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

043 Valeri, Mirela Pelizaro.

Identificação de novo e caracterização de DNAs satélites em macacos-decheiro (Saimiri, Cebidae, Platyrrhini) e peixes-boi (Trichechus, Trichechidae, Sirenia) [manuscrito] / Mirela Pelizaro Valeri. – 2022. 99 f. : il. ; 29,5 cm.

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

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

1. Genética. 2. Sequências de Repetição em Tandem. 3. Trichechus inunguis. 4. Saimiri. I. Svartman, Marta. II. Kuhn, Gustavo Campos e Silva. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 575

Ficha catalográfica elaborada pela bibliotecária Fabiane C M Reis - CRB 6 - 2680

SEI/UFMG - 1313573 - Ata de defesa de Dissertação/Tese

https://sei.ufmg.br/sei/controlador.php?acao=documento_imprimir_web...



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. Seguiuse 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:

seil assinatura eletrônica	Documento assinado eletronicamente por Edivaldo Herculano Correa de Oliveira, Usuário Externo , em 17/03/2022, às 18:08, conforme horário oficial de Brasília, com fundamento no art. 5º do <u>Decreto nº</u> 10.543, de 13 de novembro de 2020.
seil assinatura eletrônica	Documento assinado eletronicamente por Marta Svartman , Professora do Magistério Superior , em 17/03/2022, às 19:04, conforme horário oficial de Brasília, com fundamento no art. 5º do <u>Decreto nº</u> 10.543, de 13 de novembro de 2020.
seil assinatura eletrônica	Documento assinado eletronicamente por Diogo Cavalcanti Cabral de Mello, Usuário Externo , em 18/03/2022, às 19:13, conforme horário oficial de Brasília, com fundamento no art. 5º do <u>Decreto nº 10.543, de 13 de novembro de 2020</u> .
seil assinatura eletrônica	Documento assinado eletronicamente por Gustavo Campos e Silva Kuhn, Professor do Magistério Superior , em 21/03/2022, às 11:14, conforme horário oficial de Brasília, com fundamento no art. 5º do <u>Decreto nº 10.543, de 13 de novembro de 2020</u> .
seil assinatura eletrônica	Documento assinado eletronicamente por Fernando Araujo Perini, Professor do Magistério Superior, em 21/03/2022, às 22:16, conforme horário oficial de Brasília, com fundamento no art. 5º do <u>Decreto nº</u> 10.543, de 13 de novembro de 2020.
seil assinatura eletrônica	Documento assinado eletronicamente por Francisco Pereira Lobo, Professor do Magistério Superior, em 22/03/2022, às 12:47, conforme horário oficial de Brasília, com fundamento no art. 5º do <u>Decreto nº</u> 10.543, de 13 de novembro de 2020.

SEI/UFMG - 1313573 - Ata de defesa de Dissertação/Tese

https://sei.ufmg.br/sei/controlador.php?acao=documento_imprimir_web...



A autenticidade deste documento pode ser conferida no site <u>https://sei.ufmg.br</u> /<u>sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0</u>, informando o código verificador 1313573 e o código CRC DE320D8D.

Referência: Processo nº 23072.214252/2022-90

SEI nº 1313573

AGRADECIMENTOS

À minha orientadora Dra. Marta Svartman, pela oportunidade, confiança, ensinamentos, dedicação e amizade durante os sete anos de convivência que me proporcionaram profícuo desenvolvimento acadêmico.

Ao meu coorientador Dr. Gustavo campos e Silva Kuhn, pela disponibilidade, solicitude, ensinamentos e convivência indispensáveis para meu desenvolvimento acadêmico.

Ao Dr. Guilherme Borges Dias, por ter me guiado em todas as etapas, me ensinado tanto e proporcionado momentos incríveis dentro e fora do laboratório. Gui, você é uma pessoa maravilhosa.

Às queridas Alice Alves do Espírito Santo e Radarane Santos Sena que me ajudaram a desenvolver o trabalho de todas as formas possíveis, por toda nossa amizade, dores e alegrias compartilhadas. Vocês e respectivas famílias me acolheram em Belo Horizonte de tal forma que quase não percebi estar tão longe da minha própria.

Aos atuais e ex-integrantes do Laboratório de Citogenômica Evolutiva por todo apoio, auxílio, amizade, risadas e boa convivência. Vocês formam um grupo incrível e que adorei fazer parte.

À querida Tatiane Melo Preisser sempre super solícita e eficiente que me auxiliou (socorreu) em diversos experimentos, sem contar a amizade e suporte emocional.

À equipe dos laboratórios que estiveram de portas abertas e me ajudaram em vários momentos: Laboratório de Biodiversidade e Evolução Molecular (LBEM), Laboratório de Biotecnologia e Marcadores Moleculares, Laboratório de Diversidade Genética Humana (LDGH), Laboratório de Genética de Microrganismos, Laboratório de Genética de Populações, Laboratório de Genética Experimental (LGEX), Laboratório de Genética Funcional e Laboratório de Tecnologia Genética (TecnoGen).

À Yatiyo Yonenaga-Yassuda, Camila Nascimento Moreira e lara Braga Sommer por terem sido tão solícitas e responsáveis por ceder parte das amostras que usei para desenvolver essa tese.

Ao técnico do laboratório Daniel dos Santos Filho, pelo auxílio laboratorial e convivência.

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.

À CAPES, pela bolsa de doutorado e à FAPEMIG e ao CNPq pelo financiamento da pesquisa.

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 seguenciado 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 seguências 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

Figura 1.	Localização preferencial dos satDNAs nas regiões heterocromáticas e	
	organização de satDNAs em <i>higher-order repeat</i> (HOR)	4
Figura 2.	Mapa representando a distribuição do gênero Trichechus no continente	
	americano	11

Capítulo 1

Representative Saimiri karyotype	16
Metaphases of Saimiri after FISH with the alpha satDNA	
probe	17
FISH with CapA in the karyotypes of Saimiri	18
Neighbor-joining clustering and Non-metric	
multidimensional scaling (NMDS) of evolutionary	
divergence among of Saimiri alpha sequences	19
Schemes of Saimiri chromosomes showing the	
localization of the alpha and CapA satDNAs in Saimiri	20
Representative Saimiri karyotypes	25
CBG-banded Saimiri karyotypes of Saimiri	26
Agarose gel (1%) showing the PCR products of Saimiri	
species using a) Alpha and b) CapA primers	27
Phylogenetic analysis of alpha sequences of Saimiri	28
	Representative <i>Saimiri</i> karyotype Metaphases of <i>Saimiri</i> after FISH with the alpha satDNA probe FISH with CapA in the karyotypes of <i>Saimiri</i> Neighbor-joining clustering and Non-metric multidimensional scaling (NMDS) of evolutionary divergence among of <i>Saimiri</i> alpha sequences Schemes of <i>Saimiri</i> chromosomes showing the localization of the alpha and CapA satDNAs in <i>Saimiri</i> Representative <i>Saimiri</i> karyotypes CBG-banded <i>Saimiri</i> karyotypes of <i>Saimiri</i> Agarose gel (1%) showing the PCR products of <i>Saimiri</i> species using a) Alpha and b) CapA primers Phylogenetic analysis of alpha sequences of <i>Saimiri</i>

Capítulo 2

Figure 1.	TMAsat characteristics in the analyzed species	34
Figure 2.	Metaphases of <i>T. manatus</i> (TMA) and <i>T. inunguis</i> (TIN) after FISH with the TMAsat probe	35
Figure 3.	Sliding window analysis and Neighbor-joining tree with TMAsat sequences of Sirenia species	36
Supplementary Figure 1.	TMAsat consensus sequence generated by TAREAN	41
Supplementary Figure 2.	Alignment of 66 monomers manually isolated from the <i>Trichechus manatus</i> assembled reference genome	42
Supplementary Figure 3.	Agarose gel (1%) showing the PCR products of <i>Trichechus inunguis</i> using the TMAsat primers	50
Supplementary Figure 4.	Alignment of five TMAsat monomers cloned and sequenced from the <i>Trichechus inunguis</i> genome	51
Supplementary Figure 5.	Metaphases of <i>T. manatus</i> (TMA) and <i>T. inunguis</i> (TIN) after FISH with the TMAsat probe	52
Supplementary Figure 6.	Alignment of 50 monomers manually isolated from the	52
Supplementary Figure 7.	Alignment of 40 monomers manually isolated from the	55
	Hydrodamalis gigas assembled reference genome	59

LISTA DE TABELAS

Capítulo 1

Table 1 Supplementary Table S1	Identification, sex, collection site, fundamental number and morphology of chromosome pairs 5 and 15 of the analyzed specimens Clusters retrieved by RepeatExplorer from a sample of the sequencing reads of <i>Saimiri boliviensis</i>	15 29
Capítulo 2		
Table 1	The two putative CENPB box-like motifs identified in the MCS of TMAsat from <i>T. manatus</i> , <i>T. inunguis</i> , <i>D. dugon</i> , and <i>H. gigas</i>	36
Supplementary Table 1	Estimates of average nucleotide divergence over all sequence pairs within and between species	65
Supplementary Table 2	Blast searches using the TMAsat Most Common Sequence (MCS) from <i>Trichechus manatus</i> as query against the GenBank Mammalia (NCBI:txid40674) wgs database excluding Sirenia (NCBI:txid9774)	66

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	14
1.	.1 DNAs satélites	14
1.	.2 Saimiri	17
1.	.3 Sirenia	20
2.	OBJETIVOS	23
3.	CAPÍTULO 1	24
4.	CAPÍTULO 2	41
5.	DISCUSSÃO	101
6.	CONCLUSÕES	103
7.	REFERÊNCIAS BIBLIOGRÁFICAS	104

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 guatro 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 guase 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.



natureresearch

Check for updates

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

Mirela Pelizaro Valeri¹, Guilherme Borges Dias², Camila Nascimento Moreira³, Yatiyo Yonenaga-Yassuda³, Roscoe Stanyon⁴, Gustavo Campos e Silva Kuhn¹ & Marta Svartman¹⊠

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 debated³. Even the number of species is uncertain, historically ranging from one to 16 distinguished species⁴⁻¹¹. 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 7812. 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,

¹Laboratório de Citogenômica Evolutiva, Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil. ²Department of Genetics and Institute of Bioinformatics, University of Georgia, Athens, GA, United States of America. ³Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, SP, Brazil. ⁴Department 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	Α
S. sciureus	SSC 782	Male	Presidente Figueiredo - Amazonas	74	A	Α
S. sciureus	SSC 770	Female	Santarém – Pará	74	Α	Α
S. sciureus	SSC2	Male	Unknown	74	Α	Α
S. vanzolinii	SVA 321	Female	Lake Mamirauá, Tefé - Amazonas	76	SM	Α
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 in²⁰]. 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 satDNAs²¹. 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 primates²⁴, frogs²⁵, fish²⁶ and plants²⁷.

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 conclusions¹² 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 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 μ m.

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 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 = $10 \,\mu$ m.

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 mammals³¹ 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 = 10μ m.

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}.



www.nature.com/scientificreports/



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 15¹². 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 marker²⁴.





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 sequences³⁷, and in *Callitrix jacchus* alpha repeats are divided into seven clusters that showed different hybridization patterns among chromosomes²⁹.

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-over³¹. 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 sequences¹².

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 inversions³². 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 9³². 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 ²³² 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 Mya³ starting with the divergence between *S. boliviensis* and the ancestor to all other squirrel monkeys at 1.4–1.6 Mya². 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 species³ 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 NWMs³¹. 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 loss⁴¹.

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 obtained from cultured cells according to Stanyon and Galleni⁴². GTG- and CBG-banding were performed according to Seabright⁴³ and Sumner⁴⁴, 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 Chu⁴⁶.

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 (~100bp 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 parameters⁴⁹. 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 (ACTTCCCACTGACCTGTCTT) and CapA-R (GGGCTGATGCTTAATGTAAGCA). 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 MUSCLE⁵⁰ 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 *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 vegan⁵³, representing the divergence between sequences values in Euclidian distances for two dimensions. RStudio v1.1.463⁵⁴ 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.

Received: 3 December 2019; Accepted: 15 April 2020; Published online: 08 May 2020

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).
- 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).
- 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).
- 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).
- 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).
- 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).
- 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).
- Cropp, S. & Boinski, S. The Central American squirrel monkey (Saimiri oerstedii): Introduced hybrid or endemic species? *Mol. Phylogenet. Evol.* 16, 350–365 (2000).
 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).
- 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 karvotypes, 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.
- Stanyon, R. & Galleni, L. A rapid fibroblast culture technique for high resolution karyotypes. *Bolletino di Zool.* 58, 81–83 (1991).
 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

This work was supported by a grant from the Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq process 407262/2013-0) to MS and RS. MPV and GBD received fellowships from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). MS and GK are supported by fellowships from CNPq (Process 310433/2018-5 and 308386/2018-3, respectively).

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; 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.

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

© The Author(s) 2020
a)						b)					
randra A	X	2	3	X	5	6	- Contraction	2	3	5	5	6
	1	8)) 9	10	ir	12	Total 7	8	9	10		12
	13	14	15	16	1 7	a 18	13	14	15	16	17	18 B
	18 19	20	21			хү	19 an	€ ∄ 20	60 21			XY
			c)			3		100 A				
						38	10		12			
				13	14	1 5	1 6	8 17	6 18			
				展 19	A € 20	♣ ♠ 21		10	x y µm			

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.

SBO1							SSC 78	2					
5)	and a	and the second	II.	6	ł.	λ			1		5	6
K	8	9	10	11	12	1			8	and the second s	10		12
13	14	15	16	17	18	l.	13		14	15	16	17	8 18
19	20	21			xy	č	1 9		20	8 8 21			XY
SSC2							SVA	321					
		Constant Constant 3				a 100			2	88		Ŭ.	6
	K s	9	total		12	STREET,			8	17 00 8 8 9	10		8/10 8/10 12
13	4 14	15	1 16	1 7	18	6	13		គឺ ល័ 14	1 5	0 16	8 8 17	4 6 18
1 9	a a 20	4 4 21	SUS	6 739	XY XY	,	* * 19		20	21			XY
					2		Posto 4	MI S					
					11	8	10		12				
			13	8	14	15		17	18				
			19	K.	44 20	å å 21			N Y	10	μm		

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/).

		Genome	Proportion of Similarity	0
Cluster	Read Number	Proportion [%] ^a	Hits to Other Clusters	Annotation
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

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

^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.

BRIEF RESEARCH REPORT published: 24 August 2021 doi: 10.3389/fgene.2021.694866



First Description of a Satellite DNA in Manatees' Centromeric Regions

Mirela Pelizaro Valeri¹, Guilherme Borges Dias², Alice Alves do Espírito Santo¹, Camila Nascimento Moreira³, Yatiyo Yonenaga-Yassuda³, Iara Braga Sommer⁴, Gustavo C. S. Kuhn¹ and Marta Svartman^{1*}

¹Laboratório de Citogenômica Evolutiva, Departamento de Genética, Ecologia e Evolução, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, Belo Horizonte, Brazil, ²Department of Genetics and Institute of Bioinformatics, University of Georgia, Athens, GA, United States, ³Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil, ⁴Centro Nacional de Pesquisa e Conservação da Biodiversidade Marinha do Nordeste, Instituto Chico Mendes de Conservação da Biodiversidade, Brasilia, Brazil

OPEN ACCESS

Edited by:

Francisco J. Ruiz-Ruano, University of East Anglia, United Kingdom

Reviewed by:

Nevenka Mestrovic, Rudjer Boskovic Institute, Croatia Isidoro Feliciello, University of Naples Federico II, Italy

irontiers

in Genetics

*Correspondence: Marta Svartman svartmanm@ufmg.br

Specialty section:

This article was submitted to Evolutionary and Population Genetics, a section of the journal Frontiers in Genetics

Received: 13 April 2021 Accepted: 30 July 2021 Published: 24 August 2021

Citation:

Valeri MP, Dias GB, Espírito Santo AA, Moreira CN, Yonenaga-Yassuda Y, Sommer IB, Kuhn GCS and Svartman M (2021) First Description of a Satellite DNA in Manatees' Centromeric Regions. Front. Genet. 12:694866. doi: 10.3389/fgene.2021.694866

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 speciesspecific 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).

43

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 (~100bp 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 50bp 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 30s; 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).

44

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 200ng 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 v4.3.3¹ (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.

¹https://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





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 (Supplementary genome 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 56,989-75,023bp), 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 687bp 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

Centromeric Satellite DNA in Manatees



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 196–212bp	Position 518–534bp
ECD	NTTCGNNNNANNCGGGN	NTTCGNNNNANNCGGGN
L. africana and D. novemcinctus	CTTTGCCGAGAACGGAG	CTTTGCCGAGAACGGAG
T. manatus	CTTTGCATAACAGGGAA	TTTACAGCTTTCCGGGA
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<u>G</u>CAGC<u>A</u>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 Centromeric Satellite DNA in Manatees

48

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), ninebanded 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 49

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

MV and AE received fellowships from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES). MS and GK are supported by fellowships from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq; Processes 310433/2018-5 and 308386/2018-3, respectively). This work was funded with grants from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG).

ACKNOWLEDGMENTS

We thank the Programa de Pós-Graduação em Genética, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais, for their support in the execution and publishing of this work.

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

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2021.694866/ full#supplementary-material

- 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

- 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/S146479310 4006657
- 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, D., 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

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Valeri, Dias, Espírito Santo, Moreira, Yonenaga-Yassuda, Sommer, Kuhn and Svartman. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Supplementary Material

>TMAsat

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.

