, R. F., Marmontel, M., Meirelles, A. C., Carvalho, V. L., Lavergne, A., *et al*. (2019). A hybrid swarm of manatees along the guianas coastline, a peculiar environment under the influence of the Amazon River plume. *An. Acad. Bras. Cienc.* 91, e20190325. doi:10.1590/0001-3765201920190325.

Lima, C. S., Magalhães, R. F., Santos, F. R. (2021). Conservation issues using discordant taxonomic and evolutionary units: A case study of the American manatee (*Trichechus manatus*, Sirenia). *Wildl. Res.* doi:10.1071/WR20197.

Lynch, J. W., Paim, F. P., Rabelo, R. M., Silva Júnior, J. S., de Queiroz, H. . (2021). *Saimiri vanzolinii. IUCN Red List Threat. Species 2021*, e.T19839A17940474. doi:https://dx.doi.org/10.2305/IUCN.UK.2021-1.RLTS.T19839A17940474.en.

Ma, N. S., Jones, T. C. (1975). Added heterochromatin segments in chromosomes of squirrel monkeys (*Saimiri sciureus*). *Folia Primatol. (Basel).* 24, 282–292. doi:10.1159/000155698.

Marmontel, M., de Souza, D., Kendall, S. (2016). *Trichechus inunguis. IUCN Red List Threat. Species*, e.T22102A43793736. doi:https://dx.doi.org/10.2305/IUCN.UK.2016-2.RLTS.T22102A43793736.en.

Marsh, H. (2009). Dugong: *Dugong dugon. Encycl. Mar. Mamm.*, 332–335. doi:10.1016/B978-0-12-373553-9.00080-8.

Marsh, H., O'Shea, T. J., Best, R. C. (1986). Research on sirenians. *Ambio* 15, 177–180. doi:10.2307/4313244.

Marsh, H., Sobtzick, S. (2019). *Dugong Dugon* (amended version of 2015 assessment). . *IUCN Red List Threat. Species 2019*, e.T22104A97168578. doi:https://dx.doi.org/10.2305/IUCN.UK.2015-4.RLTS.T22104A81904980.en.

Martin, R. D. (2012). Primates. *Curr. Biol.* 22, R785–R790. doi:10.1016/j.cub.2012.07.015.

Meštrović, N., Plohl, M., Mravinac, B., Ugarković, D. (1998). Evolution of satellite DNAs from the genus *Palorus* - Experimental evidence for the "library" hypothesis. *Mol. Biol. Evol.* 15, 1062–1068. doi:10.1093/oxfordjournals.molbev.a026005.

Moore, C. M., Harris, C. P., Abee, C. R. (1990). Distribution of chromosomal polymorphisms in three subspecies of squirrel monkeys (genus *Saimiri*). *Cytogenet. Cell Genet.* 53, 118–122. doi:10.1159/000132909.

Nikaido, M., Nishihara, H., Hukumoto, Y., Okada, N. (2003). Ancient SINEs from African endemic mammals. *Mol. Biol. Evol.* 20, 522–527. doi:10.1093/molbev/msg052.

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

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, 792–793. doi: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 Res.* 45, e111. doi:10.1093/nar/gkx257.

Pagán, H. J. T., Macas, J., Novák, P., McCulloch, E. S., Stevens, R. D., Ray, D. A. (2012). Survey sequencing reveals elevated DNA transposon activity, novel elements, and variation in repetitive landscapes among vesper bats. in *Genome Biology and Evolution*, 575–585. doi:10.1093/gbe/evs038.

Paim, F. P., De La Torre, A., Carretero, X., Guzmán-Caro, D. C., Stevenson, P. R., Lynch Alfaro, J. W., *et al*. (2021). *Saimiri cassiquiarensis* (amended version of 2020 assessment). *IUCN Red List Threat. Species 2021*, e.T160940148A192585552. doi:https://dx.doi.org/10.2305/IUCN.UK.2021-1.RLTS.T160940148A192585552.en.

Palacios-Gimenez, O. M., Koelman, J., Palmada-Flores, M., Bradford, T. M., Jones, K. K., Cooper, S. J. B., *et al*. (2020). Comparative analysis of morabine grasshopper genomes reveals highly abundant transposable elements and rapidly proliferating satellite DNA repeats. *BMC Biol. 2020 181* 18, 1–21. doi:10.1186/S12915-020-00925-X.

Palacios-Gimenez, O. M., Milani, D., Lemos, B., Castillo, E. R., Martí, D. A., Ramos, E., *et al*. (2018). Uncovering the evolutionary history of neo-XY sex chromosomes in the grasshopper *Ronderosia bergii* (Orthoptera, Melanoplinae) through satellite DNA analysis. *BMC Evol. Biol.* 18, 1–10. doi:10.1186/s12862-017-1113-x.

Pamponét, V. C. C., Souza, M. M., Silva, G. S., Micheli, F., De Melo, C. A. F., De Oliveira, S. G., *et al*. (2019). Correction to: Low coverage sequencing for repetitive DNA analysis in *Passiflora edulis* Sims: Citogenomic characterization of transposable elements and satellite DNA (BMC Genomics (2019) 20 (262) DOI: 10.1186/s12864-019-5576-6). *BMC Genomics* 20, 1–17. doi:10.1186/s12864-019-5678-1.

Perelman, P., Johnson, W. E., Roos, C., Seuánez, H. N., Horvath, J. E., Moreira, M. A. M., *et al*. (2011). A molecular phylogeny of living primates. *PLoS Genet.* 7, 1–17. doi:10.1371/journal.pgen.1001342.

Plohl, M., Metrovic, N., Mravinac, B. (2012). "Satellite DNA Evolution", in *Repetitive DNA*, org. M. A. Garrido-Ramos (Basel: Karger Publishers), 126–152. doi:10.1159/000337122.