	110	120	130	140	150	160	170	180	190	200	210
Consensus	- TGTTAGAAGAA	ACAGTCCGC	CCACCAC-	CAGC-	TGCTTTCTCAAA	CGTTCTCTCAGTGGGA	- GAAG	ACTGGAGAGT	GCCTGCCTTGCA	- GTATTT	GAATCC
Identity				_							
	TTTTCAAC	ACACTCO	CHACCAC		TACTTAL	TTTCTCTCLCTCLC	6446	LETCOLOLOT	COTOC TTOCA	CTART.	TOTATOO
1. NW004444102.1_1		AACAGICCAC	CAACCAG	AAGC	TGCCTTTCTGAAA		GAAG	ACTGGAGAGI	ACCIGCIIIGCA	GTAATT	GAATCC
2. NW004444102.1_2	TGTTAGAAGAA	AAGAGICCGC	CCALCAL	AAGC	IGCITICICAAA		GAAG	GCIGGAGAGI	GCAGGICTAGCA	GIATTI	GAATAC
3. NW004444102.1_3	- IGTIAGAAGAA	AAGAGICCGC	ACACLAC-	AAGC	IGCTITCICAAA		GAAG	GCAGCAGAGI	GAAGGCCTAGCAA	GAATTI	GAATGC
4. NW004444102.1_4	- IGIIAGAAGAA	AAGAGICAGC	CCACCCC-	AAGC	IGCTITCICAAA	GITCICICAGIGGGA	GAAG	GCTGGGGGAGT	GCAGGCCTTGCA	GIATII	GAATC
5. NW004444102.1_5	- IGGIAGAAGAA	AACACICCIC	ACAACAC-	AAGC	IGACIICICAAA	GGTTCTCTCAGTGGGA	AAG	ACIGGAGAGI	GCCTGCCTTGCAA	AATACII	GAAICC
6. NW04444102.1_6	- I GGIAGAAGAA	AACAGICCAC	CAACAG	CAGC-	IGCTITCICAAA	AGTICICICAGIGAGA	AAG	ACTGAGAGI	GCCTGCCTTGCAA	A-GIACII	GAAICC
7. NW004444102.1_7	- IGGIAGAAGAA	ACAGICCG	ACACCAC-	- CAGA	IGACIICICAAA		AAG	ACIGGAGAGI	GCCTGCCTTGCA/	ACACII	GAAICC
8. NW004448733.1	- IGTIAGAAGAA	AAGAGICCGC	CCAACAG	AAGC	IGCTITCICAAA	GITCICICAGIGGGA	GAAG	GCTGGAGAGI	GCAGGCCTAGCA	A-GIAIII	GAATAC
9. NW004444141.1_1	- T G T T C G A A G A A	AAGAGTCCCC	CCAACAC-	AAGC	TGCTTTCTC	GTTCTCTCAGTGGGA	GAAG	GCTGGAGAGT	GCGGGCCTTGCAA	A - GGATTA	CAATCC
10. NW004444141.1_2	- I G I I A G A A G A A	AAGGGICCCC	CCAACAC-	AAGC	IGCITICICIAA	GIICICICAGIGGG-	GAAG	GCIGGAGAGI	GCAGGCCTAGCA	GIAIII	CAAICC
11. NW004444141.1_3	- I GCI GGAAAAA	AACTICCAC	ALACCAC-	CAGC -	IGACCICICAAA	GITCICICAGIGGGA	AAG	ACTGGGGAGT	GIGIGUCGIGUA	GTACTI	GAATCC
12. NW004444141.1_4		AACAGICCIG	ACACCGC -	- CAGC -	TACTITCIGAAA	ATTICICICAATGGGGG	AAAG	ACIGGGANI	GCIGCCII	A-GIACII	GAATCC
13. NW004444141.1_5	- I GGI GAACAA	ACAGICCIG	ACACCAC-	CIGC-	TACTICAGAAA	CATICICICAAIGGGG	AAL	ACIGGIGAAI	GICIGCCITICA	A-GIACII	GAATCC
14. NW004444141.1_6	- TGATAAAGC	GCAGTCC	ACACCAL	CACA	AGCGGTCTCAAA	CATTCTCTCCATGGGA	AAG	ALTIGEGAGT	I C T C C G T T C A /	- GTACTT	GAATCC
15. NW004444141.1_7	- IGAI AAHAC	AGANICCIG	ACAACA	CACA	AGCITICIAAAA	CATICCCCCAGI	AAG	ACIGGAGCAI		GGTAGTI	GAATCC
16. NW004444013.1_1	- IGTTAAAAGAA	AAGAGICCGC	CCAACAC-	AAGC	IGCTITCICCAA	GITCICICAGAGGGA	GAAG	GCTGGAGAGT	GCAGGCCTAGCAA	A-GIAIII	GACIAC
17. NW004444013.1_2	TGTGAGAAGAA	AAGICCGC	CLALCAL-	AAGC	TGCTTTCTCAAA		GAAG	GCTGGAGAGI	GCAGGCCTAGCAA	GIAITI	GAATCC
18. NW004444013.1_3	- I G I I A G A A G A A	AACAGICCGC	GGGACAC-	AAG	TGUITICICAAA	GITCICICAGIGGGA	GAAG	ACTGGAGAGT	GAGGGCCTAGCA		GAATC
19. NW004444013.1_4	TGTTAGAAGAA	AACAGICCGC	CLAC AC-	AAGC	TGCTTTCTCAAA	ATTCTCTCAGTGGGA	GAAA	ACAGGAGAGI	GCCGGTCTAGCAC	J-GIAIII	CAATCC
20. NW004444013.1_5	- IGGIAGA - GAA	AACAGICCGC	CAAACAC-	CAGC-	IGCTTTCTCAGA	AGTICICICAGIGGGA	GAAG	ACTGG-GAGT	GEETGEETGEAA	GTATT	GAATCC
21. NW004444013.1_6	- IGIIAGAIGAA	AACACICCGC	CGACCAC-	CAGC-	AGCTTTCTGAA	CGTTCTCTCAGTGGGG-	AAG	ACTGGALAGT		GTATT	GAATCC
22. NW004444013.1_/		AACAGICCIC	ACACCA	- CAGC -	TGCTTTCTCAAA		AAAG	ACTGGGAAGT	GECTGAATTECAA		GAATCC
23. NW004443969.1_1	TGTTAGAAGAA	ACCGICC	CCACCAC-	CAGC -	TGCTTTCTGAAA		GAAG	ACTGGAGAGI	GEETECTTGEAA	GTATT	GAATCC
24. NW004443969.1_2	TGTTAGAAGAA	AACIGICCAC	CCACCAC-	CAGC	TGCTTTCTCAAA	ALICICICAGIGGGA	GAAG	ACTGCAGAGT	GEETGEETTGEAA	GTATT	GCATCC
25. NW004443969.1_3	- IGTTAGAAGAA	AACIGICCGC	CCACCAC-	CAGC	TGCATTCTCAGA	COT CTCAGE	GAAG	AAAGCAGAGT	CCTGCCTGCA	GTATT	GAATCC
26. NW004443969.1_4	GGTTAGAAGAA	AACIGC	CCACCAC-	CACC	TGCTTTCTAAAA	GGTCGTCTCAGTGACA	GAAG		GEETGEETTGEAA	GTATT	GAATCC
27. NW004443969.1_5	TOTTAGAAGAA	ACAGICCGC	CCACCAC-	CAGC	A C C T T T C T C A A A	GGTTCTCTCACTCCCA	AAG	ACTGGGGAGT	GCETCCCTTCCA	GTATTT	AAATCC
20. 110004443969.1_6	TGTTAGAAGAA	ACAGICCOC	TCACCAC	CAGC	TEETTACTCAAA		AAG	ACTOGAGAGI	GCCTGCCTGCA		GAAT-C
29. 100004443969.1_7	CTGTTAGAAGAA	ACAGICCOC	TCCCTA	CAGC		COTT	AAG				GAATCC
21 NIW004443969.1_11			GTTACACC	CACCACA	CTCTTCCCTATA	CGTTCTCTATCAGGG	AAAG	TGTTCAAATT	TTCTCCCATTCA		GAAGC
32 NW004443969.1_12		ACAGTAGAC	ACAGGAT		ACTTTTCTTACA			AATGTCAAGT	TTCTCCCTTCC	GTTCTT	I GAAGCC
33 NW004443969.1_13			ACACGAT	GGGAA	ACCTTTCTTCC	CATTCTCTGAATGGGA		AATATTAAGT			GAAGCC
34 NW004443969 1 15	TGATAAAAC	ACTGTTGCC	ACACGTT		ACACTTCTTACA	CATTCTCTGAATGGGA	AAAG	AATGTCAAGC	TTCTCCCTTTCC		GAAGCC
35. NW004443980.1 1	- GAT AAAAA	ACAGTCCTT	ACACCAT-	CTACA	AGATCTCTCAAA	CATTCTTTCAATGACA	- A A G	ACTGGAGAGT	TTCTCCCTTTCAC	GTACTT	GAATCC
36. NW004443980.1 2	- TGATAAAAC	ACAGTCCTC	ACAGCAT	CTACA	AGCTCTCTCAAA	CATTCTCTCAATGACA	AAG	ACAGGTGAAT	TTCTCCCTTTCA	GTATT	GAATCC
37. NW004443980.1 3	- TGATAAAAG	ATAGACCTC	AAACCAT	CTACA	AGCTCTCTAAA	CATTC - CTCAATGGGA	- TAAG	ACTGGTGAGT	TGCTCCCTTTCA/	ATGCTT	CAATC
38. NW004443980.1 4	- TGGTAGAAGAA	ACAGTCCAC	ACAGCAC -	- AGC -	TGACTTCTCAAA	CGTTCTCTCAGTGGGA	- AAG.	ACTGGAGAGT	GCCTGCCTTGCA	-GTATTT	GAATCC
39. NW004444509.1_1	- TGTTA CC AGAA	ACAGTCCGC	GCACCAC-	- AAGC -	TGCTTTCTCAAA	AGTTCTCTCAGAGGGA	GAAG	ACTGGAGAGT	GCGGGCCTAGCA/	- GTATTT	[GAATC]
40. NW004444309.1_19	- TGATAAAAT	- ACAGTCCTC	ACACCAT	CTATA	AGCTTTCTCAAA	CATTCTGTCAATGGGA	- A A G	A C T G G 🕇 G A G T	CTCTGCCT-GCA/	GTACCT	GAATCC
41. NW004444509.1_2	- T G G T A G A A G A A	ACACTCTGC	CCACCAC -	- CAGT -	TGCTTTCTCAAA	- GATCTCTCAGTGGGA	- GAAG	ACTGGAGAGT	GCCTGCCTTGCT	GTATTT	GAATCC
42. NW004444509.1_3	- T G T T <mark>G</mark> G A A G A	AGAGCCCC	CCACCAC-	AAGC	TGCTTTCTCAAA	GTTCTCTCAGTGGGA	GAAG	G C A G T A G A G T	CAAGGCCTAGCAA	A - GAATTT	GAATCC
43. NW004444509.1_4	- T G G T A G A A G A A	ACACTCCGC	CCACCAC-	CAGC -	TGCTTTCTCAAA	GGTTCTCTCAGTGGAA	GGAG	ACTGGAGAGT	GCCTGCCTTGCC	A-GTATTT	GAATCC
44. NW004444509.1_5	– T G <mark>G</mark> T A G A A G A A	AACACACCGC	CCACTAC-	C A G C -	TGCTTTCTCAAA	GGTTTTTCTAAGTGGAA	GAG	CTTCTAGAGT	GCCTGCCTTGCC	- GTATTT	GAATCC
45. NW004444509.1_6	- T G G T A G A A G A A	AACAGTCCTG	ACACCAC-	CAGC -	TACTTTCTGAAA	CATTCTCTCAATAGGA	TAAA	ACTGACGAGT	GTCTGCCTTCCA/	- GTACTT	GAATC
46. NW004444509.1_7	TGAAAAAA	GGAGTCCT	ACTTGGT	GAAA	ACCTCTCTTACA	CGTTTTCTGAATGGAA	- AAC	AATGTCAGGC	TTCTCCCTTTCA/	GTACTT	
47. NW004444509.1_8	TGAAAAAA	ACAGTICCI	ACATGGT	GAAA	ACCTTTCTTACA	CATTGTCTGAATGGGA	AAAC	AATGTCAAGC	TTCTCTTTTCA/	- GTACTT	GAAGCC
48. NW004444509.1_9	- T G A A A A A		ACATGGT	GAAA	ACCTTTCTTACA	CGTAGTCTGAATGGGA	AAAC	AATGTCAAAC		A-GTACTT	GAGGCC
49. NW004444309.1_1	- T G T T A G A A G A A	AAGAGTCTGC	CCACC-T-	CAAC -	TGCTTTCTCAAA	T G C TCTCTCAGTGGGA	GAAG	GCAGTAGAGT	GAGGCCTGCCA/	- GAATTT	GAAT-C
50. NW004444309.1_2	- T G T T A G A A G A A	AAGIGTCCGC	CCALCAC-	AAGA	TGCCTTCTCAA	GGTTCTCTCAGTGGGA	GAAG	ACTGGAGAGT	GCCTGCCTTGCA	A-GTATTT	CAACCC
51. NW004444309.1_3	- T G T G A G A G G A A	AACAATCAGC	CCACCAC-	CAGC -	TGCTTTCTCCAA	GGTTCTCTCAGTGGGA	GAAG	ACTGGAGAGT	GCCTGCCTTGCA	A-GTATTT	CAATCC
52. NW004444309.1_4	- T G T G A G A G G A G	ACAATCIGC	CCACCAC-	CAGC -	TGCTTTCTCAAA	GGTTCTCTCAGTG					
53. NW004444309.1_5	- TCTTAGAAGAA	AACAGTCCGC	CCAGGAG	AAGC	TGCTCTCAAA	GTTGTCTCAGTGGGA	GAAG	ACTGGAGAAT	GCCTGCCGTGCA	A-GTATTT	CAATCC
54. NW004444309.1_6	- IGIIAGAGGAA	AACAGICCGC	CCACCAC-	CAGC -	IGCITICICAAA	GGIICICICCGIGGGA	GAAG	ACIGGAGAGI	GCCTGCCTTGCA/	GIAIII	GAAICC
55. NW004444309.1_7	- TGTTAGAAGAA	AACCTTCCCC	CCACCAG-	AAGC	TGCTTTCTCAAA	G TTGTCTCAGTGGGA	GAAG	CCTGGAGAGT	GCCTGCCTTGCA/	- GCTTTC	GAATCC
56. NW004444309.1_8	- T G T T A G A A G A A	AACAGTCAGC	CCAGCAG	- AAGC -	TGCTTTCTCAAA	GGTTCTCTCAATGGGA	GAAG	CTGGAGAG	GCCTGCCTTGCA	A-GTATTT	CAATCC
57. NW004444309.1_9	- TGTTAGAAGAA	AACALTCCGC	ACACC	AAGC	TGCTTTCTCAAA	GGTTCTCTCATGGGA	AAAG	ACTGGAGAG	GCTTGCA	A-GTATTT	GAATAC
58. NW004444309.1_10	TGTTAGAAGAA	AAGACICAGA	CCAGC	GAGC	TGCTTTGTCAAA	GGTCTCTCAGTGGGA	GAAG	ACTGGAGAGT	GCCTGCCTTGCA	GTATTT	GAATCC
59. NW004444309.1_11	- TGTTAGAAGAA	AACAGTCCGA	CCAGCAC -	GAGG	TGCTTTCTAAAA	CATTGTCTCAGTGGGA	GAAG	ACTGGAGAGT	GCTTGCCTTGCA/	- GTATTT	GAATCC
60. NW004444309.1_12	- TGGTAGAAGAA	ACAGTCCGC	ACACAAC-	CAGC -	TGCTTTCTCA - A	CCTTCTCTCCTCGGGA	AAG	A - GGGAGAGT	A CTGCTTTGCA/	GTACTT	GAATCC
61. NW004444309.1_13	- GGTAGAAGAA	A - AATCCGC	ACGCCAC-	CAGC-	IGCTTTCTCAAA	AGIGCTCTCAGTGGGA	A G	ACTGGAGAGT	G CTGC TTGCA/	GAACTT	GAATCC
62. NW004444309.1_14	- IGGTAGAACAA	AACIGTCIC	ACACCAC-	GC -	TGCTCTCAAA	CATTCTCTCAATAGGA	AAG	ACTAGCAAGT	GCAGCCTTCA	- GTACTT	ACTCC
63. NW004444309.1_15		AAGAGTCC	ACACCAC-	AAGC	IGC-IICCCAAA	CATTCCCTCAATGCGA	AA	ACTGGTAAGC	GACIGCCTTCA	- GIICIT	GAAACC
64. NW004444309.1_16		AAAGICC	ALALCAC-	CAGC-	TGCTTTCTCAAA	CALLAIGICAAIGGGA	AAG	CIGGCGAGT	GECTECT CA	GIACIT	GAAI
66 NW004444309.1_17		ACAGICC	ACACCAC-	CAGC	AGCTTTCTCAAA		AAG	ACTOCCOACT	CTCTCCCTTTCA/	GTACT	GAATCC
THE INVERTIGATION AND A				ALA	MAN THE THE AGE		e4 1 1	Ch. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1. 1.			

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

	220	230	240	250	260	270	280	290	300	310
Consensus	CCTG	GTTTTCAAA	GGGAACTTTGCA-	TAACAGGGAA	TGAAGGTAC	TTTTGG	CATTC	A A A G C A A A A C A G -	- GCAGATTG	- TTTGCCT
Identity										
1 NW004444102 1 1	ССТО		GAGAACTTTCTA-	TAACAGGGAA	AGAAGGTAG	TTTTGG	C T TTC	AAAGCAAAACAG-	- GCAGATTG	TTTGCCT
2. NW004444102.1 2	CCTG	GTTTTCAAA	GCGAACTTTGCA-	TACCTGTGAA	TGAAGGTAC	TACTGG	CTTTC	AAAGTCAAACAG -	- GCAGATCT	TCTCTCT
3. NW004444102.1 3	CCTG	- GTTTTCAAA	GGGAACTTTGCA-	TTACAG-GAA	TGAAGGTAT	TGCTAG	CTTAG	ATAGCCAAACAG-	GCAGATTG	TCTGCCT
4. NW004444102.1_4	ССТС	- ATTTTCAAA	GGGAACTTTACA-	TACCAGGGAA	TGAAGGTTC	TAATGG	C T TTC	AAAGC CAAACAG -	GCAGATTG	TCTGCCT
5. NW004444102.1_5	CCTG	- CTTTTCAAA	GGGAACTTTGCA	TAACAGGGAA	TCAAGATAC	TTTTGG	CATTC	ACAGCAAAACAG -	- GCAGATTG	CTTGCCT
6. NW04444102.1_6	CTTA	- CTTATCAAA	GGGAACTTTGCA-	TAACAGGGAA	TCAAGCTAC	TTTTGG	CATTC	A A A G A A A T A C A A -	GCAGATTA	TTTGCCT
7. NW004444102.1_7	CCTG		GGGAACTTTGCA-	TAACAGGGAA	TGAAGGTAC			AAAGCAAAACAG -	GCAGATGG	TETCCCT
9 NW0044440733.1	CCTG		GGGAACTTTGCA-	TTCCAGGGAA	TGAAGGTAC				GCAGATTG	TTTCCCT
10. NW004444141.1 2	ССТС	GATTTCAAA	GGGAACTTTCCA-	TTCCAGGGAA	TGAAGGTAT	TTCTGT	CATCC	TAAGCCAAACAG -	GCAGATTG	TTTGCCT
11. NW004444141.1 3	CGTG	GTTTTCAAA	GGGATCTTTGCA-	TTACAGGGGA	TGAAATTAC	CTTTGC	CATTG	AAGCAGAACGG-	- TTAGATTG	TTTGCCC
12. NW004444141.1_4	CCTG	G T T T T <mark>G</mark> A A A	GGGAATTTGCG	TTACAGGGAA	TGAAGGAAC	TTTTCC	C T T T C	AAGAAAACAG -	CAGATTG	TTTGCCT
13. NW004444141.1_5	CCTC	GTTTTCAAA	GGGAACTTTGCA	TACAGGAA	TGAAGTTAA	CTTTCG	CATTC	AAGGAAAACAG-	GCAGATTG	ТСТ
14. NW004444141.1_6	CCCGG	GTTTTCCAA	G-GAACTTTGCA-	TACAGGGAA	TGACTGTTC	CTTTCG	CATTC	AGGAAGAAAAAG	GCAGATT	GTTCCCCT
15. NW004444141.1_/	CCTC	GTTTTCAAA	GGGAGCTTTCCC-	TACCOGGAA	TGALGGTAL	CACTCC		AGGAAGAACAAG	GCAGATAA	TOTOTOT
17 NW004444013.1_1	CCTG	GTTTTTAAA	GCAACTATGCA-	TAGCAGGGAA	TGAAGGTAC	TACTGG			GCAGATTO	TETGCCT
18 NW004444013.1_3	CCTG	GTTTTCAAA	GGAAACTTTGCA-	TAACAGGGAA	TAAAGGTAC	TACTIG	GTTTC	AAAGCAAAACAG-	- GCAGTTG	TTTGCTT
19. NW004444013.1 4	CCTG	GTTTTCAAA	GGGAACTTTGCA-	TAACAGGGAA	TGAAGGGTC	TACAGG	CTTTC	AAAGCAAAACAG-	GCAGATTG	TTTGCCT
20. NW004444013.1_5	ССТС	G T T T T C A A A	GGGAAC <mark>G</mark> TTGCC	TAACAGGGAA	TGAAGGTAA	TTTTGG	CATTC	AAAGCAAAACAG -	ACAGATTG	- TTTGC <mark>G</mark> T
21. NW004444013.1_6	CCTG	GTTTTCAAA	GGGAACTTTGC	TTACAGGGAA	AG-GGGTAA	CATTCG	CATTC	CAAGCAAAACAG-	GCAGATTG	TTTGCTT
22. NW004444013.1_7	CCTG	GTTTTCACA	GGGAACTTTGC	TAAAGGGAA	TGAAGGTGA	CATTCG	CATT	A A G A A A A A A G G -	GCAGATTG	TTTGCCT
23. NW004443969.1_1	CTG	- GITAICAAA	GGGGAACTTTGCA-	TAAGAGGGAA	ACAAGGTAC	TTTTCC	CATTC	AAAGCCAAACAG -	CAGATTA	TTTCCCT
24. NW004443969.1_2 25. NW004443969.1_3		GTTTTCAAA	GAGAACTTTGTA -	TAAGAGGGAA	TGAAGGTAG	TTTTGC	CATTC			TTCCCCT
26. NW004443969.1 4	CCTG	GTTTTCAAA	GGAACTTTGCA-	TACCAGGGAA	TGAAGG		CTTTC	AAAGCCAAACAG -	- GCAGATTG	TCTGCCT
27. NW004443969.1 5	ČČTĞ	ATTTTTGAA	GGGAACTTTGCA-	AAACAGGAAA	TGAAAGTAC	ATTTGG		AAAGCAAAACAG -	ATTG	TTTGCCT
28. NW004443969.1_6	ССТС	GTTTTCAAA	GGGAACTTTGCA	TAACAGACAA	TGAAATTAC	CTTTGG	CATTC	AAAGCAAAACAG -	GCACATTG	- ТТТ Б 🔳 С Т
29. NW004443969.1_7	CCTG	GTTTTCAAA	GGGAACTTTGCC	TAACAGGGAA	TGAAGGTAC	CTTTGG	CATTC	AAAGCAAAACAG-	GCAGATTG	TTTGCCT
30. NW004443969.1_11	ATATGGGGGTATAG	GIGTGITGAAC	CCGTACTCTGTGC	IGICIGIIIC	TGAALGTGC	CTTACAGAC	GAALCITTG	AAGGAGAAAACAA	GCTTCTT-	TICCCAT
31. NW004443969.1_12			AGAAACTCTGTG	TTACTGTGAC	TGAACGTGC	CTTTCT	AGAALCICIG		GCAGGTTG	
33 NW004443969.1_13	CCCGGGCTTTCAG	ATGTTTTCTAT	GGGAGCTCTGGT	TTACTGGGAA	AGAATGTGC	CTTTCA	CATTC		- GCTGATTG	TTCCCAT
34. NW004443969.1 15	CCCAGGCTTTCAG	ATGTTTTTAT	GGGATCTCTGGT -	TTACCAGTGA	TGAATGTGC	CTTTCA	ATTC	AGGAAAAACAAG	GCTGATTG	TTCCCAT
35. NW004443980.1_1	GCCG	GTTTTCAAA	TGGAACATTGCA-	CTACAGAGAA	TGACGGTTC	CTTTCA	CATTC	TAGGAAGAAAAAG	GGCAGATTG	GTTCCCTT
36. NW004443980.1_2	CCTG	GTTTTCAAA	■GGAACTTT■CA-	TTACAGAGAA	GAAGGTGT	TTT CT G	CATTC	TAGAAATAAAAAG	GGCCGATTG	CTGCCCC
37. NW004443980.1_3	CCCG	CTTTTCAAA	GGGAACTTTGCA-	TTATATGAAC	TGAAGGTAC	TTTTCA	CATTC	A A A G A A A A C A A G -	GCAGATTG	GTTCATCT
38. NW004443980.1_4			GGGAACTTGGA-	TAACAGGGAA	TGAADATAC				ACAGATIG	TTTGCCT
40 NW004444309 1 19	CTTG	GTTTTCTAA	GGGAACATTGC	TTACAGGGAA	TGAAGGTAT	TTTTCG	CAGTA		GATTG	TTCATCT
41. NW004444509.1 2	ČĊŤĞ	GTTTTCAAA	GGGAACTTTGCA-	TTACAGGGAA	TGAAGGTAC	TTTTGG	CATTC	AAAGC AAACAG -	- GTACATTG	TTTCCCT
42. NW004444509.1_3	C C T G	GTTTTCAAA	GGGAACTTTGCA-	TAACAGGGAA	TGAAGGTAC	TCTTGG	CATTC	AAAGCAAAACAG -	GCACATT	TTTCCCT
43. NW004444509.1_4	CCTG	GTTTTCAAA	GGGAAAGTTCAA-	TTACAGGGAA	TGAAGGTAC	TCTTGG	CATTC	AAAGCAAAACAG-	GCACATTG	TTTCCCT
44. NW004444509.1_5	CCIG	GTTTTCAAA	GGGAACGTAGCA-	TAACAGGGAA	TGAAGGTAC	ITTTGG	CATIC	AAAGAAAAACAA-	GCAGATIG	
45. NW004444509.1_6	CCTGTGTGTTTTCAC	GTTTTCAAA	CTAAATTTTACC		TGAAGGAAC				GCAGATIG	
47. NW004444509.1 8	ACCOGGTTTTCAC	ATGTTTTCTGT	CTAAATTTTAGG-	TTACTGGCCA	TGAATTGC		AATTC	TAGAAAGAACAAA	GCAGTTTG	TTCTCCT
48. NW004444509.1 9	CCTGGGTTTTCAC	ATGTTTTCTA-	AATTTTAGG	TTACCGGCAA	TGAATGTGC	CTTTCA	AATTC	AAGAAGAACAAG	- GCAGTTTG	TTCCCTT
49. NW004444309.1_1	ССТС	- G T T T T G A A A	GGGAACTTTGCA-	TAACAG-GAA	TGAAGGTA	TGCTAG	C T TAC	A A C G C C A A A C C G -	GCAGATAG	TTTGCCT
50. NW004444309.1_2	CCTG	GTTTTCAAA	GGGAACTTTGTC	TAACAGACAC	TGAAGGTAC	TACTGG	CTTTC	AAAACAAAACAG -	GCAGATTG	TTTGCCT
51. NW004444309.1_3	ССТG	GTTTTCAAA	GGGAACATTACA-	TGACAG-GAA	TGACGGTAT	TACTAG	CTTAC	A A A G C C A A A G A G -	- GCAGATTG	TTTGCCT
52. NW004444309.1_4	CCTG	GTTTTCAAA	GGGAACTTTGCA-	TAACAGACAC	TGAAGGTAC		ATTTC	AAAGCAAAACAG -	GCAGATIG	TTTGCCT
54 NW004444309.1_5	ССТС	GTTTTCAAA	GGGAACTTTGCA-	TAACAGGGAA	TGAGGGTAT			AAAGCAAAACAG -	- GCAGATTG	TTTTCCT
55. NW004444309.1 7	CCTG	GTTTTCAAA	GGAAAGTTTGCA-	TAACAGAGAA	TGAAGGTAC	TACTGG	C T TTC	AGAGCAAAACAG -	GCAGATTG	TTTGCCT
56. NW004444309.1_8	C T G	GTTTTCAAA	GGGAAGTTTTCA-	TAACAGGGAA	TGAAGGTAC	TTCTGG	CTTTC	AAAGCAAAACAA	GCAGATTG	TTTGCCT
57. NW004444309.1_9	CCTG	- GTTTT G AAA	GGGAACTTTGCA-	TAACAGGGAA	TGAAAGTAC	TTCTGG	C T TTC	AAAGCAAAACAG-	GCAAATTG	TTTGCCT
58. NW004444309.1_10	CCTG	- GTTTTCAAA	GGGAATTTGCA-	TACAGGGAA	TGAATTAC	CITTAG	CATTC	AAAGCAAAACAG-	GCAGA	TTTGCCT
59. NW004444309.1_11	CCTC	GTTTTCAAA	GGGAACTTCGCA-	TAACAGGGAA	I GAAGGTAC	ATTTCC	CATTC	AAAGCCAAACAG-	GCAGCITG	TTTCCCT
61 NW004444309.1_12	ССТС		GGGAACTTTGCA-	TAACAGGAAA				GAAGGGGAAACAG -	- GCAGATIG	ATTGCAT
62. NW004444309 1 14	ССТС	GTTTTCAAA	GGGAACTTTGCA-	TTACAGGGAA	TGAAGGTAC	CTTTCG	CATTC		- GCAGATTG	CTTCCCT
63. NW004444309.1 15	ČČTĞ	GTTTTCAAA	GGGAACTTTGA-	TTACAGGGGA	TGAAGGTAC	CTTTCA	CATT	AACAG-	- GCAGATTG	ATTGCCT
64. NW004444309.1_16	CCTG	GTTTTGAAA	GGGAAGATTGCA-	TTACAGGGAA	TGA TAC	CTTTCA	CATTC	TATGAAAAACAG -	GCTGTTTG	TTCACCT
65. NW004444309.1_17	CCTC	- GTTTCAAA	GGGAACTTTGCG	TACAGGGAA	TGAAGGTAC	ATTTIG	CATTC	AAGAGAAACAG	GCAGATTG	TTTGCCT
66. NW004444309.1_18	GAG	GTTTTCAAA	GGGAAATTTTCA-	TACAGTTAA	TGAA-GTAA	CTTTCT	CATTC	A G A A G A A C A A -	GCAGATTA	TTTCCCT

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

	320	330	340	350	360	370	380	390	400	410	420
Consensus Identity			AGGTTCAGTT	TTGCTGACTA.				- TGTAGCTT		AGAAAGAGAAA	
1. NW004444102.1 1	TCAGCGAGAGGA	TGGCAGAAG	AGGTTCAGTT	TTGCTGACTA	AACTTTGTG		GGAGCA	- TGTAGCTT	- CTTCCCTAT	AGAAAGAGAA	AGCAA
2. NW004444102.1_2	TCAGCTAGAGGA	CAGCAGAAG	AGGTTCAGTT	TGGCTGACTA	AACCTTGTGI	TGCTTAACA	GGAGCA	- TGTAGCTT	- CTTC - CTGC	GAAATAGAAA	AGTAA
3. NW004444102.1_3	TCAGCTAGAGA	CAGCAGAAG	AGGTTCAGTT	TTGCTGACTA	AACCTTGTGT	TGCTTAACA	CGAGCA	- TGTACCTT	- CTTCCCAGCO	GAAAGAGAA	A-CAA
4. NW004444102.1_4	I CAGCIAGAGGA	CAGCAGAAG	AGGIACAGII	TGGCTGACTA,	AACCIIGIGI	IGCITAACA	GGAGCA	- IGIAGCII		GAAAGAGAA	AGCAA
5. NW004444102.1_5 6. NW04444102.1_6	TCAGCTAGAGGA	TGGCAGAAG	AGGTTCAGTT	TTGCTGACTA	AAGCTTGTGT	IGCTTAACG	GGGCA	TGTAGC	- CTTCCCTGC	AGAAAGAAGAAGAA	AGCAA
7. NW004444102.1 7	TCAGCTAGAGGA	TGGCAGAAG	T G T TTCAGTT	TTGCTGACTG	AAGCTTGTGT	TGCTTAACG	CGAGCA	- TGTAGCTT	- CTTCCCTGC	AGAAAGAGGA	AG-AA
8. NW004448733.1	TCAGCTAGAGGA	CA GCAGAAG	AGGTTCAGTT	TGGCTGACTA	AACCTTGTG	TGCTTCACA	GGAGCA	- TGTAGCTT	CTTCCCTGC	G A C A G A G A A /	AGC <mark>G</mark> A
9. NW004444141.1_1	TCAGCGAGAGGA	ACAGCAGAAG	AGGTTCAGTT	TTGCTGACTA.	AACCTTGTGT	T C T T A A C A	CGAGCA	- TGTAGGTT	- CCTCCCTGC	GAAAGAGAAA	AGCAA
10. NW004444141.1_2 11. NW004444141.1_3	TCAGCTAGAGGA	A G C A G A A G	AGGIICAGII	TTECTGACTA		I GCTTAACA	CGAGCA		- CTTCCCTGCC		AGCAA
12. NW004444141.1 4	TCAGCTAGAGGA	TGGCAGACG	CGGTTCAGTT	TTGCAAACTA	AAGCTTGTGT	TGGTTGACA	CGAGCA	- TGTATCCT	ATTCACTG	AGGAAGAGTA	ATGAA
13. NW004444141.1_5	TCAGCTAGACA	GACAGAAG	CAGTTCAGTT	TTGCTGACTA	AAGCTTG-GI	TGGTTAACA	CGAGTA	- TGTATCCT	ATTCCCTGC	AGAAAGAGGA	AGAAA
14. NW004444141.1_6	TCAGCTAAAGAA	GGCAGAAG	ATTTCAGTT	TTGCTGACGA.	AAACTGGTGI	TGTAACCA	CGTGCG	- TGTTGCCT	ATTCACTGC	AGAAAG <mark>G</mark> GAA	CCAA
15. NW004444141.1_7	TCACCTACAGGA	ACGACAGAAG	GGIICAGII	TIGCAGACIA,	AACHIIGI	I GIAACCA	GCAGCA		ATTCACTGC	AGAAAGAGAAA	ACCCA
17 NW004444013.1_1		TAGCAGAAG	AGGTTCAGTT	TTGCTAATTA	AACCTAGTGI	TGCTTCACA	GGAGCA	- TGTAGCTT	- CTTCCCTGC		AGCAA
18. NW004444013.1_3	TCAGCCAGAGG	GGGCAGAAG	AGGTTCAGTT	TTGCTGACTA	AACCTTGTGT	TGCTTAAGG	CGAGCA	- TGTAGCTT	- CTTCTCCTGC	GAAAGAGAAA	AGCAA
19. NW004444013.1_4	TCATCCAGAGG	GGCAGAAG	AGATTCAGTT	TTG <mark>G</mark> TGACTA,	AACCTTGTGT	TGCTTAACA	CGAGCA	- TGTAGCTT	- CTTCTCCCC	GAAAGAGAA	AGCAA
20. NW004444013.1_5	TCAGCTAGAGGA	GGGCAGAAG	AGTTCAGTT	TTGCTGACTA,	AACTTGTGI	TGCTTAACA	CGAGIG	- TGTAGCTT	- CTCCAGTGC	AGAA <mark>G</mark> GAGAA/	AGCAA
21. NW004444013.1_6			CGTTCAGTT	TTGCTGACTA			CAAGCA			AGAAAGAGAAA	ACCCA
23 NW004443969 1 1	TCAGCTAGAAGA	TGGCAGAAG	AGGTTCAGTT	TTGCTGACTA	AACTTTGTG	TGTGTGTAACG	CGAGCA	- TGAAGCTT	- CTTCCCTGC	AGAAAGAGAAA	AGCAA
24. NW004443969.1 2	TCAGCTAGAAGA	TGGCAGAAG	AGGTTCTGTT	TTGCTGACTA	AACCTTGTGT	TGTGTAACG	CGAGCA	- TGTAGCTT	- CTTCCCTGC	AGAAAGAGAAA	AGCAA
25. NW004443969.1_3	TCAGCTAGAGGA	TGGCAGAAG	AGGTTCAGTT	TTGCTGACTA	A A C C T T T T G T	TGCTTAATG	GAGCA	- TGTAGCTT	- CTTCTCCCC	CAGAAAGAGAAA	AGCAA
26. NW004443969.1_4	TCAGCTAGAGGA	CAGCAGAAG	AAGTTCAGTT	TTGCTGACTA	AACC-TGTGT	TGCTTAATG	GAGCA	- TATAGCTT	- CTTGCCTGG	AGAAAGAGAA	AGCAA
27. NW004443969.1_5	TCAGCTAAAGGA	A T G G C A G A A G	AGGIICAGII	TTECTGACTA			GAGCO				AGCAA
29. NW004443969.1_0	TCAGCTAGAGGA	TGGAAGAAA	AGGTTCAGTT	TTGCTGAGTA	AAGCTTGTGTG	IGCTTAAC	GAGCC	-TGTAGCCT	- CTTCCCTGC	AGAAAGAGAGAA	AGCAA
30. NW004443969.1_11	TCAAGTACAGGA	GACAATAG	AGATGTAGTT	TTGCTAAGCA	AACACTATGI	TACCAGAA	CCTGCATA	TGTTCACA	ATTCACTGC	TTAAAGAGAAA	AGTAA
31. NW004443969.1_12	TCAAGTACAGGA	GGCAATAG	AGATGTAGTT	TCCCTAAGTA.	AACACTGTGT	TACCAGAA	CATGCA	TGTTCCCA	ATTAACTGC	GAAAGAGAAA	AGCAA
32. NW004443969.1_13	TCTACTACAAGA	ATGGCAAAAG	TCTTGTAGTT	TTGTTGAGTA	ACCTTCTG	TGTAAGCA	CGIGCA	- TGTTTCCA	GTTTACTGC	CAAAAAAAAAAAAAA	AGCAA
33. NW004443969.1_14 34. NW004443969.1_15			CTAT CTCCTT	TTGCTGACTA	AGCCTTGTGTG	TTTTATCCA	CATGCA		ATTCALTCC		AGCAA
35. NW004443980.1 1	TCAGCTAAAGAG	CAGCAGAAG	CTGTTCAATT	TTGCTGACTA	ACCTTGTGI	TGCAACCA	CGTTTA	- TCTTACTT	-ATTCAGTG-G	AAAAGGCAA	TCAA
36. NW004443980.1_2	TTAGCTAAAGAC	GCCGCAAAAG	GGTTCAGAT	TTGCTGACTA	AACCTTGGG	TGTAACCA	GGGCA	- TGTTCCCT	TTCAATGC	AGAAAGAGAA	TCCAA
37. NW004443980.1_3	TCAGATTAAAGA	CAACAGAAG	CGGTTCAGTT	TTGCTGATTA	AACCGTATGI	TGTAACCA	GGGCA	TATTCCT	ATGCACTGA	AGAAAGAGAA	ACCTA
38. NW004443980.1_4	ACAGGIAGAGGA	ATACCAGAAG	AGGIICAGII	TIGCIGACIA	AACCIIGIGI	IGCITAACG	CAAGCA			AGAAAGAGGAA	AGCAL
40 NW004444309.1_1	TCAGCTAAAGCA	TGGCAGAAG	GGGGTCAGTT	TTTCTGACTA	AACCTTGTGT	TGTAACCA	CGTGCA	TGTGTTCT	- ATTCACTGG		ACCAA
41. NW004444509.1_2	TCTGCTAGAGGA	CAGCAGAAG	AAGTTCAGTT	TTAATGAATA	AAATATGTGT	TGGTTAACG	GGAGCA	- TGTAGTTT	- CTTCCATAT	AGAAACAGAAA	AGCTA
42. NW004444509.1_3	TCTGCTAGAGGA	ATGGAAGAAG	CT GTT T AGTT	TTCCTGA-TG	AATTTTCAGI	TGCTTAACA	CAAGCA	- T G <mark>G</mark> A G T T T	CTTCCATAT	AGAAAGAGAAA	AGCGA
43. NW004444509.1_4	TCTGCTAGATGA	ATGGCAGAAG	AGTTCAGGC	GTGCTGAATA,	AAGTTTGTGT	TGCTTAACG	CGAGTA	- TGTAGCTT	- CTTCCATGI	AGAGAGAAA	AGCGG
44. NW004444509.1_5 45. NW004444509.1_6			CAGTTCAGTT	TTACTCACTC	AAGCTTGGG	I GCTTAACG	GAGCA				AGCAA
46. NW004444509.1 7	TGAACTACAGCA	AGGCAAAAA	GATGTAGTT	TTCCTGAATA	AACCTTGTG	GTG	CAAGCACT	GGTAACTTTGC	ACTTCTCTGA	CTAAAGAGAAA	ACCAA
47. NW004444509.1_8	TCAACTACAGCA	GGCATAAG	GGATATAGTT	TTGCTGAATA	AACCTTGTG	TGCA AGCA	CGCGTT	CGTTTGCA	- CTTCACTGC/	CTAAAGAGAAA	ACAAA
48. NW004444509.1_9	TCAACTACAGCA	TGGCAAAAG	GGATGTAGTT	TTGCTGAATA	AACCTTGT	GCAAGAA	CGGGCA	GTTCACA	- CTTCACTGC	GAAAGAGAAA	ACCAT
49. NW004444309.1_1		ACAGCAGAAG	AGGIICAGII	TAGCTAACTA	AACCIIGIG	IGCITAACA	CGAGCA				AGCAA
51 NW004444309.1_2	TCAGCTAGAGG	CAGCAGAAG	AGGTTAGTT	TTGCTGACTA	AACCTTGTGT	IGCTTAACO	CGAGCA	- TGTAGCTT	- CTTCCCTGC		AGCAA
52. NW004444309.1_4	TCAGCTAGAGGA	TGGCAGAAG	AGGTTCAGTT	TTGCTGACTA	ACCTTATG	TGGTTAACG	GAGCA	- TGTAGATT	- CTTCACGGG	AGAAAGAGAAA	AGTAA
53. NW004444309.1_5	TCAGCTAGAGGA	ATGGCAGAAG	ACATTCAT	TTGCTGACTA	CACCTTTTG	TG <mark>G</mark> TTAAC <mark>G</mark>	GAGCA	- TGTAG <mark>AA</mark> T	- CTTCACGGG	CAGAAAGAGAAA	AGCAA
54. NW004444309.1_6	TCAGCTAGAGGA	TGGCAGAAG	ACATTCA T	TTGCTGACTA		TGGTTAAAG	GAGCA	- TGTAGAAT	- CTTCACGGGG	AGAAAGAGAAA	AGCAA
55. NW004444309.1_7	TCAGCTAGAGGA	ATAGCAGAAG	AGATTCAGTA		AALAIIGIG		GGAGCA				AGCAA
57. NW004444309.1_8	TCAGCTAGAGGA	TGGCACAAG	AGGTTCAGTT	TGGCTGACTA	ATCCTTGTG	TGCTTCAAG	CGAGCA	- TGTAGCCT		AGAAGGAGAAA	AGCAA
58. NW004444309.1_10	TCAGCCAGAGG	TGGCAGAAG	AGGTTCAGTT	TTGCTGACTA	AACCTTGTGT	TGTTTAAGG	CGAGCA	GTACCT	-CGTCCCTGC	AGAAAGAGGA	AGCAA
59. NW004444309.1_11	TCAGCGAGAGGA	ATGACAGAAG	AGGTTCAGTT	TTGCTAACTA.	AACCTTGTGT	TGCTTAACG	CCAGCA	- TGTAGCTT	- CTTCCCCGA	AGAAGGAGAAA	AGCAA
60. NW004444309.1_12	TCAGCTAGA	TGGCGGAAG	AGGTTCAGTT	TTGATCTCTG	AACCTTGTGT	TGCTTAACG	AGAGCA	- TGTAGCCT	- CTTCCGTGCC	AGAAAGACAAA	AGCAA
61. NW004444309.1_13 62 NW004444309.1_14	TCCAGTAGAGGA	TGGTAGAAG	AGGTTCCGTT	TTGCALLA	AAGCTTGTA	TGCTTTACA	CGATCG				ACCAA
63. NW004444309.1 15	TCAGCTGGAGGA	GCGGAAG	CTGTTCAGTT	TTGCTGACTA	AAGCTTGTA	TGCTTAACA	GGAGCA		GTTCCCTGC	AGAAAGAGAAA	ACCAA
64. NW004444309.1_16	TCAGCTAGAGGA	CCGCAGAAG	CGGTTCAGTT	TTGCTGACTA	AAGCTTGTGT	TGCCTCACA	GAGGA	TGTATCAT	ATTCCCTGC	AGAAAGCGAAA	ACCAA
65. NW004444309.1_17	CCAGCTAGATGA	TGGCAGAAG	CAGTTCCGTT	TCCCTGAGTA.	AATCTTGTC	TGTCAGCA	CGTGAA	TGTTTCCT	ATTCACTGC	GAAATGAGAA	AACAA
66. NW004444309.1 18	TCAGCTATAGGA	ATGGCGAAAG	GGTTCAGTT	TIGCTGACTA	AACTTATG	GTAAGCA	CGGGCA		ATTCACTGC	AGAAAGGAAA	ACCAA