Rocchi, M., Archidiacono, N., Schempp, W., Capozzi, O., Stanyon, R. (2012). Centromere repositioning in mammals. *Heredity (Edinb).* 108, 59–67. doi:10.1038/hdy.2011.101.

Rylands, A. B., Mittermeier, R. A. (2013). "Family Cebidae (capuchines and squirrel monkeys)", in *Handbook of the Mammals of the World, vol. 3. Primates*, orgs. R. A. Mittermeier, A. B. Rylands, e D. E. Wilson (Barcelona: Lynx Edicions), 348–389.

Schneider, H., Sampaio, I. (2015). The systematics and evolution of New World primates - A review. *Mol. Phylogenet. Evol.* 82, 348–357. doi:10.1016/j.ympev.2013.10.017.

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). *Sci. Rep.* 10, 1–11. doi:10.1038/s41598-020-76199-8.

Shapiro, J. A., von Sternberg, R. (2005). Why repetitive DNA is essential to genome function. *Biol. Rev.* 80, 227–250. doi:10.1017/S1464793104006657.

Silva, B. S. M. L., Heringer, P., Dias, G. B., Svartman, M., Kuhn, G. C. S. (2019). De novo identification of satellite DNAs in the sequenced genomes of *Drosophila virilis* and *D. americana* using the RepeatExplorer and TAREAN pipelines. *PLoS One* 14, e0223466. doi:10.1371/journal.pone.0223466.

Silva Júnior, J. S., Ravetta, A. L., Lynch Alfaro, J. W., Valença-Montenegro, M. M. (2021). *Saimiri collinsi* (amended version of 2020 assessment). *The IUCN Red List Threat. Species 2021*, e.T70610928A192585417. doi:https://dx.doi.org/10.2305/IUCN.UK.2021-1.RLTS.T70610928A192585417.en.

Solano-Rojas, D. (2021). *Saimiri oerstedii. IUCN Red List Threat. Species 2021*, e.T19836A17940807. doi:https://dx.doi.org/10.2305/IUCN.UK.2021-1.RLTS.T19836A17940807.en.

Stanyon, R., Consigliere, S., Müller, S., Morescalchi, A., Neusser, M., Wienberg, J. (2000). Fluorescence in situ hybridization (FISH) maps chromosomal homologies between the dusky titi and squirrel monkey. *Am. J. Primatol.* 50, 95–107. doi:10.1002/(SICI)1098-2345(200002)50:2<95::AID-AJP1>3.0.CO;2-8.

Sujiwattanarat, P., Thapana, W., Srikulnath, K., Hirai, Y., Hirai, H., Koga, A. (2015). Higher-order repeat structure in alpha satellite DNA occurs in New World monkeys and is not confined to hominoids. *Sci. Rep.* 5, 1–10. doi:10.1038/srep10315.

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

Tautz, D. (1993). Notes on the definition and nomenclature of tandemly repetitive DNA sequences. *EXS* 67, 21–28. doi:10.1007/978-3-0348-8583-6_2.

Thakur, J., Packiaraj, J., Henikoff, S. (2021). Sequence, Chromatin and Evolution of Satellite DNA. *Int. J. Mol. Sci.* 22, 4309. doi:10.3390/ijms22094309.

Thorington, R. W. (1985). "The Taxonomy and Distribution of Squirrel Monkeys (*Saimiri*)", in *Handbook of Squirrel Monkey Research* (Springer US), 1–33. doi:10.1007/978-1-4757-0812-7_1.

Upham, N. S., Esselstyn, J. A., Jetz, W. (2019). Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLOS Biol.* 17, e3000494. doi:10.1371/journal.pbio.3000494.

Utsunomia, R., Silva, D. M. Z. de A., Ruiz-Ruano, F. J., Goes, C. A. G., Melo, S., Ramos, L. P., *et al*. (2019). Satellitome landscape analysis of *Megaleporinus macrocephalus* (Teleostei, Anostomidae) reveals intense accumulation of satellite sequences on the heteromorphic sex chromosome. *Sci. Rep.* 9, 5856. doi:10.1038/s41598-019-42383-8.

Valeri, M. P., Dias, G. B., Pereira, V. D. S., Kuhn, G. C. S., Svartman, M. (2018). An eutherian intronic sequence gave rise to a major satellite DNA in Platyrrhini. *Biol. Lett.* 14, 20170686. doi:10.1098/rsbl.2017.0686.

van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., Thermes, C. (2018). The Third Revolution in Sequencing Technology. *Trends Genet.* 34, 666–681.

doi:10.1016/j.tig.2018.05.008.

Vianna, J. A., Bonde, R. K., Caballero, S., Giraldo, J. P., Lima, R. P., Clark, A., *et al*. (2006). Phylogeography, phylogeny and hybridization in trichechid sirenians: Implications for manatee conservation. *Mol. Ecol.* 15, 433–447. doi:10.1111/j.1365-294X.2005.02771.x.

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, 137–144. doi:10.1007/s00412-004-0301-9.

Wicker, T., Sabot, F., Hua-Van, A., Bennetzen, J. L., Capy, P., Chalhoub, B., *et al*. (2007). A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.* 8, 973–982. doi:10.1038/nrg2165.

Willard, H. F. (1985). Chromosome-specific organization of human alpha satellite DNA. *Am. J. Hum. Genet.* 37, 524–532.

Yonenaga Yassuda, Y., Chu, T. H. (1985). Chromosome banding patterns of *Saimiri vanzolinii* Ayres, 1965 (Primates, Cebidae). *Pap. Avulsos Zool.* 36, 165–8.