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

		430	440	450	460	470	480	490	500	510	520
Consensus	AAGCCAGT		TGCTCTCAGTCTGG	GGATGAAG	GAGAAGGA	TCACTTACTGTG	TTGAAAATCCC	- TCTGTG	- TTACATGC	CTAAGAA	AGTTGCCTGTGCA
Identity											
1. NW004444102.1 1	GAGCCTGT		TGCTTTCAGTCTGG	GGTT	TGAAAAGGAA	TCACTTATTGTGT	TTGAAAATCGT	TCTGTG	TTAAATGC	CTAAGGA	AGATGCCTGTGCA
2. NW004444102.1 2	GAGCCTGT		TGCTCTCCGTCTGG	GGATAAAG	GGAGAACGAT	TCACTTTCTGTGI	TTGAAAATCCC	TCTGTG	TTACATGO	CTAAGAA	AGTTGCCTGTGCA
3. NW004444102.1_3	GAGCTTGC		TGCTCTCAGTTTGG	GGATGAAG	GGAAAAAGAT	ТСССТТТСТБТБІ	TTTGAAAATCCC	TCAGTG	TTACATGC	CTAAGAA	AGTTGTCTGTGCA
4. NW004444102.1_4	GAGCTTGT		TGGTCTCAGTCTGG	GGATGAAG	GGAGAACGGT	TCACTTTCTGTGT	TTGAAAATCCC	TCTGTG	TTACTTGC	CTAAGAA	<pre>GTTGCCTGTGCA</pre>
5. NW004444102.1_5	AAGGCAGT	TGC	TGCTCTCAGTCTGG	G G G T T G A A G	TGAGAAGTTT	TCCCTTACCTTA	TTGAAAATCCC	TCTATG	TTACATGC	CTAGGAA	GTTGCCTGTGCA
6. NW04444102.1_6	AAGCCIGT	TCC	T-CTCTCAGTCTGG	GGATGAAG	TGCAAAGGAT	TCACCTACTGTGT	TTGAAAATCCA	TCTGTG	TTACATGO	ATAAGAA	GTTGCCTGTG-A
7. NW004444102.1_7	CACCTTCT	IGC	TECTETCAGICIGA	GGATGAAG	C C A C A A C C A T			TCTCTC		CTAAGAA	AGTTGCCTGTGCA
8. NW004448733.1	GAGCTTGT		TECTETCAGTETEE	GGATGAAG	GGAGAACGAT		TTAAAAATCCC	TCTGTG	TTCCATCC	CTAAGAA	AGTTGCCTGTGCA
10. NW004444141.1 2	GAGCTTGT	'	TGCTCTCAGTCTGG	GGATGAAG	GGAGAACGAT	TCACTTTCTGTGT	TTGAACATCCC	TCTGTG	TTACATGO	CTAAGAA	AGTTGCGTGTGCA
11. NW004444141.1 3	AAGGCAGT		TGCTGTCAGTCTGG	AGGTGAAG	TGAGAAGGAT	TCAATTACTGTGT	TTGAAAATCCC	TCTGAG	TTACATGO	CTAAGAA	AGTTGCCTGTGCA
12. NW004444141.1 4	AAAGT		TGCTCTCAGTTTGG	GGATGAAG	TGAGAAAG	TCACTTACTGTG	TTGAAAATCIC	TATGTG	TTTCATGC	CCAAAAAA	- TGGCCTGTGCA
13. NW004444141.1_5	AAGACAGT		TGCTCTCAGTCTGG	AAAATGAAG	TGAGAA - AAT	TCACTTACTGTG	TTGAAAATCTC	TCTGTG	TTAGATGC	CCAAGAA	AGTGGCCTGTGCA
14. NW004444141.1_6	AAGACAGT		TGCTCTCAGTCTGG	TACATGAAA	TGAGAATGAT	TTACTTACTGTG	TTTGAGAATCTC	TCTGT	TTTCATGC	CCAGGAA	GGTCGCCTATGCA
15. NW004444141.1_7	AATACAGT		TIGTCTCATTCTAG	GGGGAAA	TCAGAAGGAT	TAACTTAGTGTG	CTGAAAATCC	TCTGTG	TTACATGO	ACAAAAA	GTTGGCTCTACG
16. NW004444013.1_1	GAGCIGT		TGCTCTCAGACTGG	GGATGAAG	GGAGAACGAT		TIGAAAAICCC	TCTGTG	TTACATGO	CTAGAA	AGTIGCCIGIGCA
17. NW004444013.1_2	GAGCTTGT			GGATGAAG	GAGAGAAGGAT			TCTCTC	TTACATGO	CTAAGAG	AGTTGCCTGTGCA
19 NW004444013.1_3	GAGCTTGT		TECTCICAETCIES	GGATGAAG	GAGAAGGAT	TCACTTACTGTAT	TTGAAAATCCC	TATGTG		CTAAGAA	AGTTGCCTGTGTA
20 NW004444013.1_4	GAGCCAGT		TGCTCTCAGTCTGG	GGAAGAAG	TGAGAAGGAT	TCACTTACTTAG	TTGAAAATTC	- TCTGTG	TTACTTGC	CTAAGAA	AGTTGTCTGTGCA
21. NW004444013.1 6	ACGACAGT		TATCTCCCTCGGGG	GGATGAAG	TGAGTATGAT	TACTTACTGTGT	TTTGAAAATCGC	- CTGTG	TTACATCO	TCAAGAA	T GTTGCCTGTGCA
22. NW004444013.1_7	ATGACAGT		TGCTCTCCTCTGG	GGATGAAG	TGAGTAGGAT	AGACTTACTCTGI	TTTGAAAATCCC	- CCGTG	TTACATGO	TCAAGAA	■GTTGCCTGAGC■
23. NW004443969.1_1	GAGCCTGT		TTCTCTCTTTCTGG	GGATGAAG	AGAGAAGGAT	TCA <mark>GG</mark> TACTGTG1	TTGAAAATCCC	TCTGTG	TTACATGC	CAGAGAA	ATTTGCCTCTAAA
24. NW004443969.1_2	GAGCCAGT		TGCTCTCAGTCTGG	GATGAA	TGAGAAGGAT	TCACTTACTGT	TTTGAAAATCCC	TCTGTG	TTACATGC	CTAAGAA	AGTTGCCTGTGCA
25. NW004443969.1_3	AGCCAGT		TGCTCTCACTCTGG	GGATGAAG	TGAGAAGGAT	TCACTTACTGTG	TTGAAAATCCC	ATCTGTG	TTACTTGC	CTAAGAA	AGATGCCTGTGCA
26. NW004443969.1_4	GAGCCAGT		TACICICAGICIGG	GGATGAAG	TGAGAAGGAT	TCACTTACTGIGI	TTGAAAATCCC	TCTGTG		CTAAGAA	AGTIGUCIGIGUA
27. NW004443969.1_5	- AGCCAGT				TGAGAAGGAT		TTGAAAATCGC	TCTGTG		CTAAGAA	AGTTGCCTGGGCA
29 NW004443969.1_0	-AGCCAGT		TGCTCTCAGTCTGA	GGATGAAG	TGAGAGGGAT	TCACTTACTGTGT	TTGAGAATCGC	TCTGTG	TTACATGO	CCAGGAA.	
30. NW004443969.1 11	AATAGAGT		TTCCATCACTTGGC	GAGTAAC	TCAGAAGGAA	TAACTAACTGTGT	TTGAATACCCC	TCTGCG	TTACATGO	AAAAGAT	GATTGCATATACA
31. NW004443969.1 12	AATCCAGT		TTCTTTCACTTGGC	TGACTAAC	TCAGAAGGAT	TAACTTACTGTGT	TTGAAAACCCC	TCTGTG	TTACATAC	ACGAGAG	T GTTGCCTACACA
32. NW004443969.1_13	AATACAGT		TTCTCTCATTCTGC	TTGACTAAG	TCAGAAGGAT	Τ 🗛 Α С Τ Τ Α 👖 Τ G Τ G Τ	TTTGAAAACCA-	TCTGTG	TTACATGC	AAGAGAG	TGTTGCTTATACA
33. NW004443969.1_14	AACACC		TTCTCTCATTCTGC	TTGAGTAAC	TCAGAGGAAT	TAACTTACTGTGT	TTTGAAAACCCC	TCTATG	TTACATGO	ATGAGAG	<pre>GTTGCCTATACA</pre>
34. NW004443969.1_15	AATACTAT		TTCTTTCACAGTCA	GAGTAAG	TCATAAGGAT	TAAGTTACCGTGI	TTGAAAACCCC	TCTGTG	TTACATGO	ACAAGAA	GTTGCCTATAC
35. NW004443980.1_1	AAGACACT		IGCICICAGICIGA	GGAIG	- GAGAAGGAT		TIGAAAAICCC	TCTGTG		CCAGGAA	GGIIGACIAIGAA
36. NW004443980.1_2	AGGACAGT		TECTETCAGICIGG	GGATGAAG	TGAGAGAGAT			TCTGTG	TTACATGO	CCAGGAA	CATCATCIATGCA
37. NW004443980.1_3	AACCCAGT		TGCTGTCAGTGTGG	GGATGAAG	TGAGAAGGAG		TTGAAAATCCC	TCTGTG		CTAAGAA	AGTIGCCTGTACA
39. NW004444509.1 1	GATCTTGT		TGCTCTCAGGCTGG	GGATGAAG	GGAGAACGAT	TCACTTACTGTGT	TTGAAAATCCC	TIGGTG	TTACATGO	CTAAGAA	TGCTGCCTGTGCA
40. NW004444309.1_19	AAGACAGT		TGCTCTCAGTCTGG	GAATGAAG	TGAGAAGGAT	TCACTTACTGTGT	TTTGAAAGTGCC	TGTG	TTACATGO	CCAAGAA	TGTTGCCTATGCA
41. NW004444509.1_2	GAGCCAGT		T <mark>T</mark> CTCTC <mark>G</mark> GT <mark>G</mark> TGG	GGATGAAG	TGAAAAGGAT	TCACTTATTGAGI	TTGAAAATCCC	TCTGTG	T T A C 🔳 T G 🗛	CTAAGAA	AGTTGCCCGTGCA
42. NW004444509.1_3	GAGCCAGT		TTCTCTTGGTGTGG	GGATGAAG	TGAAAAGGAT	TCACTTATTGAGI	TTGAAAATCC	TCTGTG	TTAAATGC	CTAAGAA	AGTTGCCTGTGCA
43. NW004444509.1_4	AGCGT			GGATGAAG	TGAGAAGG	TCACTTACTGTG	GIGAGAATCCC	- TCTGTG	- TTACITGC	CTAAGAA	AGTTGCCTGTGCA
44. NW004444509.1_5	GAGECAGE				TGAAAGGAT	TAACTTACTGTGTG	TTGACAAICCC			CTAAGAA	AGTIGUCUGIGUA
45. NW004444509.1_0	AAGACAGT			TGAGAAAG	TCTGAAGGAT			TCTCTC	TACATGO	AGGGGAA	
47. NW004444509.1 8	AAGACAGT		TTCTCTCATTGTGC	GAGAAAA	TCTGAAGGAT	TACTTACTGTG	TTGAATACCCT	TCTGT	CTACATGO	AGGAGAA	CTAGTCTTTATG
48. NW004444509.1 9	AAGACAGT		TTCTCTCATTGTGC	TAAGGAAG	TCTGCAGGAT	TACTTACTGTG	TCGAACACCTG	TCTGTG	CTACATG	AGGAGAA	CCTTGCATTTCT
49. NW004444309.1_1	GAGCTTGC		TGCTCTCAGTCTGG	GGATGAAG	GGAGAACGAT	TCACTTTCTGTGT	TTGAAAATCCC	TCTTTG	TTACATGO	CTGCAAA	TGCTGCCTGTGGA
50. NW004444309.1_2	GAGCTTGT		TGCTCTCGGTCTGG	GAATGAA	TGTGAAGGAT	TCACTCTCAGTGI	TTGGAAAATCCC	TGTATG	TTCCATGC	CTAAGAA	AGTTGCCTGTGCA
51. NW004444309.1_3	GAGCTTGC		TGCTCTCAGTCTGG	GGATGAAG	GGAGAACGAT	TCACTTCCTGTG1	TTGAAAATCCC	TCTGTG	TTCCATGC	CTAAGAA	AGTTGCCTGTGCA
52. NW004444309.1_4	GAGCCIGT			GGATGAAG	TGAGAAGGAT	I C A G I C C A G I G I	TGGAAAATCCC	TCTATG	TTCCATGC	CTAAGAA	AGTIGCCIGIGCA
53. NW004444309.1_5	GAGCCTGT		TECTETCAGTETEE	GGATGAAG	TGAGAAGGAT		TGAAAATCCC	TCTATG	TTCCATGC	CTAAGAA	AGTTGCCTGTGCA
55 NW004444309.1_0	GAGCGTGC		TGCTCTCAGTCTGG	GGATGAAG	TGAGAAAGAT	TCACTTCCAATGI	TTGAAAATCCC	TCTGTG	TTTCATGO	CTAAGAA	AGTTCCCTGTGTA
56 NW004444309.1_7	GAGCTTTT		TGCTCCCGGTGTGG	GGATGGAG	GAT	TCACTTACTGTGT	TTGAAAATCTC	- TCTGAG	TTACATGO	CCAAGAA	AGTTGATTGTGTA
57. NW004444309.1 9	- AGCATTT		TGCTGTCTGTCCTT	GGAAGAAG	TGAGAAGGAT	TCACTAACTGTGT	TTTGAAAATCCC	TCTGTG	TTCCATGC	CTAAGAA	AGTTGCCTGTGCA
58. NW004444309.1_10	- ATCCTGT		G GCT <mark>G</mark> TCAGTCTGG	GGATGAAG	TGAGAAAGAT	TCAGTTACTGTGT	TTTCAAAATCCG	TCTGTG	TTCCATGC	CTAAGAA	AGTTGTCTCTGCA
59. NW004444309.1_11	A CAAGC		T G C T <mark>G</mark> T C A G T C T <mark>T</mark> G	GGAAGAAG	TGAGAAGGAT	TCACTTAGTGTGT	TTTGAAAATCCG	TCTGTA	TTCCATGC	CTAAGAA	AGTTGCCTGTGCA
60. NW004444309.1_12	GAGCCTGT		TGCTCTCAGTCTG	GGTGAAG	TGAGAAGGAT	TCACTCACTGTG	TTGAAAATCCC	TCTGTG	TTACATGO	CTGATAA	AGTTGCCTGTGCA
61. NW004444309.1_13	AAGACCGT		I G C T C T C A C T C T G G	GGGGGGGAAG	IGAGAAAGAT	GACTTACTGTGT		- ICTGTG	TACATAC	CTAAGAA	TT GCATTGCA
62. NW004444309.1_14	AACACAGT			ALGAAG	TGAGAAGGAT		TTGAAAATCCA	TCTCTC	TTACATCO	CAAGAG	AGTIGECIGIGEA
64 NW004444309.1_15	AAGACAGT			GGATGAAG	TGAGAAAGGAT	TCACTTACTOR	TTGAAAAICCG	TCTGTG	TTACATGO	CCARGAG	AATIGUUGAIGU
65 NW004444309 1 17	AAGATAGT		TGCCCCCAGTCCAG	GCATGAAC	TGAGAAGGAT	TCACTTACTGT	TTGGAAATCCC	TCTGTG	TTACTTGC	CCAGGAC	TGTTCCGTGTGCA
66. NW004444309.1_18	AATACAGT		TGCTCTCATCTCG	AGGACGAAT	TGAGAAGGAT	TCACTTACTGTGT	TTGAAAATCCC	TTTGG	TTACATG	TCAGGAA	GTTGCCTGTACA

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

	530	540	550	560	570		580	590	600		610	620	630
Consensus	TTGTCTAC	CTGGATAAAG	IGTTTGACTT	TTCTCCCTT	TCAAGTAGT	TGAAGCCA	ATTCAGGGT	TTACAGCTI	TCCG	- GGAGGCAT	TCTGT	GCTGCGCGG	GAGAA
Identity													
1 NIMODAAAAAOO 1 1	TTCTTTAC	ATCCATAAAC	TCTATCACTT	TTCACCCTT	TCAACTCCT	TEAACCO	ATCACACT	TTCCACCA	TCCC	CCACCCAT	TCCCT	CCTCCCCCCC	CTCACAA
2 NW004444102.1_1	TGGTCTAC	CTGGATAAAG	TGTATGACTT	ATC-CCCTT	TCAAGTAGT	TGAAGCC	TTCAGGGT	TTCCAGCAT	TTCC	GGAGGCAT	TCTGT		CTGAGAA
3 NW004444102.1_2	TTGTCTAC	CTEGATAAAG	GTTTGACTT	TTCTCCATT	TGAAGTAGT	TGAAGCC	TTCAGGGT	TTGCAGCAT	TCCG	GGAGGCAT	TCTGT	GCTGCGCGG	CTGAGAA
4 NW004444102.1_5	TTGTCTAC	CTCTGTAAAG	TGTATGACTC	TTCTGCGGT	TCAAATAGT	TGATACC	TTCAG	TTACAGAGT	TCCG	- GTAGGGAG	TGTCT	GCTGCCCAG	TGAGAA
5. NW004444102.1 5	TTGTCTAC	CTGGATAAAG	GTTTGATTT	TTCTCCTGT	TCGAGTAG	TGAAGCCA	ATGCGGGGT	TTAAAGCTT	TCCA	GGAGTCAT	TCTGT	GCTGTGCAG	GAGAA
6. NW04444102.1 6	CTGTGTAC	CTGGGTAAAG	GTTGGATTT	TTCTCCCGT	TTGAGTAGC	TAAAGCC	TGCTGGAT	TTACAGCTT	T-AG	GGAGGCAT	TCTAC	GTTGTTCAG	CGAGAA
7. NW004444102.1_7	TTTTCTAC	GTGAGTAAAG	GTTTAACTT	TTCTCCCAT	TAGAGTAGC	TGAAGCC	T G C G G A T T	TTACAGCTT	TCCG	GGAGGCAT	TCTGT	ACTGTGCCG	CTGAGAA
8. NW004448733.1	TTGTCTAC	CTGGATAAAG	TGTATGACTT	TTCCCCTT	TCAAGTCGT	TGAAGC	STTCAGGGT	TTCCAGCAT	TCCC	G G A G G G A T	TCTGT	GCTGCGCGG	C T G <mark>G</mark> G A A
9. NW004444141.1_1	TTGTCTAC	CTGGATAAAG	TGTATAACTT	TTCTCCCTT	TCAAGTAGT	TGAAGCC	STTCAGGGT	TTCCAGCAT	TCCG	GGAGGCAT	TGTGT	GCTGCGCG	CTGAGAA
10. NW004444141.1_2	TIGICIAC	CIGGATAAAG	GIAIGACII	TICICCCII	TCAAGTAGT	TGAAGCC		TTACAGCAI	TCCG	GGAGGC	TCTGT	GCTGCGCGG	- IGAGAA
11. NW004444141.1_3	TTTTTCTAC	CTCCCTCAAGG	JGITT-ACTI	TTCTCCCC		TAAAGCC	TCCCCCCT	TTACAGETT		GGAAGCAT	TCTCT		- CGAGAA
13 NW004444141.1_4	TTGCCTAC	CTCGGTAAAG	GTTTGACTT	TTCTCCCAT	TCGAGTATG	TGAAGCC	TGCGGGGT	TACGGCTT	TCCA	GGAGGCAG	TGTGT	GCTGCGCAC	CGAGAA
14. NW004444141.1 6	TTCTCCAC	CTGGGTAAAG	GTTTGATTT	TTCTCCTAT	CCGAGTGCG	TGA - GCT	TGCGGGGT	TTACAG	TCGG	- GGAGGCAC	TCCGT	GTTGCACAG	CAAGAA
15. NW004444141.1 7	TTCTCCAC	CTTGGTAAAG	AGATGACCTT	TTCTCCCAT	TTGAGTACG	TGAAGCCA	ATGTGGGGT	TTATGTGTT	TTGA	GTAGGCGC	TCTGT	CCTGCCCGG	CGAGAA
16. NW004444013.1_1	TTGTCTAC	CTGGATAAAG	TGTATGACTT	AT-GCCATT	TCAAGTAGT	TGAAGCC	STTCAGGGT	TTGCAGCAT	TCCG	GGAGGCAT	TCTGT	GCTGCGTGGG	CTGAGAA
17. NW004444013.1_2	TTGTCTAC	CTGGATACAG	ATATGACTT	TTCAACCTT	TCAAGTCGT	GAAGCAC	G T T C A G G G C	TTCCAGCAT	TCCG	GGAGGCAT	TCTGT	GCTGCGCGG	CTGAGAA
18. NW004444013.1_3	TTTTCTAC	CTGGATAAAG	TGTATGACTT	TTCTCCCTT	TCAAGTAGT	TGAAGTCC	JTTCAGG TT	TGACAGCCT	TCCG	GGAGGCAT	TCTGT	GCTTCACGG	CTGGAA
19. NW004444013.1_4	TEGTATAC	CIGGATAAAG	I G I A I G A C A I	TTCTCCCTT	TCAAGTAGT	TGAAGCC	a TTCAGGGT	TTACAGCAT	TCGG	GGAGGCAT	TCTGT	GETGEGEGEG	TGAGAA
20. NW004444013.1_5	TTGTCTAC	CTGGGTAAAG		TTCTCCCTT	TCCACTACT	TGAAGCC		TTACAGCAT	TCCA	GGAGGCAT	TCTGT	GCTGCTTGG	- TGAGAA
21. NW004444013.1_0	TTGTTCAC	CTGGGTAAAG	ACTTTAACTT	TTCTGTCTT	TAGAGTAC	TGAGGCC	TGCTATGT	TTACAGCTI	TCCG		TCTGT	GCTGCGCGG	GAGAA
23. NW004443969.1 1	TTGTCTAC	CTGGATAAAG	TGTATGACTT	TTCTCCCTT	TCAAGTAGT	TGAAGCCA	ATTCAGGGT	TTAAAGCAT	TCCG	GGAGGCAT	TCTGT	GCTGCCAG	TGAGAA
24. NW004443969.1_2	TTGTCTAC	CTGTATAAAG	TGTATGACTT	TTCTCCCTT	TCAAGTAGT	TGAAGC	TTCAGGGT	TTGCAGCAT	TCGG	GGAGGCAT	TCTGG	GCTGCGCGG	CTGAGAA
25. NW004443969.1_3	TTGTCTAC	CTGGAGAAAG	TGTATGACTT	TTGTCCCTT	TCAAGTACT	TGAAGCCA	ATTCAGGGT	TTACAGCAT	TCC	GGACGCAT	TCTGT	ACTGCGTGG	CTGAGAA
26. NW004443969.1_4	TTGTCTAC	CTGGATAAAG	TGTATGACTT	TTCTCCCTT	TCAAGTAGT	TGAAGCCA	ATTCAGGGT	TTACAGCAT	TCCG	- G A G G C A T	TCGGT	ACTACGTGG	CTGAGTA
27. NW004443969.1_5	TTGTCTAC	CTGGGTAAAG	GATCCACTT	TTCTCCCTT	TCAAGTCGG	TGAAACCA	ATTCAG	TTACAACAT	TCCG	GGAGGCAT	TCTGT	GTTGCGCGG	CGAGAA
28. NW004443969.1_6	TTTTCTAC	CTGGGTAAAG	GATCCAGTT	TTCTCCCTT	TCAAGTAGG	TGAAGC	JTTCAGCAT	TTACAGCTI	TCCG	GGAGGCAT	TCTGT	GCTGCGTGGG	CAGGAA
29. NW004443969.1_7		CCCCCTAAAG	TTTCAACTT			TGAAGCC		ATACATCTA				GATGUALGG	JCAAGAA
30. NW004443969.1_11	TTCTGTAC	C-AGGTAAAG	TGTTAAACTT	TTCTCCCAT	TCAAATATA	TGAAGCC	TTTGGGGT	TTTGATGI	TTTC	AGCAGTAC	TCCAT	GCTGCCAT	CATTGA
32. NW004443969.1 13	TTCTCTAC	CAGGGTAAAG	TGTCTAACTT	TTCTCCCAT	TCAAGTAGG	TGACGCCA	ATGCGGGGT	TTAGATGT	TCA	AGATGCAC	TTTGT	GCTGCCGGG	CTTGAA
33. NW004443969.1_14	TTCT-CAG	CATAGTAAAG	GTATAACTT	TAATCTCAT	TCAAATACA	TGAAGTCA	ATACTGGAT	TTAGATGTT	ATCA	AGCAGTAC	TATGT	GCTACCCTT	TGGAA
34. NW004443969.1_15	TTCTGTAC	CAGAGTAACG	GTATAATT	TAATCCCAT	T - A A A T A TG	TGAAGTCA	ATACGGGAT	TTAGACATT	TTCA	AGCAGTAT	TCTGT	GCTGCCCAT	ICTT GAA
35. NW004443980.1_1	TTCTCCAC	CTGGGTAAAG	🛾 G T T T 👖 T T	TTCTCCCTT	CCGAGTACA	TGAAGCC	TGCGCTGT	TTACAGCTT	TCTG	GGAGGCAC	TCTGT	GCTGCGAGG	CTGAGAA
36. NW004443980.1_2	TTCTCTAC	ATGGGAAAAG	AGTTTGACTT	TTCTCCTT	CCAAGAACG	TGAAGCCA	AGCGCAGT	TTATAACCT	TCTG	GGAGGCAC	TCTGT	GCTGTGAAG	SCCACAA
37. NW004443980.1_3	I G G I C I A C	CIGGGIAA-G	GITIGACII	TTCTCCCC	CAGAGAIIG	TGAAGC	I GCAGGGI	TTACAGCTI	TCAC	GGAGICGI	TCCCT	GCTGCACAG	GAGAA
38. NW004443980.1_4	TTTTTCTAC	CTGGATAAAG	TGTATGACTT	TTCTCCCTT		TGAAGCC		TTCCAGCAT	T	GGAAGAT	TCTTT	GCTGCACGG	CTGAGAA
40 NW004444309.1_1	TTCTCCTC	CTGGGTAAAG	GTTTGACTT	TTCTCC	TGGAGTAGT	TGAAGCC	TGCAGGGT	TTACAGCTT	TCCG		TCTGT	GCTGTCTGG	GAGAA
41. NW004444509.1 2	TTGTCTAC	CTGGAGAAAG	TGTATGACCT	TTCTCCCTT	TCAAGCAGT	TGTAGCC	TTCAGGGT	TCATAGCCT	TCCG	GGAGGCAT	TCTGT	GCTGCGCGG	CTGAGAA
42. NW004444509.1_3	TTGTCTAC	GTGGACAAAG	TGTATGACTT	TTCTCCCTT	TCAAGTAGT	TGAAGCCA	ATTCAGGGT	TCATAGCCT	TCCG	GGAGGCAT	TCTGT	GCTGCGCAG	CTGAGAA
43. NW004444509.1_4	TTGTCTAC	CTGGATAAAG	T G T 🗛 T G 🕇 C T T	- TCTTCCTT	TCAAGTAGT	TGTAGCCA	ATTCAGGGT	TCATAGCCT	TCCG	AGAGGCAT	TCTAT	GCTGCGCGG	ATGAGAA
44. NW004444509.1_5	TTGTCTAC	CTGGACAATG	TGTATGACTT	TTCTCCCTT	TCAAGTAGT	TGTAGCC	ATTCAGGGT	TCATAGCCT	TCCG	AGAGGCAT	TCTAT	GCTGCGTGG	AGGAGAA
45. NW004444509.1_6	TTGTCTAC	CTGGGTGAAG	A G T T T T C C T T	TTCTCCCGT	TCGAGTATG	TGAAGCC	TGCATGGT	TTACAGCTT	TCCC	AGAGGCTC	TTGT	GCTGCGGTG	CGAGAA
46. NW004444509.1_7	TTATCAAC	CAGAAAAAAG		TTCTCCCAT		TGAAGUUA	TAACACCT			AGAIGCAL	T-TGA	GCTGCCCCA	
47. NW004444509.1_8	TTCTATAA	CAAGGAAAAG		TTCTGCCAT	CCAAATCCG	TGAATCC	ATACGGTGT	TTAGATOCT		AGATGCGC	T-TGA	GCTGCGCAA	CGTGAA
49. NW004444309.1 1	TTGTCTAC	CTGGATAAAG	TGTATGTCTT	TTCTACCTT	TCAAGTAGT	TGAAGAC	TTCAAGG	TTCCAGCAT	TCCG	GGAGACAT	TCTGT	GCTGCGCGG	TGAGAA
50. NW004444309.1 2	GTGTCTAC	TGGCTAAAG	TGTATGACTT	TTCTCCCTT	TTGAGAAAC	TGAAGCCA	ATTCAGGTT	TTACAGCTT	TCCG	GGAAGCCT	TCTGT	GCTGCGTGG	CTGCGAA
51. NW004444309.1_3	TTGTCTAC	CTGGATAAAG	TGTATGACTT	TTCTCCCTT	TCAAGTAGT	TGAAGCC	S TTCAG T GT	TTCCAGCAT	TCCG	GGAGACAT	TCTGT	GCTGCGCGG	CTGAGAA
52. NW004444309.1_4	GTGTCTAC	CTGGATAAAA	TGT <mark>A</mark> TGACTT	TTCTCCCTC	TCAAGTAGT	GGAAGCC	G T G CAGG <u>G</u> T	TTACAGCAT	TCCG	GGAGGCCT	TCTGT	GCTGCGCGG	CTGAGAA
53. NW004444309.1_5	GTGTCTAC	CTGGATAAAG	TGTATGACTT	TTCTCCCTT	TCAAGTACT	GAAGCAA	ATTCAGGCT	TTACACCAT	TCCG	GAAAGCAT	CCCAT	GCTGCGCG	CTGCGTA
54. NW004444309.1_6	GIGICIAC	I G G A T A A A G	IGIAIGACII	TTCTCCCTT	TCGAGIAGI	TGAAGCC	TTCAGGGI	TTACAGCAI	ICCG	GGAAGCCI	TCCCT	GCTGCGCGG	- I GCGAA
55. NW004444309.1_7	TTCTCTAC	CTCCTTAAAG	TGTATGACTT	TTCT	TCAAGTAGT	TCAAGECA		TTACAGCAT	TCCC	GGAAGCAT	TCTTT	GETGEGEGEG	- TGAGAA
57 NW004444309.1_8	TTOTCTAC	CTCCGTAAAG	GATCCACTT	TACTCCTTT		TGAAGEC	ATTCAGOGT	TTACAACTT	TCCA	GGAGGCAT	TCTGT	GCTCTGC-G	CTGAGAA
58 NW004444309.1 10	TTGTCTAC	CTGGGTAAAT	GATCCACTT	TTCTCCTTT	TCAAGTGGG	TGAAGCC	TTCAGCAT	TTACAGCTI	TCCC	GGAGGCAT	TCTGT	GATGCACGC	GAGAA
59. NW004444309.1 11	TTGTCTAC	CTGGATGAAG	GACCCACTT	TTCTCCCTT	TCAAGTAGG	TGAGGCCA	ATTCAGCGT	TTACAGCTT	TCCA	GGAGGCAT	TCTGT	GCTGTGCGC	CAAGAA
60. NW004444309.1_12	TTGTCTAC	CTGGGTAAAC	GTTTGACTT	TCATCCCGT	TAGAGTTGG	TGAAGCC	TGTGCGGT	TTACAGCT	TCTG	GGAAGCGT	TTGT	GCTGCATGG	CAAGAA
61. NW004444309.1_13	ΤΤGΤΤΤΑΤ	CTG <mark>AG</mark> TAAAG	ACTTCGACTT	ATATCCCGT	TGGAGTACG	TGAAGCC	T GT G G G G T	ΤΤΑ Α Α Τ ΟΤΤ	TCTG	GGAGGCAC	TCTTT	GCTGTGGGGG	CTGAGAA
62. NW004444309.1_14	TTGTCTAC	CTGGGTAAAG	AGTTTGAATT	CTCTCCAGT	TCGAGTAGG	TGAAGAC	TGTGGGGT	T-ACAGCTT	TCTG	GGAGGCAA	TCTGT	GCCGCCAG	CAGAGAA
63. NW004444309.1_15	TTGTCTAA	GTGGATAAAG	GTTTGACTT	TTCTCCCCT	TCGAGTATG	TGAAGGC	TGTTGGGT	TTACAGCTT	TCCA	GGAGGCAC	TCTGT	GCTGTGCCG	TGAGAA
64. NW004444309.1_16	I GGICIGC	CTCCCTAAAG	GITIGACTT	TTCTCCCCT	TCCAGICCG	TGAAGCC	ACGGGGT	TTCACCTT		GGAGGCAC	TCTAT	GETETEC	AGAGAA
66 NW004444309.1_17	TTCTCTAC	CTGGGTAAAG	GTTAGACCT	TTCTTCCTT	TCGACTACA	TGAAGC	TGCAGGGT	TTATAGCTT	TCCA	GGAGACAC	ATTGT	GCTGTGTGTGG	TGAGAA
NY 1811 10 1919 1919 20 2.1 10				A DATE OF A DATE					1 2 2 7				- HOUDA

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

			640	650	660	670	_	680	690	700	710	720	730	D
	Consensus			CGGTATCCC	AAAGGAGAAG	AGAAACC-	GC	GTTTCTCC		AMACGACAG	- CGGCAGAGC	CGTAGTTTTGCT	GAGT - A	AACAC
	identity		GT TAGAGE				6.6							-
	1. NW004444102.1_1 2. NW004444102.1_2	GACCG	CTATACAGA			AGAAACA	GC	- ATTTGICCO	CATTCIGGA	ACACGAALG	AGGCAAAAC	CATAGTIGIGCI	GCGT-A	AAAAC
	3. NW004444102.1 3	TACCG	CTTTACAGA	CGGTATCGCI	TAAAGAAGAAG	AGAAACC-	GC	GTTTCTGC		ACACGAATG	AGGCAAAAG	CGTAGTTGTGCT	GCGT - A	AAAAC
	4. NW004444102.1_4	TACTG	CTTTACAGA	CGGTATCGCT	TAAAGGAGAAG	AGAAAC	GT	GTTTCTGC	GTTCTGG	ACACGAATA	AGCCAG-GC	CACAGTTTGGCT	GAGT - A	AAAAG
	5. NW004444102.1_5	AGCCG	CCTTACAGA	CGGTATCGCT	TAAAGGAGAAG	AGAAACG	GC	GTTTCTCC	CGTTCTGGT	GAGGACAG	- AGGCAGCGC	GTAGTCTTGCT	GAGTGA	AAGAC
B. NOD244441.1 TACC M. CETTLA A GARGETATE GETLB A AMOAGA AGAA GAA ALCALE GC GTTTET CONTENT OF ALCALE GARAGE	6. NW04444102.1_6 7. NW004444102.1_7	TGCCG				AGAAACC-	AC			ALGCAACAG		GIAGICIIGCI	GAGT-A	AAGAC
9. INVOX44411.1 11. INCGG CTTTA CAGG CRITTC CTTA CAGG CRITTC CTG CAGG CAGA CAGA CAGA CAGA CAGA CA	8. NW004448733.1	TACCA	CTTTACAGA	ACGTATCGCI	T G A A A G A G A A G	AGAAATC-	GC	GTTTCTCC	CCTTCTGCA	ACACGAATG	AGGCAAAAC	CTTAGTTGTGCT	GTGT	AAAAC
10. NUMBER 11. STECCE CTTM ACGACCEGTATTE CTAAGGACAAACGAAACGAAACGAAACGAAACGAAACGA	9. NW004444141.1_1	TAGCG	CTTTACAG	GCAGTATCGCI	TAAAGGAGAAG	AGAAATC	GC	GTTTCTGC	CATTCTGGA	GCACGAATG	AGGCAAAAC	CGTATTTGTGCT	GAGGG	AAACC
	10. NW004444141.1_2	TACCG	CTTAACAGA	CGGTATTGCT	TAAAGCAGAAG	AGAAACC	GC	GTTTC	GCT	AGGACAG		GGTAGTTTTTT	GAGT - A	AAAAC
	11. NW004444141.1_3 12 NW004444141.1_4	TGCCG			I GAAGGAGAAG	GGAAAC	AT			ALAGGACAG		GTAGTTTTGCT	GAGT - A	AAAAC
It MONOHANDAL THE CHIC CERTIFICACIÓN	13. NW004444141.1 5	TGCCG	CCTTACAGA	CGGTATCCCI	GAAGAAGAAG	ATACAGC -	AC	GTTTCTCC	CGTTCTGGT	GTATGACAG	- CAGTAGAGA	TATAGTTTTGCT	AAGA A	AACAC
11 W100444411 T T T C C T C <td< th=""><th>14. NW004444141.1_6</th><th>TACTG</th><th>CCTTACAG</th><th>GGTAT<mark>GT</mark>CT</th><th>T <mark>G</mark> A A G G A <mark>A</mark> A A A</th><th>GCAAAGC -</th><th>AC</th><th>GTTTCTCC</th><th>CCTTCT</th><th>CCCAG</th><th>- CAGTAGAGA</th><th>GTAGTTTTGCT</th><th>AAGG A</th><th>AACAC</th></td<>	14. NW004444141.1_6	TACTG	CCTTACAG	G GTAT <mark>GT</mark> CT	T <mark>G</mark> A A G G A <mark>A</mark> A A A	GCAAAGC -	AC	GTTTCTCC	CCTTCT	CCCAG	- CAGTAGAGA	GTAGTTTTGCT	AAGG A	AACAC
12 Involution 12 Involution 13 Involution	15. NW004444141.1_7	T - T G	CCTTACCGA	CIGTITIC	G AAAGAGAAA		AA	GTTTCTGC	CTTTCAAGT	ATACGGCAG	- CAAGAGAGA	TGTCTGTTTGCT	GAGGA	AACAC
IN NON-DATABLE NECCG CTT GACAGANE GTATIC CTAAAGGAGAAAC GGG GTTTTTCTCCGTTTGGAAGANE GAATAC GGG GAATAC GGG GAATAC GGG GAATAC GGG GATTTTCCCGTTTGGAAGANE GAATAC GGG GAA	16. NW004444013.1_1 17. NW004444013.1_2	TACAG	CTTTACAGA	GGTATCGCI		AGAAACC-	GC	GTTTCTCC	ATTCTGGA				GCGTGA	AAAAC
19. Modulation CG CTTTACCGARGEGTATTCCTAAAGGAGAAACU GC GTTTTCCCGTTTGCGAAU GCGCAGAGCGAAU GCGCAGAGCGAAU GCGCAGGGCGAAU GCGCAGGGCGAAU GCGCAGGGCGGAAU GCGCAGGGCGGAAU GCGCAGGGCGGAAU GCGCAGGGCGGAAU GCGCAGGGCGGAAU GCGCAGGGCGGAAU GCGCAGGGCGGAAU GCGCAGGGCGGAAU GCGCAGGGCGGAAU GCGCAGGGCGGGAAU GCGCAGGGCGGAAU GCGCAGGGCGGGAAU GCGCAGGGCGGAAU GCGCAGGGCGGCAUU GCGCAGGGCGGAAU GCGCAGGGCGGAAU GCGCAGGGCGGAAU GCGCAGGGCGGAAU GCGCAGGGCGGAAU GCGCAGGGCGGAAU GCGCAGGGCGGCAUU GCGCAGGGCGGCAUU GCGCAGGGCGGCAUU GCGCAGGGCGGCAUU GCGCAGGGCGGCAUU GCGCAGGGCGGCAUU GCGCAGGGCGGCAUU GCGCAGGGCGGGCAUU GCGCAGGGCGGCAUU	18. NW004444013.1 3	-ACCG	CTTTACAGA	TCGTATCCCT	TAAAGGAGAAG	AGAAAC	GT	TTTCTCC	CGTTCTGGA	ACATGAATG	AGGCAAAAC	CGTAGTTGTGCT	GAGT-G	AAAAC
20 NW044499311 T T CC C C TTT A CAGACGETATE CC TA A GGAGAGAGAGAGAGAGAGAGAGAGE C GE G TTT CTC CC TTT C GGAGAGAGGE C G TTT GC T A A GGAGAGGE C C TTT C C C TT A CAGACGE C G TT C C C TT A CAGACGE C G TT C C C TT A CAGACGE C G TT C C C TT A CAGACGE C G TT C C C TT A CAGACGE C G TT C C C C TT A CAGACGE C G TT C C C TT A CAGACGE C G TT C C C TT A CAGACGE C G TT C C C TT A CAGACGE C G TT C C C TT A CAGACGE C G TT C C C TT A CAGACGE C G T T C C C T A CAGACGE C G T T C C C T A CAGACGE C G T T C C C T A CAGACGE C G T T C C C C T A CAGACGE C G T T C C C C T T C C C G T C G C G C	19. NW004444013.1_4	GACCG	CTTGACAGA	CGGTATCCCT	TAAAGGAGAAG	AGAAAC	GC	GTTTCTCC	CGTTCTGG	ACACGAATG	AGGGAAAAC	C G <u>T</u> A G T T <mark>G</mark> T G C T	GAGT - A	AAAAC
12 10 <td< th=""><th>20. NW004444013.1_5</th><th>TACCG</th><th>CTTTACAGA</th><th>CGGTATCCCI</th><th>TATAGGAGAAG</th><th>AGAAACC-</th><th>GC</th><th>GTTTCTCC</th><th>CGTTCTGG</th><th>ACATGAACC</th><th>AGGCAGAGG</th><th>CGCAGTTTTGCT</th><th>GAGT - A</th><th>AAGAC</th></td<>	20. NW004444013.1_5	TACCG	CTTTACAGA	CGGTATCCCI	TATAGGAGAAG	AGAAACC-	GC	GTTTCTCC	CGTTCTGG	ACATGAACC	AGGCAGAGG	CGCAGTTTTGCT	GAGT - A	AAGAC
Image: State Contracts Image: State Contracts<	21. NW004444013.1_6 22. NW004444013.1_7	TGCCG		AGGTATICC	T G A A G G A G G A G	AGAAAGC -	AG			ALAGGACAG		GTAGITITGCI	GACC - A	AACAC
24. MV00444399.1 TAICLG CTTTACAGAGAGAAAAGAGAGAAAACG GC GTTTTCTCCCATTCTGGAABACGAAAGG GC GC GTTTTCTCCCGTTGGGAABACGAGAGAGAAACC GC GC GTTTTCTCCCGTTGGGAABACGAGGAGAGAAACC GC GC GTTTTCTCCCGTTGGGAABACGAGGGGGGGGGGGGGGGGG	23. NW004443969.1 1	CACCT	CCTTACAG		TAAAGGAGAAG	AGAAAC	GC	GTTTCTCC	CGTTCTGG	ACA - GAATG	AGGCAGAGC	CGCAGTTGTCCT	GAAT - A	AAGGC
25. WV0444399.1.5 MGCC G CTTTTALAGACGUTATCGCTAAAGGAGAAAACC G C GTTTTCTCCCGTTUGGGAAALGAGAGGACAGGAGAGAGAAACC G C GTTTTCTCCCGTTUGGGAAALGAG GACAGGAGAGAGAAACC G C GTTTTCTCCCGTUTUGGGAAALGAG GACAGGAGAGAGAAACC G G GTTTTCTCCCGTUTUGGGAAALGAG GACAGGAGAGAGAAACC G G G TTTTCTCCCGTUTUGGGAAALGAG GACAGGAGAGAGAAACC G G G TTTTCTCCCGTUTUGGGAAALGAG GACAGGAGAGAGAAACC G G G G TTTTCTCCCGTUTUGGGAAALGAG GACAGGAGAGAGAAACC G G G G TTTTCTCCCGTUTUGGGAAALGAG GACAGGAGAGAGAAAACC G G G G TTTTCTCCCGTUTUGGGAAALGAG GACAGGAGAGAGAAAACC G G G G TTTTCTCCCGUTUGGGAAALGAG GACAGGAGAGAGAAAACG G G G G TTTCTCCCCGUTUGGGAAALGAG GACAGGAGAGAGAAAACG G G G G TTTCCCCCGUTUGGGAAALGAG GACAGGAGAGAGAAAACG G G G G TTTCCCCCGUTUGGGAAALGAGGAGAGAGAAAACG G G G G G TTTCCCCCGUTUGGGAAAGGAGAGAGAAAACG G G G G G TTTCCCCCGUTUGGGAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	24. NW004443969.1_2	TACCG	CTTTACAGA	CGGTATCACT	TAAAGGAGAAG	AG-AAC	GC	GTTTCTCC	CATTCTGGA	ACACGAATG	AGGCAAAAG	C G T A G T T <mark>G G</mark> G C T	GAGT-A	AAAAC
92 NMGC426 CUT TGC CG GC TTT GC CG GG AT TC CC TG AG GG GG AG AG AG AG AG ACC GC GC TTT TC CC GG TA TA GG GG CG GG CG GG GG GG GG GG GG GG GG	25. NW004443969.1_3	TACCG	CTTTATAGA	CGTATCGCI	TAAAGGAGAAG	AGAAAAC -	GC	GTTTCTCC	CGTTTGGGA	ATATGAATC	AGGCACAGC	CGCTGTTTTCCT	GAGT - A	AAGGC
28. WW00444396915 TGCCG CCM TGCAGACKMTTATCCCTGAAGGAGAAGAAGAAGT GC AAAGT GC 29. WW0044396915 TGCCG CCM ACTTACAGGCCGGCACAGAGAGAAAAGT GC GGTTTTCCCTGAGTAAGGGAGAGAGAGAGAGAAAAGT GC 30. WW00443969112 TGTGA CCTTACAGACGGAGAAAAGGAGAAAAAGT GC GGTTTTCCCTGAGTAAGGGAGAGAGAGAGAGAGAAAAGT GC 30. WW00443969112 TGTGC CCTTACAGACAGAGAGATCTCCTGAAGGAGAAAAGAAGC GGGGAGACAAAGAAGAGC GGTTTTCCCTGAGTAAGGGGAAAAGAGGAGAGAAAAGAAGAGC 31. WW00443969114 TGTGC CCTTACAGACAGAGAGAAAAGGGAGAAAAGAAGC GGGGGAGCCG GTTTCTCCTGAGTAAGGGAGAAAAGAAGAGGGGAGAAAAAGAAAG	26. NW004443969.1_4 27. NW004443969.1_5	AGCCG		AGGTATAGC		AGAAACC-	AC		AGTTCTGGT	ATAGGACAG			GAGT - A	GAG
29. NW004443969.11 TIGTG - CGTTACGACG GG ATCHCCTG AAGG GG AAGAA BAAGG GG GTTTCCCCGG TTTGCGTTCGG GTACAGG GG CGG GG AAGAAGG GG AAGAAGG GG GTTTCCCCGGTTGG GTACAGG GG GG AAGAAGG GG AAGAAGG GG GTTTCCCCGGTG AGG GG AAGAGG GG GG AAGAAGG GG GG AAGAAG	28. NW004443969.1_6	TGCCG	CCTTGCAGA	CATTATCCCT	T G A A G G A G A A G	AAAAC	GC	ATTTCTCA	GGTTCTGGT	ATATGACAG	- CGGCACAGC	AGCAGTTTTGCT	GAGT - A	AAGAC
	29. NW004443969.1_7	AGCAG	CTTTTCAG	G G G C A T C T C T	T <mark>G</mark> A A G G A G A A <mark>A</mark>	ACAAAGT	GC	CTTTCTCC	CGTTCTGGT	ATACGACAG	CAGTAGAGA	TATAGTATCTCT	AAGC A	AACTC
121 NV004439951.14 TARTS - C ET LA CA GA CAGATA CA AGG CA ALA CA ALA GA CALA CA CAGATA CA CAGATA CA CAGATA CACATA CACA	30. NW004443969.1_11	TGTGA	- CTTACACA		T G A A G G A G A A A	ACAATGC -	G	GTTTC-CTC	CGTTCCAGT	ACAGGACTG	CAACAGAGA	GTAGATTCGCT	AAGGA	AACAC
33. NW004439951,15 TGG - CCTTACAGACAWAATCHCTGAAGGAGAAWACAAGC GW GTTTCTCCCCGTGAAGTACAGAAACAAGAC CAW CAAWAATCAGAGATTAGTTAGTTAGCTAAGTAAACCAC 35. NW004439801,1 TGCTG CCTTACAGACGATAGACAAAAAGC GG GTTTTCCCCTAGGAAAAGCACAW TAGGAGAAATTACACAW TAGGAGAGAGAGAAATTACACAW TAGGAGACACAW TAGGAGAAATTACACAW TAGGAGAGAGAGAAATTACACAG GG GTTTTCCCCTATACAGACGA GTAGTTTTTCCCTAAAGGAGAAAAGC GG GTTTTCCCCCTTCTGGTAAATTAGACACAW TAGGAGAGAGAGAAATTACACAW TAGGAGAGAGAGAGAAAATTACACAG GG GTTTTCCCCCTTTCGGTAAATTAGACACAW TAGGAGAGAGAGAAAATTACACG GG GTTTTCCCCCTTCTGGTAAATTAGACACAW TAGGAGAGAGAGAAAATTACACG GG GTTTTCCCCCGTTCTGGGTAAATTAGACAG GG GGGGAGAGGGAGAGAGAGAAAAAGTG GG GG GG GTTTTGCCCCGTTCTGGGTAAATTAGGAGCGG GG	31. NW004443969.1_12 32. NW004443969.1_13	TAT		GCI	TGAAGGGGAAC	ACAATGC -	AA	GTTTCTCC	CAGTCAAG	AGGACAG		TATACTTTTGCT	AAGT	AAGAC
94. NV00443999.11 5 C GT TACAGAC MGTATTC TC TG AAGGACAAAA TG C GA G TT CTT CC TT CG TA TG AA GAC GA AACAC C 36. NV00443980.12 TG CG G C CTT TACAGA MGTATTG CT G AAGGAC GAAAA TG AG GC G G G TT TC TT CC C CATTTC GG TA AAGAC GA GAAGAAAAAAAAAA	33. NW004443969.1_14	TGTG-	CCTTACAGA		TGAAGGAGAAA	ACAAAGC -	GA	GTTTCTCC	CCGTGAAGT	ATAGGACAA	- CAATAGAGA	GTAGTTTCCCT	AGGT - A	AACAC
35. NW004443980.12 TGCIG C CTTALAGAAGGIA TGACIGAAGALAGAAGC GG GTTTTCTCCGTTTLCCGTALAGAGAGACAGC C CAGTTTCTCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGA	34. NW004443969.1_15	TGTG-	CCTTACAGA	CAGTATCTCI	T G A A G G A G A A A	ACAATGC	GA	GTTTCTCC	TCGTGAAGT	ACAGAA	CAATACAGA	TGCAGTTTCCCT	AAGT A	AACAC
39. NUNDERLANDER LE THACA GALEGE THE COLUMA AGEN ALAGE AND AGEN UNDERLANDER LE COLUMA AGEN ALAGEN	35. NW004443980.1_1	TGCTG	COTTACAGA	AGGTATGACI	TGAAGACAAAA	ACAAAGC -	GC	GTTTCT		ACACTACAA	- TAGTGGAAA	GTAGTTTTCT	AAGG A	AACAC
38. NW00444399.14 TG CC M CGTTACAGG CGTTATTGCTGAG AGGAGAAACC GC GATTTGCCCGTTTGGAATGAGC GAGGCGGAGAGC GGAGTTTGGTGGG AAAACAG GC GATTTGGCTGGG AAAAACAG GC GATTTGGCTGGG AAAAACAG GC GATTTGGGAGAAACG GAGAGAAACG GATTTGGCCGTTGGGAATAAAACC GAATGAAACG GAGGCGGAGAAACC GAATGAAACG GAGGCGGAGAAACC GAATGAAACG GAGGCGGAGAAACC GAATGAAACG GAGGCGGAGAAACC GAATGAAAGC GAATGAAGC GAATGAAGC GAATGAAGC GAATGAAGC GAATGAAGC GAATGAAACC GAATGAAGC GAATGAAGC GAATGAAGC GAATGAAACC GAATGAAGC GAATGAAGC GAATGAAGC GAATGAAGC GAATGAAGC GAATGAAGC GAAGAGAAACC GA GATGATTGGCGGCGGAGAAGGAGAAACC GAA GATGAATGAAGGCGCGGAGAGGCGGGAGAGGGAGAAACC GAC GATTTTGCCGATTTGCGGGAGAAAGC GAACC GAATGAAGAGAAACC GAA GGAAGGAAACC GAC GATTTTGCGGGAGAAAGC GAAGGAGAAACC GAC GTTTTGCGGAGAAAGGGAGAAACC AGC GTTTTGCGGGAGAAAGGAGAAACC AGC GTTTTGCGGAGAAAGGGAGAAACC AGC GTTTTGCGGAGAAAGGAGAAACC AGC GTTTTGCGGAGAAAGGGAGAAACC AC GATTTTGCGGAGAAAGGGAGAAACC AC GATTTTGCGGAGAAAGGAGAAACC AC GATTTTGCGGAGAAAGGGAGAAACC AC GATTTTGCGGAGAAAGGGAGAAACC AC GATTTTGCGGAGAAAGGAGAAACC AC GATTTTGCGGAGAAAGGGAGAAACC AC GATTTTGCGGAGAAAGGGAGAAACC AC AGC GTTTCCCGGTTCGGAACGAGAGGAGAG	37 NW004443980.1_2	TGCAG	CCTTACAGA		TAAAGGAGAAAG	AGAAAG	G	GTTTCTCC	CGTTCTGGT	ATATGACAG		CGCAGTTTTGCA	GAGT - A	AAGAC
99. NW00444399:11 TACCG CTTTACAGA MGGTATCNCTAAAGAGAGAAACC CG GATTTCTCCCATTCCGAACAGAACCGAAAACC GATTTCTCCCCATTCTGGGAACAGAGAGAAAACC GATTACAGAGAGAGAGACC GAACAGAACAGAGAGAAAACC GGGTTTCCCCAAAGAGAGAGAGAAAACC GGGTTTCCCGAACAGGAGAGAGAGAAACC GATTTCCCCCAAAGAGAGAGAGACACC GACGAACAGAGAGAGAGAGAAACC GACGAACAGGAACAGGAGACGC GACGAACAGGAACAGGAGACAGGAAACC GACGAACAGGAAACC GACGAACAGGAAACC GACGAACAGGAAACC GACGAACAGGAAACC GACGAACAGGAAACC GACGAACAGGAACAGGAACAGGAACC GACGAACAGGAACAGGAACC GACGAAGAGAGAAGAGAACC GACGAACAGGAACAGGAACC GACGAGAGCGCGCGAAGCCGGAAGCCGGAGGCCGGAGGCCGGAGGCCGGAGGCCGGAGGCCGGAGGCCGGAGGCCGGAGGCCGGAGGCCGGAGGCCGGAGGCCGGAGGCCGGAGGCCGGAGGCCGGAGGCGGAGGCGGAGGCGCGAAGGCAAGGAAAGGAAAGGAAAGC GC GTTTACAGAAGAGAAAGGAAAGGAAAGC GTTTCCCCGAGCAGGAACGGAAGGGAAACC GACGAGGAAGGGAAACC GACGAGGAAGGGAAACC GTTTCCCCGAGCGCGAAGGCGCGGAGGCGGAGGCGGAGGAAGGAAAGC GTTTCCCCGAGGAGGGCCGAAGGGCAATTTCCCGAAGAGGAAAGGGAAAGC GTTTCCCCGAGGAGGGCGCGGCGGCGGCGGGGGGGGGGG	38. NW004443980.1_4	TGCCA	CCTTACAG	CGTTATTGCT	T G A T G G A G A A G	AGAAACC	GC	GTTTGTCC	CCTTCTGGT	ATAGGACCG	AGGCAGAGT	CGCAGTTTTGCT	GAGT-A	AAGAC
40, NW00444309.1] 19 TACCG CTTTACAGACGTATTCCCTAAAGGAGAAAACC GA GTTCCCCCGTGTGGAACAGGAACAG AAGCAGAAACC GA GTTCCCCCGTGGGT AAACAG 41, NW00444509.1] TACCG CTTTACAGACGTATCCCTAAAGGAGAAACC GC GTTCTCCCCATTCAGGAACAGGAACAG GAGCGGCGCGCGCGTTTGCTGGGT AAACAC 43, NW00444509.1] TACCG CTTTACAGAAGGGTATCCCCTAAAGGAGAAACC GC GTTCTCCCCATTCAGGAACAGGAACCGGCGCGCGCGCGCG	39. NW004444509.1_1	TACCG	CTTTACAGA	AGGTATCACT	TAAAGAAGAAG	AGAAAC	GC	ATTTCTCC	CATTCTGGA	ACACGAATG	AGGCGAAAC	C G T A G T T <mark>G</mark> T G C T	GAGT-A	AAAAC
11. INVOUNDENDED IN A CAGAGE CONTRACTOR AND CONTRAC	40. NW004444309.1_19	TACCG	CTTTACAA	CGGTATCCC	I GAGGGGGAGAAA	ACAAACC	GCC	GTTTCTCC		ALALGACAG		GIAGICICGGI	AAAI - A	AACAC
43. WW004444599:1-4 TACCG - CTTTACAGA/BGGTATCCCTAAAGAGAAAGC - AC - AC - GTTCTCCCTGGAACAGAAAG GAAAGC - GTTACTCCCAGTCTGGAACAGAAACG - GC - GTTCTCCCGGTCTGGAACAGAACGAACG - GC - GTTCTCCCGGTCTGGAACAGAACGAACGAACG - AAAGAC 44. WW004444599:1-5 TACCG - CTTTACAGA/BGGTATCCCTAAAGAGAAAGC - GC - GTTCTCCCGGTCTGGAACAGAACGAACGAACACAGAAACAC 45. WW004444599:1-7 TGGG - CTTTACAGA/AGGAATCTCTGAAAAGAGAAAGC - GC - GTTCTCCCAGTCAAGGATAGGAACGA CTGTAGTTTT - CGA/AGCAAACAC 46. WW004444599:1-7 TGGG - CTTTACAGA/AGGAATCTCTGAAAGAAAGAAAGC - GC - GG - GGTTCTCCCAGTCAAGGATAGGAATGG C CTATAGAAGATGTAAGTTT - CGA/AGCAAAACAC 47. WW004444599:1-7 TGGG - CTTTACAGA/AGGAATCTCTGAAAGAAAGAAAGC - GC - GA - GGTTCTCCCAGTCAA/GGAATGG C CTATAGAAGATGTAAGTTAACGAACAC 48. WW004444599:1-7 TGCG - CTTTACAGA/AGGTATCCCCTAAAGGAAAGGAAAGG - GC - GTTTCTCCCAGTCAA/GGAAGGAAACGAGAAACAC GC - GTTTACAGA/AGGCGTAGTCTGCAAA 49. WW004444309:1-7 TACCG - CTTTACAGA/AGGAAAGGAAAGGAAAGG - GC - GTTTCTCCCGTACTAGAACAAAAGG - CG GC GGCGTTTTGCTGAGT - AAGAAC GA GGCAAAAGCCGCGCGCGTTTTGCTGAGT - AAGAAC 50. WW004444309:1-7 TACCG - CTTTACAGA/AGGAAAGGAAAGGAAAGC - GC - GTTTCTCCCGTAGCTA/GGAACACAGCA - CG GCCGCGGCGCTTTTGCTGAGT - AAGAAC GA GCAAAAGCCGCGCGCGCGTTTTGCTGAGT - AAGAAC 51. WW004444309:1-7 TACCG - CTTTACAGA/AGGAAAGAAGAAAGACC - GC - GTTCTCCCCGTACTA/GAAACACACAGCA - CG GCCGCGGCGCGCGCGCGCGTTTGCTGAGT - AAGAAC GA GCAAAAACCGCGCCGCGCGCGTTTGCTGAGT - AAGAAC 52. WW004444309:1-7 TACCG - CTTTACAGA/AGGAAAGAAGAAAGAACC - GC - GTTCTCCCCGTACTA/GAAACACACACACAACAGCGCCGCGGCGCGCCGCGC	42. NW004444509.1 3	TGCCC	CTTTACAGA	CGTTATCCCI	TAAAGGAGAAG	AGAAACC -	GA	GTTTCTCC	CATTCAGGA	ACATGAATG	AGGCAGAGC	CGCAGTTTTGCT	GGGT - A	AACAC
<pre>44.WW004444599.15 TACCM G CTTTACAGAAGGTATCCCTAAAAGAGAAAACCC AC G TTTTCTCCCGTTCTGGAACAGCAGAGCCGCGCGGCGTTTTCTCTGAGT AAAGAC 45.WW00444599.17 TGGG - CTTTACAGACGAGAATCMCTGAAGGAAAGGCAGTGC - GA GGTTCTCCCAGTTCAGGAGGGCCGCGCGGCGGAGTTTTCTCGAAGAACAC 47.WW00444599.19 TGTG - CTTTGCAGACAGGAATCMCTGAAGGGAAAGGCAATGC - GA GGTTCTCCCCAGTCAAGGAGGCCGCGGCGTAGTCTGTGCAGAGAGAG</pre>	43. NW004444509.1_4	TACCG	CTTTACAGA	AGGTATCCCT	TAAAGGAGAAG	AGAAACC	AC	GTTTCTCT	GTTCTGG	ACACGAATG	AGGCAGAGC	CGCAGTTTTTCT	GAGT - A	AAGAC
45. NW0044445991_6 TAGCE CTTTACAGACAGAATCTCTGAAAGAGAATGC GA GGTTCTCCCCAGTCAAGTATAGGATGG CTATAGAGATGTAGTTTT-CGAAGCAAAACCAC 47. NW0044445991_8 TATG - CTTTGCAGACGGAATCTCTGAAAGAGAATGCAGTGC GA GGTTCTCCCCAGTCAAGTATAGGATGG CTATAGAGAATGTAGTATTTTGAAGCAAACAC 48. NW004444591_1 TACCG - CTTTACAGACGGTATTCCCGAAGGAATGCCGA GGCGATTTCCGA GG TCTCCCCCAGTCAAGTATAGGATGG CTATAGAGAATGTAGTTTTTGAAGCAAACAC 49. NW004443991_1 TACCG - CTTTACAGAAGGGTATCCCGAAAGGAGAAGGGAAAGC GC GC GTTCTCCCCGTCCAAGTAAAGCAGAAAGC GCGCGTCTTGGGTGAGATAAGCGATGG CAAAAAGC 50. NW004443991_3 TACTG - CTTTACAGAAGGTATCCCTAAAGGAGAAAGG GAAAGC GC GC GTTTCTCCCCGTCCAAGCAAACG GAAGGC - CGGCGAGAGC GCGCGTTTTGGTGGTGAGA 51. NW004443991_3 TACTG - CTTTACAGAAGGTATCCCTAAAGGAGAAAGG GAAACC - GC - GTTTCTCCCCGTACTAGGAACACGGA - CGGCGAGAGC GCGCGTTTTGGTGAGT AAAGAA 52. NW004443991_4 TACCG - CTTTACAGAAGGTATCCCTAAAGGAGAAAGG GAAACC - GC - GTTTCTCCCCGTACTAGGAACACGGAC - CGGCGAGAGCGGCGTTTTGGTGGGGAGAAACC - AAGAAC 53. NW004443991_5 TACCG - CTTTACAGAGGTATCCCTAAAGGAAGAGAAAACC - GC - GTTTCTCCCCGTACTAGGAACACGGAC - CGGCGCGCGTTGTGCTGAGT - AAGAAC 54. NW004443991_6 TACCCG - CTTTACAGAAGGTATCCCTAAAGGAAGAAAACC - GC - GTTTCTCCCCGTACTAGGAACACGGAC - CGGCCGCGCGTTGTGCTGAGT - AAGAAC 55. NW004443991_6 TACCCG - CTTTACAGAAGGTATCCCTAAAGGAAGAAAACC - GC - GTTTCTCCCCTTCTCGGAACACGGAC - CGGCCGCGCGTTGTGCTGAGT - AAGAAC 56. NW004443991_6 TACCCG - CTTTACAGAAGGTATCCCTAAAGGAAGAAAACC - GC - GTTTCTCCCCTTCTGGAACACGGAC - CGGCCGCGCGCGTTGTGCTGAGT - AAGAAC 56. NW004443991_8 TACCCG - CTTTACAGAAGGTATCCCTAAAGGAAGAAACC - GC - GTTTCTCCCCTTCTCGGAACACGGACGCGCGCGCGCGCGTGTGTGT	44. NW004444509.1_5	TACTG	CTTTACAGA	AGGTATCCC	TAAAAGAGAAG	AGAAACC-	AC	GTTTCTCC	CGTTCTGG	ACACGAATG	AGGCAGAGC	CGCAGTTTTCCT	GAGT-A	AAGAC
47. NV00444450915 TATG - CTTTGCAGACAGAATTTCTGAAGGAAATGCAGTGC GA GGCATTTCCCCAGTCAAGGAAGGA AAGGC CATTGCAGAAGGAAGAAACAC 48. NV0044445091 TACCG - CTTTACAGACGCATTTCCCGAAAGGAGAAGGAAAAGC GC GC GTTCTCCCCGATCAAGGAGAAGGAAAAGC GC GC GTGTTTGCTGGAAGGAAAACAC 49. NV0044445091 TACCG CTTTACAGACGGTATCCCTAAAGGAGAAGAGAAAGG GC GC GTTTCTCCCCGATCTGGAAAGGC AAAGGC GC GC GTTTTGCTGGAGT AAAGGAGAAAGG 50. NV0044443091 TACCG CTTTACAGACGGTATCCCTAAAGGAGAAGAGAAAGG GC GC GTTTTCCCCGATCTGGAACAGGAAACGG GC GC GTTTTGCTGAGT AAAGGAGAAACC GC GC GTTTTCCCCGATCTGGAACAGGAACACGGC GC GC GTTTTGCTGGAGT AAAGGAGAAAACC GC GC GTTTTCCCCGATCTGGAACAGGAGACACGGC GC GC GTTTTGCTGGAGT AAAGGAGAAAACC GC GC GTTTTCCCCGATCTGGAACACGGAACGGGACGCGCGCGCG	45. NW004444509.1_6	TACC	CTTTACAGA		I GAAAGAGAAG	GCACTCC-	GA		CACTCAAGT	ALALGACAG			AACC - A	AACAC
48. NV004444309.19 TGTG - CTTTGCAGACCAGTATTCCGAAGGAGAAGGAAGGAGAGGCGCGGTTCTCCCGTTCTGCAAGGAGGGCGAAGAGCGTGTGTGCGAGATGTAACCACACACA	47. NW004444509.1 8	TATG -	CTTTGCAGA	CAGAATCTCT	GAAGGAGAAG	GCATTTC-	GA	GTTCTCC	CAGTCAAGT	ATAGGATGG	- CTATAGAGA	GTAGTTTTT	GAAGCA	AACAC
49. NW004444309:11 TACCG - CTTTACAGACGCTATCCCTGAGAGAGAGAGAGAGAGAGACC GC - GTTTCTGCCATCTGGAAACGGAAGCCGACGCTGTTGTGCTGAGT - AAGAAC 50. NW004444309:12 TACCG - CTTTACAGAGGTATCCCTAAAGGAGAAGAGAAACGG - GC - GTTTCTCCCGTACTAGGAACACGGCA - CTGGCAGAGCCGCGTTTTGGTGAGT - AAGAAC 51. NW004444309:14 TACCG - CTTTACAGAGTTATCCCTAAAGGAGAAGAGAAACC - GC - GTTTCTCCCGTACTAGGAACACGGCA - CTGGCAGAGCCGCGCGTTTTGGTGAGT - AAGAAC 52. NW004444309:15 TACCG - CTTTACAGAAGGTATCCCTAAAGGAGAAGAGAAACC - GC - GTTTCTCCCCTTCTGGAACACAGGCA - CGGCCGCGCGTTGTGGTGGTGAGT - AAGAAC 53. NW004444309:16 TACCG - CTTTACAGAAGGTATCCCTAAAGGAGAAACC - GC - GTTTCTCCCCTTCTGGAACACAGGA - CGGCCGCAAAACCGTAGTTGTGTGTGGTGGGTGAAAAGAGAGAAACC - GC - GTTTCTCCCCTTTGTGGAACACAGGAAAGCG - AGGCCGAAAACCGTAGTTGTGTGTGGTGGGTGAAAAGAGAGAAACC - GC - GTTTCTCCCGTTCTGGAACACAGAAAGCG - AGGCCGCAAAACCGTAGTTGTGTGTGGTGGGGAAAAACC - GA - AAAAC 54. NW004444309:16 TACCG - CTTTACAGACGGTATCCCTAAAGGAGAAAACC - GC - GTTTCTCCCGTTCTGGAACACGAAAGC - AGGCGCGCGCGAGATTGTGCTGAGT - AAAAAC 55. NW004444309:18 TACCG - CTTTACAAAGGAGAAAGAGAAAACC - GC - GTTTCTCCCGTTCTGGAACACGAAAGCG - AGGCGGCGCGCGCGCGAGGTTTGCTGAGT - AAAAAC 56. NW004444309:19 TG CCG - CTTTACAAACGGTATCCCTGAAGGAGAAAACC - GC - GTTTCTCCCGTTCTGGAACACGAACCG - CGGCAGAGCGAGGAGTTTTGCTGAGT - AAAAAC 57. NW004444309:19 TG CCG - CTTTACAAACGGTATCCCTGAAGGAGAAAACC - GC - GTTTCTCCCGTTCTGGAACACGAAACCG - CGGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGAGGTTTTGCTGAGT - AAAAAC 58. NW004444309:19 TG CCG - CTTACCGAGGGTATCCCTGAAGGAGAAAGGAGAAACC - GC - GTTTCTCCCGTTCTGGTATACGACGG - CGGCAGGGCGCGCGGCGGCGCGGAGGGAGAGAGAGA	48. NW004444509.1_9	TGTG-	CTTT <mark>G</mark> CAGA	CAGTATTTCT	T G A A G G A G A A T	GCAGTGC	GA	GCTTCTCC	CGTTCAAGT	A T A G G A T G G	CTATCGAGA	T G T A G T T T T G T G	AAGT - A	AACAC
30. NW004444399:1_2 TACCG - CTTTACAGAAGGCTATCCCTAAAGGAGAGAGAGAGAAGCG GC - GTTTCTCCCGTACTAGGAACACGGCGC - CTGCCAGAGCCGCGTTTGGTGAGT - AAGAAC 51. NW004444399:1_3 TACCG - CTTTACAGAAGGTATCCCTAAAGGAGAAGAGAAACC - GC - GTTTCTCCCCTTCTGGAACACCGACCGCAAAACCGGTGTGTGT	49. NW004444309.1_1	TACCG	CTTTACAGA	CGCTATCCC	G A A A C G A G A A G	AGAAACC	GC	GTTTCTGC	CATTCTGGA	ACACGAATG	AGGCAAAAG	CGTAGTTGTGCT	GAGT - A	AGAAC
12 11 12 <td< th=""><th>50. NW004444309.1_2 51. NW004444309.1_3</th><th>TACEG</th><th></th><th>AGGTATCCC</th><th></th><th>AGAAAGG-</th><th>GC</th><th>GTTTCTCC</th><th></th><th>ACACAGCA-</th><th></th><th></th><th>GAGT - A</th><th>AGAAC</th></td<>	50. NW004444309.1_2 51. NW004444309.1_3	TACEG		AGGTATCCC		AGAAAGG-	GC	GTTTCTCC		ACACAGCA-			GAGT - A	AGAAC
53. NW004444309:15 TACCA CTTTACAGAAGGAATGTATCCCTAAAGGAGAAAGAGAAACC GC GTTTCTCCCGTAGTAGAACACGACAC ACAGACCGGAGCCGTATTTGTTGTGAGT - AAGAAC 54. NW004444309:16 TACCG CTTTACAGACGGTATCCGTAAAGGAAAAACC GC GTTTCTCCCGTTCTGGAACACACACACCG - AGGCAAAACCGTAGTTGTGCTGGGT - AAAAAC 55. NW004444309:17 TACCG CTTTACACACGGTATCCGTAAAGGAGAAAAACC GC GTTTCTCCCGTTCTGGAACACGAATC - AGGCAAAACCGTAGTTGTGCTGAGT - AAAAAC 57. NW004444309:19 TGCCG CTTTACACACGGTATCCGTAAAAGAAGAAAAAACC - GC - GTTTCTCCCGTTCTGGAACACACACACGCGCAGACCG - AGGCGGCAAAACCGGACG - AAAAAC 57. NW004444309:19 TGCCG CTTTACACACGGTATCCGCTGAAGTAAGAGAAAAACC - GC - GTTTCTCCCGTTCTGGAACACGACG - GGCCAGAGCCGCAGATTTGCTGAGT - AAAGAC 58. NW004444309:11 TGCCG CCTTACAGAGGGTATCCGCTGAAGGAGAAGAGAA	52. NW004444309.1 4	TACCG	CTTTACTGA	CGCTATCCCT	TAAAGGAGAAG	AGAAACC -	GC	GTTTCTCC	CTTTCTGGA	ACACGGATG	AGGCAAAAC	GTAGTTGTGCT	GAGT - A	AAAAG
54. NW004444309:16 55. NW004444309:10 56. NW004444309:10 57. NW004444309:10 56. NW004444309:10 56. NW004444309:10 56. NW004444309:10 57. NW004444309:10 58. NW004444309:10 58. NW004444309:10 59. NW004444309:11 50. CCTTACCGACGTATCCCTGAAGAGAAGAGAAAAGCC-GC-GTTTCCCCGTTCTGGAACACCACACC	53. NW004444309.1_5	TACCA	CTTTACAGA	ATGTATCCCT	TAAAGGAGAAG	AGAAACC	GC	GTTTCTCC	CGTAGTAGA	ACACAGCA-	- C G G C C G A G C	CGCCGTTTTGTT	GAGT - A	AGAAC
55. NW004444399:1/2 56. NW004444399:1/2 57. NW004444399:1/2 58. NW004444399:1/2 59. NW004444399:1/2 50. Second Contractor Contrector Contractor Contractor Contractor Contr	54. NW004444309.1_6	TACCG	CTTTACAGA	AGGTATCCCT	TAAAGGAAAAG	AGAAAC	GC	GTTTCTCC		ACACAAACG	AGGCAAAAC	CGTAGTTGTGCT	GGGT - A	AAAAC
35. NW004444309:19 AGCCG - TTTACAAACGGTATCGCTGAAGAAGAAGAGAAAACC GC ATTTCCCCGTTCTGGTATACGACCA - CGGCAGAGACGAGTTTTGCTGAGT - AAAGAC 58. NW004444309:10 TGCCG - CTTACAGAGGGTATCCCTGAAGGAGAAGAGAAAAGC GC - GTTTCTCCCGTTCTGGTATACCACCG - CGCCAGAGCCGCAGTTTTGCTGAGT - AAAGAC 59. NW004444309:11 TGCCG - CTTACAGAGGGTATCCCTGAAGGAGAAGAGAAAAGC GC - GTTTCTCCCGTTCTGGTATACCACCG - CGCCAGAGCCGCAGTTTTGCTGAGT - AAAGAC 60. NW004444309:11 TGCCG - CTTACAGAGGGTATCCCTGAAGGAAGAAAAGC - GC - GTTTCTCCCGTTCTGGTATACCACCG - CGCCAGAGCCGCGCGCGGCTCGTTTGCTGAGT - AAAGAC 61. NW004444309:11 TGCCG - CCTTACAGACGGTATCCCTGAAGGAGGAGAAAGC - GC - GTTTCTCCCGTTCTGGTATACCACGGCGCGCGGCTCGTTTGCTGAGT - AAAGAC 62. NW004444309:11 TGCCG - CCTTACAGGCGGTATCCCTGAAGGAGGAGAAAGC - GC - GTTTCTCCCCGTTCTGGTATACCGCGGCGCGGCGGCTGTAGTTTTGCTGAGT - AAAGAC 63. NW004444309:11 TGCCG - CCTTACAGGCGGTATCCCTGAAGGAGAGAAAGC - GC - GTTCTCCCCTTCTGGTATGGCGGCGGCGGCGGCGGTGGTGTTTTGCTGAAG 64. NW004444309:11 TGCCG - CCTTACAGGCGGTATCCCCTGAAGGAGAAAGC - GC - GTTTCTCCCCTTCTGGGGATGGCGGCGGCGGCGGCGGCGCGTGTAGTTTTGTAACC - AAACAC 65. NW004444309:11 TGCCG - CCTTACAGGCGGTATCCCCTGAAGGAGAAAAGC - GC - ATTCTCCCCTTCTGGGCAGCGGCGGCGGCGGCGGCGCGCGC	55. NW004444309.1_7	TACCT				AATAACC	GC	GTTTCTCC	CGTTCTGG	ACACGAATC	-AGGCAAAAG	CGTAGTIGTGCT	GAGT - A	AAAAC
58. NW00444309.1_10 TGCCG-CCTTACAGACGTGAAGCGCTGAAGGAGAAGAGAA	57. NW004444309.1 9	AGCCG	TCTTACAAA	CGGTATCGCT	T G A A G T A G A A G	AGAAACC-	GC	ATTGCCCC	CGTTCTGCT	ATACGACCA	- CGGCAGAGA	CGCAGTTTTGCT	GAGT-A	AAGAC
59. NW004444309:111 TGCCG-CCTTACAGAGGATATCCCTGAAGGAGAAGAGAAG	58. NW004444309.1_10	TGCCG	CCTTACCGA	GTAATCGCI	T <mark>G</mark> A A G G A G A A G	AGAAAGG	GC	CTTTCTCC	CGTTCTGGT	ATACCACCG	CGGCACAGC	CGCAGTTTTGCT	GAGTA	AAGAC
00. NW004444309:1/1 TGCCG CCTTACAGACGGTATCCCTGAGGAGAAAAAAAAGGAAGGCACGC ACTTTCTCCCCTTTCTGGGAGCGCGCGCGCGCGCGCGCGC	59. NW004444309.1_11	TGCCG	CCTTACAGA	GGGTATCCC	TAAAGGAGAAG	AGAAAGC -	GC	GTTTCTCTC	GTTCTGGT	AACAACGG	- CGICAGCGC	CGCAGTTTTGCT	GAGT - A	AAGAA
62. NW00444309.1_14 TGCCA CCTTACAGCACTATCACTGAAGGAAGAAGAAGGAAGAAAGC AC TTTTCTCCCTTTCAGGTATGGACAG AGGCACGCGCTGTAGTTTTGTTAACC AAACAC 63. NW00444309.1_15 TGCCG CCTTACAGTCGCTATCTCTGAAGGAAGAGGAAGAGGAAAACCATCTTTCTCCCTATTCTAGTATACGACAG CAGCGCTGTAGTTTTGCTAACT AAACC 64. NW00444309.1_16 TGCTG CCTTACAGGCGGTATCCCTGAAGGAAGGAGAAAGGCGC CC CATTTCTCCCATTCTGGTATACGACAG CGCGCGCGCGCTGTAGTTTTGCTAACT AAACC 65. NW004444309.1_17 TGCCG CCTTACAGCCGGTATCCCCTGAAGGAGAAAGGCGCGC CC CATTTCTCCCATTCTGGTACGGCACGCGCGCGCGCGCGCG	61 NW004444309.1_12	TGCCT	CCTTAAAGA	CGGTATCCC	TGAAGGAGGAGGAG	AGAAAG	GC			A ACGAGEG	CCGCCAGAGC	GTAGTTTTTCT	A A C T - A	AACTC
63. NW00444309.1_15 TGCCG-CCTTACAGTCGCTATCTCTGAAGGAGAAGGGAAAGC-AT-CTTTCTCCCATTCTAGTATACGACAG-CAGCACGCTGTAGTTTTGCTAACT-AAACC-C 64. NW00444309.1_16 TGCTG-CCTTACAGCCGGTATCCCTGAAGGAGGAGGAGGAGAGGC-GC-ATTTCTCCCATTCTGGTACAGGACAG-GGGCGCGCGCGCTTTTATAACT-AAACC- 65. NW004444309.1_17 TGCTG-CCTTACAGCCGGTATCGCTGAGGGAGAGAAAAAAGCC-GC-ATTTCTCCCATTCTGGCATACGGCGCGCGCGCGCGCTGTAGTTTTGTAAGT-AAACC- 65. NW004444309.1_17 TGCTG-CCTTACAGCCGGTATCGCTGAGGGAGAAAAAAGGC-GC-ATTTCTCCCATTCTGGCATACGGCCGCGCGCGCGCGCTGTAGTTTTGTAAGT-AAACC- 66. NW004444309.1_18 TGCTGCCTGAAGCCGCTGTACCGCTGAGGGAGAAAAAAAGGC-GC-ATTTCTCCCATTCTGGCATACGGCCGCGCGCGCGCGCGCG	62. NW004444309.1 14	TGCCA	CCTTACAGA	CACTATCACT	T G A A G G A G A A G	AGAAAGC -	AC	TCTCTCC	CTTTCAGGT	ATGTGACAG	AGGCAGCGC	GTAGTTTTGTT	AACC - A	AACAC
64. NW004444309.1_16 TGCTG-CCTTACAGACGGTATCCCTGAAGGGAGGAGGAGGAGGC-GC-ATTTCTCCCATTCTGCTACAGGACAG-GGGCGCGGCACTGTAGTAGTTTTATAACT-AAACC 65. NW004444309.1_17 TGCAG-CCTTACAGCCGGTATCGCTGAGGGAGAAAAAAGGC-GC-ATTTCTCCCATTCTGGCATACGACAG-CAGCACGATGTAGTTTTGTAAAGT-AAACAC 66. NW004444309.1_18 TGCAA-CCTTACAGCCGGTATCGCTGAGGGAGAAAAAAAGGC-GC-ATTTCTCCCATTCTGGCATACGACGCCAGCGGATGTAGTAGTTTTGTAAGT-AAACAC 66. NW004444309.1_18 TGCAA-CCTTACAGCCGGTGTCCCTGAGGGAGAAAAAAAGGCGA-GTTTCTCCCATTCTGGCATACGACGATGCAGGATGTAGTTTTGCCAAAGT-AAACAC	63. NW004444309.1_15	TGCCG	CCTTACAG	CGCTATCTCT	T G A A G G A G A A G	GGAAAAC -	AT	CTTTCTCC	TATTCTAGT	ATACGACAG	CAGCAACGC	GTAGTTTTGCT	AACT A	AACCC
	64. NW004444309.1_16	TGCTG	CTTACAGA	CGGTATCCCT	GAAAGAGGAG	AGAAAGC-	GC	ATTTCTCC		ACAGGACAG	GGGCGGCAC	GTAGTTTTATT	AACT - A	AACCC
	66. NW004444309.1_17	TGCAA	CCTTACAG	CTGTATCCCT	T G A G A G A G A A A	AAAAAAAAA	GA	GTTTCTCC		ATAGGATAG		GTATTTTGCT	AAGT - A	AACAC

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.

	1 10	20	30	40 50	60	70	80	90	100	110	120	130	140
Consensus	ATTCACTGCATGAAAGA	-GAAAG-	-CAAAATACAGTGT	-TCTCATTCAAGCTGG-	TTAACTTAAT	CAACTG	AAG-TT	TTCAGTTTTGAT	AGAANAAACAGTO	CTCACACCAC	CAGC-TGCTTT	-TCAAACAT-	-TCTCTCA
Identity				International Contraction of the International Contractional Cont	II-ma-la-								
						-							
1. BMBL01009774.1_1	ATTCACCACATGAAAGA	-GAAAG-		- TCTCC-TCAAGCTGC-	TTAACTGAAT		AAG-II	TICAGITIIG	AGAAGAAAG	GCCCAGCAC	AAGC-IGCIII	C-TCAALGT-	TCCCTGC
2. BMBL01009774.1_2	ATTCACCATCAAAGA	GAAAG		TCTTCAAGCTGC	TTAACTGAAT	CACTO	AAG	TTCAGTTTTC	AGAAGAAAGAGTC	CGUCAGCAG	AGC-TGCTTT		TCTCTGA
3. BMBL01009774.1_3	ATTCACEGCATGAAACA	CAAAG			TTAAGTGAAT	CLACIG	AAG TT		ACAACCAACCGIC				TETETCA
4. BIVIBL01009774.1_4	ATTEACCACATGAAAGA	- GAAAG			TTAACTAAAT	ACTG	AAG TT		AGAAGGAAG GTC		AAGC-TGCATT		TCTCTCA
6 RMPL01009774.1_5	ATTCACCACATGAAAGA	-GAAAG-	-CAAGATACAGEGT	-TCTCCTTCAAGCTGC	TTAACTGAAT	ACTG	- AG G TT	TTCGGTTTTCT	AGAAGAAAGGGTC	TGCCCAGCAC	AGC-TGCTTT	-TCAAACGT-	TCTCTCA
7 RMPL01009774.1_0	ATTCACCACATGGAAGA	-GAAAG-		-TCTCCTTCAAGCTGC	TTAACTGAAT	ACTG	AG-TT	TTCAGTTICTT	GGAAGAAAGTGTC	GGCCCAACAC	AAGC TGCTTT	C-TCCAACGT-	TCTCTCA
8 BMBL01002074.1_1	ATTCACCACATGAAAGA	-GAAAG-	-CAAAACACAGOGT	-CTTCCTTCAAGCTGC	TTAAGGGAAT	GAACTG	-AAG-TT	TTCAGTTTGGAT	AGAAGAAAGTTTC	TGCCCAGCAC	AGC-TGCTTT	C-TCAAATGT-	TCTCTCA
9 BMBL01002074.1_1	ATTCACTACATGAAAGG	-GAAAG-	-CAAAATACGCCGT	-TCTCCTTCAAGCTC	TTAACAGAAT	ACTG	-AAG-TT	TTCAGTTTTG	GGAAGAAAGAGTC	GCCCAACAC	AAGC-TGCTTT	C-TCAAGAGT-	TCTCTGA
10 BMBL01012666 1 1	ATTCACEGCGTGAAAGA	-GAAAG-	AAAAATACAGCAT	-TCTCCTTCAAGCTGC	TTAACAGAAG	ACTG	-AAG-TT	TTCAGGTATGT	AGAAGAAACAGTC	CTOCACAAA	CAGC-TGCTTT		TCTCTCA
11 BMBL01012666 1 2	ATTCACCACATGAAAGA	CATGG	GAAAATACAGCGT	-TCTCCTTCAAGCTGC	TTAACTGAAT	CTACT	-AAG-TT	TTCAGTTTTG	AGAGGAAAGAGTC	GCCCAGCA-	AAAC - TGCTTT	-TCCAATGT-	TCTCTGA
12 BMBL01099827.1.1	ATTCACTGCATGAAAGA	-GAAAG-	-CAAAATACAGTGT		TTAACTTAAT	CCACAG	-AAG-TT	TTCAGTTTTGGT	AGAACAACAGTO	CTCACACCAC	CAGC-TGCTTT	C-TCAAACGT-	TCTATCA
13 BMBI 01099827.1.2	ATTCCCAGTATGAAAGG	GAAGG	-CAAAATACAGTGT	TCACATTCAAGCTGG	TTAAGTTAAG	CAACT	-AAG-TT	TTCAGTATTGGT			CAGC TCATTT	C-TCAAACAT-	TCTCTCA
14 BMBI 01099827 1 3	ATTCAATGCAGGAAAGA	-GAAAG-	-CAAACTACAGTGT	-TCTCCTTCAAGCTGG-	TTAACTTAAC	CCACTG	AAG CT	TTCAGTTTTGGT	AGAATACACAGTO	ATAACACAAAC	CTCC-TGCTTT		TCTCTCA
15. BMBL01099827.1 4	ATTCACT CATGAAAGA	-GAAAG-	-CAAAATACAGCAT	TCTCCTTCAAGCTGC	TTCACTTCAT	CCACTG	-AAG-TT	TTCAGTTTTGGT	AGAAGAAACAGTO	GCACACCAC	CAGC-TGCTTT	C-TCAAAGTT-	-TCTCTCA
16. BMBL01099827.1 7	ATTCATTGCCTGAAAGA	-GAAAG-	-CAAAATACAGTGT	TCTCATTCAAGCTGG	TTAACTTAAT	CAACTG	-ACA-TT	TTCAGTTTTGGT			CAGTGTGCTTT	C-TCAAACAT-	TCTCTCA
17. BMBL01099827.1 8	TCACTGCATAAAAAA	GAAAAA	CAGAAT CAGTGT	GCTCATTCAAGCTGG	TTAACTTAAT	CAAC	TT	ATTAGTTTTGAG	-AAACACAGT	CTCCCACCAC	GCAAGATTT	C-TCAAACAT-	TCTCT
18. BMBL01099827.1 9	TTCACT CATGAGAGA	-GAAAG-	-CAGAGTACAGTGT	GTTCATTCAAGCTGG	TTAACTTAAT	CAAC	TT	ATCAGTTTTGAG	AAAACACAGTO	CTCCCACCAT	GCAAGCTCT	C-TCAT-	TCTCTCA
19. BMBL01099827.1	TTTACAGCATGAAAGA	GAAA	CAAAAATATATTGT	-TCTCATTCAAGTTG-	TTAACTTAAT	CAATTG A	ATAAG CT	ATCAGTTTTGAT	AAAACACTGT	CCACACAA	GAGAAAATTT	C-TCATACAT-	TCTCTGA
20. BMBL01099827.1	ATTCACTGCATGAAAG	GAAA	-CAAAACTGTTT	TCTCATTCTAGTTGA	TTAACTTAAT	GTGGTTGA	AAG-TT	ATC GTTTTGAT	AAA		CACTCT	TCCCATACGT-	TCTATGA
21. BMBL01099827.1	ATTCACTGCATGAAAGA	GAAA	-CACAATACAGT	CTTTCATTCTAGTTGA	TTATGTCAGA	GGGTA-	TT	ATC GTTTTGAT	GAAA CACTGT -		T	C-CCACACGT-	TCTCTCA
22. BMBL01099827.1	ATTCACTCATGAAAG	GAAAA	AAAATACAGT	TCTCATTTAGTTGA	TTAACTTAAT	AGT	AAG-TT	GTC GTTTTG T	AAACACTGT	CCACAGGAT	CAAAAACATTT	C-TTATATATAT-	-TCTCTGA
23. BMBL01099827.1	ATTCACCACATGGAAGG	GAAAC	-CAAAATACAGT	CTCTTGTTCTAGTTGA	TTAAGTCAGA	- AGGTG	TT	ATCAGT-TTGAT	AAACACTGT	CCACATGAC	CGAGAACCACT	T-TTATACCT-	-TCTCT <mark>G</mark> A
24. BMBL01099827.1	ATTCACTG	GAAA	CAAAATACAGT	CTCTCATCTGC	TTGATTAAGT	CACATG	GTATT	ATC GTATTGAT	AAACTCTGTT	CCTACATGAT	GGAGGATCACT	C-TTATACAT-	TTCTGA
25. BMBL01105564.1_1	ATTCAGTGCATGATA	GAAAG	CAAAATACAG	-TCTCCTTCAAGCTGG-	TTAACTTAAT	CCACTG	AAG-TT	TTCAGTTTTG	AGAA <mark>G</mark> AAAC <mark>T</mark> GTC	TG CCCACCAC	CAGC-TGCTTT	C-TCAACGGT-	TCTCTCA
26. BMBL01105564.1_2	ATACAGTGCCTGAAGGA	-GAAAG-	-CAAAATACACTGC	-TCTCGTGGAAGCTGG	TTAAC		GT	TTC T GTTTTG G T	AGAAGAAACAGTC	CTCACACCAC	CAGC-TGCTTC	C-TCAAACAT-	TCTCTCA
27. BMBL01043776.1_1	ATTCACTGCATGAAAGA	-GAAAG-	AAAATAGTGT	-TCTCATTCAAGTTGG-	TTAACTTAAT	CAA <mark>G</mark> TG <mark>GA</mark>	GTT	ATGAGTTTTGAT	AAAACACTCTC	CTCACCCAT	CCACAAGCTTT	C-TCAAACAT-	TCTCTCA
28. BMBL01043776.1_2	ACTCACTGCAAGAAAGA	-GAAAG-	-CAAAATACAGTGT	-TCCCATTCAAGTTGG-	TTAACATAAT	CAAGTAGA	AAC-TT	ATGAGGTTTGAT	AAAACACACTO	CTCACACCA	CACAAGCTTT	C-TCAAACAT-	TCATTCA
29. BMBL01043776.1_3	ATT TC CTGCA <mark>G</mark> GAAAGA	GAAAG	CAAAATACAATGT	-TCTCATTCAAGTTGG	CTAAGT	TCAGITCA	CAAG-TT	ATCAGTTTTGAT	AAAACACAGTO	CTCACACCA	CGAGAAGATTT	C-TCAAACAT-	TGTTCA
30. BMBL01043776.1_4	ATTCACCCCAGGAAAGA	-GGAAG-	-CAAAATACGCCGT	TCTCCGTCAAGCTGC	TTAACTGAAT	CTACGG	AAG-TT	TTCAGTTTTG	AGAAGAAACAGT	GCCGACCAG	AGC-TGCTTT	GCAAAGGC	TCGCTCA
31. BMBL01006510.1_1	ATTCACCGTATAAAAGA	-GAAAG-	CAAAATACAGCGT		TILACIGAAT	CACT			-GAAGAAAGAGTC		CAGC-TCTTG		CICICICA
32. BMBL01061436.1_1	ATTCAGTCTATCA	GAAGG			THETTAAT	GAACIG	AAC TT	TTCACTTTTG	AGAACAAGCAGTC	CIGICACCA	CAGC TACTT	G TGAAACAT-	TETETEA
33. BMBL01061436.1_2		GAAAG	AAAAATACCGTGT	TETEATTCAACCE	TATACTTAAT	GAACTG	AAG	TTCACTTICC	AGAACAAGCAGCC	CTGACACCAC	CAGE TACTTO	- TGAAACAC	TETCTCA
34. BIVIBLUTU01430.1_3		-CCAAC-			TTACTTAAT	CAACTG	AAG	CTCCCTTTTCCT	AGAGECAACOGIC	GIGACACCAC		C-TGAAAGAT-	TCTCTCA
36 BMBL01061436.1_4	ATTTCCTGCATGAAAGA	-GAAAG		TTTCATTTAGTAG	TTAGCTTAAT	CAACIG		ATAT GTTTTGAT		GCACGCCAT	GAAAAGCATT		TCTCT
37 BMBL01061436.1_5	ATTCACTGCATGAAAGA	GAAA	-CAAAATACAGGTT	TTTCATTTAGTAGT	TTAGCETAAT	CAAGTG	AAG-TT	ATCENTITICAT	AAAA CGCAGT	GCCACAAGAT	AGAATACCATT	C-TTACACAT-	TTTTGA
38 BMBL01061436 1 7	ATTCACAACATGAAAGA	-GAAAG-	-CAAAATGCAGTTT	TCTCATTCAGCTTGA	GTATCTCAGA	AGGAT	AA- CT	CACTGTGTTTGA	AAAGCCCTCTGTGTG	TACATGCAT	GAGAATGTIGC		TCTCTAC
39 BMBI 01082959 1 1	ATTCACTGCTGAAAGA	-GAAAG-	-CAAAACAGTGT	-TCTCTTTCAAGTTGG-	TTAACTTAAT		AAG-TT	ATCAGTGTT	AAAA-ACACAGT	CTGCACAAT	CACAAACTTT	C-CCATGCAT-	TCTCTAA
40 BMBI 01082959 1 2	ATTCACTGCATGAAAGA	-GGAAG-	-CAAAATACAGTGT	TCTCATTCAAGTTGG	TTAACTTAAT	CAATTG A	AAG-TT	ATCAGTTTTGAT	AAAACACTGT	CCACATGAT	CAATAAACTTT	-TCATA-AT-	TCTCTGA
41 BMBI 01041892 1 1	ATTCACTGCATGAAAGA	-GAAAG-	AAAATAGTGT	-TCTCATTCAAGTTGG-	TTAACTTAAT	CAAGTGGA	TGTT	ATGAGTTTTGAT	AAAACACTOTO	CTCACCCCAT	CACAAGCTTT	C-TCAAACAT-	TCTCTCA
42. BMBL01041892.1.2	ACTCACTGCAAGAAAGA	-GAAAG-	CAAAATACAGTGT	TCCCATTCAAGTTGG	TTAACATAAT	CAAGTAGA	AAC-TT	ATGAGGTTTGAT	AAAACACACTO		CACAAGCTTT	C-TCAAACAT-	TCATTCA
43. BMBL01041892.1 3	ATT TC CTGCA <mark>G</mark> GAAAGA	-GAAAG-	-CAAAATACAATGT	-TCTCATTCAAGTTGG	CTAAGT	TCAGTTCA	CAAG-TT	ATCAGTTTTGAT	AAACACAGTO		GAGAAGATTT	C-TCAAACAT-	TGTTCA
44. BMBL01041892.1 4	ATTCACCCCAGGAAAGA	-GGAAG-	-CAAAATACGGGGGG	TCTCCGTCAAGCTGC	TTAACTGAAT	CACGG	-AAG-TT	TTCAGTTTTG	AGAAGAAACAGT	TGCCGACCAG	AGC-TGCTTT	GCAAAGGC	-TCGCTCA
45. BMBL01022471.1 1	ATTCAGTGCATGAAAAAA	-GAAAG-	-CAATATGCAGTGT	TCTCTTCAAGCTGC	TTCACTTAAA	CAACT	AAG-TT	TTCAGTGTTGGT	AGAACAAAGCGTC		CGGC TGCTTT	C-TCCACCTT-	-TCT TG CA
46. BMBL01022471.1 2	ATT GAAA GCATGAAAGA	AAAG	CAAAACACAGTGT	TTCATTCAAGCTG	TTCACTTAAG	CAACT	AAG-TT	TTCCGTTTTGGT	AGAAAAACTGTC	CTCACACCAC	CAGC-TGCTTT	C-TCCAACAT-	TCTTCA
47. BMBL01022471.1_3	GTTGAAAACATGAAAGA	AAGAG	CAAAATACAGTGT	TCTCATTCAAGCTG	TTCACTTTAG	CAACT	- AAG-TT	TTCAGTTTTG <mark>G</mark> T	-GAACATACCGTC	CTCACACCAC	CAGC TGCTTT	C-TCCACCTT-	TCTTCA
48. BMBL01022471.1_4	ATTCAAAGCATGACAGA	AAAG	-CAAAATA-AGTGT	-TCTCATTCAAACTGC	TTCACTTAAG	CAACT	AAGTT	TTCAGTTTTG <mark>G</mark> T		CTCACACCAC	CAGC TGCTTT	C-TCCCATGT-	TCTTCA
49. BMBL01022471.1_5	ATT GAAA GCATGAAAGA	AAAG	-CAAAATACAGTGT	-TCTCATTCAAGCTG	TTCACTTAAG			CAGTTTTGGT	GGAATATACCGTC	CTCACACCAC	CAGC TGCTTT	C-TCCACCTT-	TCTTTCA
50. BMBL01022471.1 6	ATTCAATGAATGAAAAA	-GAAAT -	-CAAAATACAGTGT	TCACATTCAAGCTGG	TTAACTTAAT	CTACTT	-AAT-TT	ATTGT TTTGCAT	AAACACACTG	CTCACACCAT	TTATAAGCTTC	-TCCAACAT	TCTCTCA

Supplementary Figure 6. Alignment of 50 monomers manually isolated from the *Dugong dugon* assembled reference genome (accession GCA_015147995.1). Part one of six.

Consensus interiour MARCARGO CALUE CALUE CALUE CALUE CALUE CALUE AND CONSTRUCT CONSTRU		150	160	170	180	190	200	210	220	230	240	250	260	270	280	290
Udentity Destination Destination Destination Destination 1 DRIMED(1005774.12) CTGGGG MAGGETGGGAGGTGGAGGTGGGGG THE TRAATC CCGGG THE CAAGETGGGG GTGGGG THE TRAATC CCGGG THE CAAGETGGGG GTGGGG THE CAAGETGGG THE CAAGE CCGGG THE CAAGETGGGG GTGGGG THE CAAGETGGG THE CAAGE CCGGG THE CAAGETGGG GTGGG THE CAAGETGGG THE CAAGE CCGGG THE CAAGETGGG GTG THE CAAGETGGG THE CAAGETGG THE CAAGETGGG THE	Consensus	GTGGGATAAGA	CTGGCGAG	GTCTGCCTT	TCAAGTAY	TTGAATCC-	CCTGGTTTTCAAAG-		GGAACTTT	GCATTACAGGG-	AATGAAGGTA	CTTTCGCA-		AAAAA	AGGCAGATTGTT	T-GCC
BVBB0100977.11 OTGGOR AND/ETGGOR CONTROL OF THE ANALY GGOR AND/ETGGOR THE AND/ETGGOR AND/ETGGOR THE AND/ETGGOR AND/ETGGOR THE AND/ETGGOR AND/ETGGOR THE AND/ETGGOR	Identity			Lange The						the second second	Illerall_mas.					
																-
	1. BMBL01009774.1_1	GTGGGA <mark>G</mark> AAG	CTGGAGAGT	GCAGGCCTA	GCAAGTAT	TTGAATCC-	CCTGCTTTTGAAAG-		GGAATTTT	TC-CTACCAGG	AATGAAGGTA		TTCCAAG	CCAAAC	AGGTAGATTGTC	T-GCC
3 BMBL01002741 S GEGGAAMACTEGEGACTERCARIATECTARACCE CONSTITUCAMS CAACTEGACTARACCE AND CAACTEGAC CONSTITUCAMS CAACT	2. BMBL01009774.1_2	GTGGGA <mark>G</mark> AAG	CTGGAGAGT	GCAGGCCTA	GCAAGTAT	TTGAATCC	CCTGGTTTTCAAAG		GGAATTTT	IGCATACCAGGG	CATGAAGGTA	TATTGGCT	TTCAAAG	CCAAAC	AGGTAGATTGTC	T-GCC
4. BMB.01009774.1.4 GTGGGGAAGGTGGGGGGTGGGGGGGGGGGGGGGGGGGG	3. BMBL01009774.1_3	CTGGGAGAAG	CTGGAGAGT	GCTGCCTT	GCAAGTAT	TTGAATCC	GGTGGTTTTCCAAG-		GGAACTTT	IGCATATTAAGG	AATGAATGGA	TACTGGCT	CTCAGAG	CCAAAC	AGGCAGATTGCC	T-GCC
5. BMB.01009774.15 GREGGRAAGE MAGE GRAAGE TIGE AACHEET TRACKARE GRAACTICE CATEGORY TRACKARE GRAACTICE CATEGORY ACHIEVE THE CATEGORY ACHIEVE AND CHARACTER CATEGORY ACHIEVE THE CATEGORY ACHIEVE AND CHARACTER CATEGORY AC	4. BMBL01009774.1_4	GTGGGA <mark>G</mark> AAGA	CTGGAGAGT	GGGGGCTA	GCAATTAT	TTGAATCC-	CCTTTTTTTCCAAG-		GGAACTTT	FGCATACCAGGG-	AATGAAGGTA			CCAAAC	AGCAGATTGTT	- GCC
6, BMB.01009774, 1_0 GTGGG, BAAGATTCROW, GARTAGARGE CLIMAC, AND GTTRATCC, CLIMAGTTTRATCC, CLIMAGTTRATCC, LIRGANG, ANTONIGUE, LINGAG, CLIMAGT, ANTONIGUE, CLIMAGT, CL	5. BMBL01009774.1_5	GAGGGAGAAG	ATGGAGAGT	GCGGGCCTA	GCAAGTG	TTGAATAC-	CCTGGTTTACAAAG-		GGAACTTT	GCAT AT CAGGG	AATAAATGTA	GCTGGCT-	TTC <mark>AG</mark> AG	CTGAAC	AGGCAGATTGT	T-G T C
7. BMB01009274.1 CTGGGAT BAGE TOBERAGE TOBER	6. BMBL01009774.1_6	GTGGGAGAAGA	CTGGAGAGT	GAGGCCTA	GCAAGTG	TTGAATCC-	CCAGGTTTACAAAG-		GGAACTTT	IGCAT AGCAGGG	AATGAATGTA	TACTGG-T	TTCAGAG	CCAAAC	AGGCAGATTGT	T-GCC
B, BMB 100 02274,1 1 GEGG&AAAAGETGGGAAAGETGGGGCTGGEGGCTGGCAAGETGTTGCAAGE CTGGTTTCAAAG BBB 100 02266,1 2 GEGGGAAAGETGGGGGGGGGGGGGGGGGGGGGGGGGGGG	7. BMBL01009774.1_7	GTGGGATTAG	CTGCAGAGT	CCGGGCCTA	GCAAGGAT	TTAAATCC	CTTGGTTTTCAAAA-		GGAACTTT	IGCATACCAGGG	AATGAGGGGA	TACTGGCT	TTCAACG	CCAAAT	GGCAAATTGTC	T-GCC
9, BMBL01002274,1,2 GEGERARAAGE TIGERARAGE TIGERARAGE TIGERARAGE TIGERARAGE TIGERATE CALEGE TITERAAGE GEARATTIGCATECCAGE ANALGERARGE TIGERARGE TIGERARGE AND TITERAATEC GEARATTIGCATECCAGE AND AGAE TIGERARGE TIGERARGE AND AGAE TITERAATEC GEARATTIGCATECCAGE AND AGAE TIGERARGE AND AGAE TITERAATEC GEARATTIGCATECCAGE AND AGAE TIGERARGE AND AGAE TITERAATEC GEARATTIGCATECCAGE AND AGAE TIGERARGE TIGERTE JEBMBL0109827.1 JEBM	8. BMBL01002074.1_1	GTGGGAGAAG	CTGGAGAGT	IGCGGGCCTA	GCAAGTAT	TTGAATCC-	CCTGGTTTTCAAAG-		GGAACTTT		AATGAATGGA	STACTGGCT-	TTCAAAG	CCAAAC	AGGCAGATCGTC	T-GCC
	9. BMBL01002074.1_2	GTGGGAGAAG	CTGGAGAGT	GCTGGCCTA	GCAGGTAT	TTGAATCC	CTGGTTTTCAAAG-		GAACTTT	GCA-TCCAGGG	AATTCAGGTA		TTCAACG	CCAAAC	GGCAGAGT	T-GC
11. BMBL01012666.1_2 GRAGTTGCARGETGCARGETGCARGETGCARGETGCARGETGCARGETGCARGE GRAGTTGCATACCAGGCARGTAREGTGCARGETGCAR	10. BMBL01012666.1_1	GTGGGGTAAGA	CTGTAGAGT	GCTTGTCTT	GCAAGTA	TTGAAT C-	CCTGGTTTTCAAAG-		GGAAATTT	GCAT CCAGGG	AATGAAGGTG	GTTTTGGCA-	TTCAAAG	CAAAGC	AGGCAGATTGTT	T-GCC
12. BMBL0199827.] GTGGGATAAAACTGICGAGTGICTCCTTGCAGTATTGAATCC GTGGTTTCAAA BMBL0199827.] GTGGGATAAAACTGICGAGTGICTCCTTCCAGTATTGAATCC GTGGTTTCAAA GGAACTTGCATTACAGGGA ABCGGTGGCCTCTCCAGTGGTGTCCAGTGCTTTCAAA BMBL0199827.] GTGGGATAAAACTGICGAGTGICTCCTTCCAGTGTTTGAATCC GTGGTTTCAAA GGAACTTGCATTACAGGGA ABCGCTGGCCTCTCCAGTGTTTGCAGTGTTTGCAG BMBL0199827.] GTGGGATAAACTGICGAGTGICTCCTTCCAGTGTTTGAATCC GTGGTTTCCAAA BMBL0199827.] GTGGGATAGACTGGCGCTCTCCGATTGTTTCAAATGCGCGGGTTTCCAAA GGAACTTGCATTACAGGGA ABCGCTGCCCTCTCCGATTGTTTCGAATGCGCGGGATGCCCCCGGGTTTCCAAA GGAACTTGCATTACAGGGAAGGGCGCTCCCCTTCCAGTGCGCCCCGGGTTTCCAAAG GGAACTTGCATTACAGGGAAGGGCGCCCCCCGGGTTCCGCCTTTCCAAAG GGAACTTGCATTACAGGGAAGGGCGCCCCCCGGGTTCCGCCTTTCCAAATGCGGAGGCGCGCCGCGGATGCCCCCGGGTTTCCGACGGCGGACGGCGGCGCCCCCGGGATGCCCCCGGGATGCGCCCCGGGATGCGCCCCGGGATGCGCCCCGGGATGCGCCCCGGGATGCGCCCCGGGATGCGCCCCGGGATGCGCCCCCGGGATGCGCCCCGGGATGCGCCCCGGGATGCGCCCCGGGATGCGCCCCGGGATGCGCCCCGGGATGCGCCCCCGGGATGCGCCCGGGAGGGCGCCCCCCGGGATGCGCCCCGGGATGCGCCCCGGGATGCGCCCCGGGATGCGCCCCCGGGATGCGCCCCGGGATGCGCCCCGGGATGCGCCCCGGGATGCGCCCCGGGATGCGCCCGGGAGGGA	11. BMBL01012666.1_2	GTTGGACAAG	CTGGACGGT	GCAGGCCCA	GCAAGTAT	TTGAATCC-	CCTGGGTTTCTAAG-		GGAAGTTT	IGCATACCAGGG	AATGAAGGTA	GTTGGCT-	TTCAAAG	CCAGAC	AGGTAGATTGTG	T-GCC
13. BMBL0199827.1 Green LAARACTOGCANGTOTTCCCTTCANTATTENATCE CICCOMTTCANAGE GAAGE TACAGE GAAGETTAGE CAN GREEN CONTRACTOR CAN GREAT CONTR	12. BMBL01099827.1_1	GTGGGATAAGA	CTGTCGAGT	GTCTGCCTT	GCAAGTAC	TTGAATCC	CTGGTTGTCAAAG-		GGAACTTT	IGCATTACAGGG-	ACCGAAGGTA	TTTGCA-	TTCAAAG		AGACAGATTGTT	T-GCC
14. BMBL01099827.1 GTGGGATAGACTGGCGAGGTCTTCCTTGCAGGTTTTCAAGG GGACATTCCAGTGCAGGGTACCTTTGGCA TTCTAGGGAAACAGCGGCGAGTCTTCGCAGGTTTTCAAGG 15. BMBL01099827.1 GTGGGATAGACGTGGCGAGTGTTCCTGCCTTTCAAGGTTTTCAAGG GGACTTTCCATTCAGGGAAGGTACCTTTCGCAGGAGGTGTTTTCGCAGGAGGTACCTTTGGCA 16. BMBL01099827.1 GTGGGATAGACGTGGCGAGTGTTCCGCAGTTTCGAGGAGGTTTCGAGGGAGG	13. BMBL01099827.1_2	GTGGAATAAAA	CTGGCAAGT	GTCTGCCTT	TCAAGTAC	TTGAATCC	CTGCTTTTCAAA		GGAACTTT	CATTACAGGG	ACTAAAGGTA	GCTTT GCA-	TTCTAAG	CAAAAC	AGGCAGACTGTT	T-GCC
15. BMBL010998271 7 GeGGATEAGACTEGACABACTEGACTEGACAGATEGET (CATEAGAGET TECAAGET ALAGAGEGACTEGACAGETAGET CECACETTECAAGET ALAGAGEGACTEGACTEGACAGETAGET (CACAGETTECAAGET ALAGAGEGACTEGACTEGACAGETAGET) 15. BMBL01098271 P GEGGATEAGACTEGACAGETECTECACTECAAGETTECAAGET ALAGAGEGACTEGACTETTECAAGE GEGACTEGACTAGEAGEGACTEGACTETTECACAGET 18. BMBL01098271 P GEGGATEAGACTEGACGETECTECACTECAAGETTECAAGET GEGACTEGACGETAGECAGETGETTECAGET TECHAGE 19. BMBL01098271 P GEGGATEAGACTEGACGETECTECTECACGETACTEGACGESCECTETTECAAGE GEGACTEGACGETAGECAGEGACTETTECACE TECHAGE ALGACAGECAGETTETTEGACGACTEGACCETECTEGACGESCECTETTECAAGE 20. BMBL01098271 P GEGGATEAGACTEGACGETECTECTECACGESCECTETTECAAGE GEGACTEGACGETAGECAGEGACTETTEGACGACTEGACGESCECTETTEGACGACTEGACGETTEGACGESCECTETTEGACGACTEGACGETTEGACGESCECTETTEGACGACTEGACGETTECAGES GEGACTEGACGETAGECAGEGACGETECTEGACGETTECAGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTESTECAGETTECAGESCECTETEGACGESCECTETTEGACGESCECTETEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETEGACGESCECTETEGACGESCECTETTEGACGESCECTETTEGACGESCECTETCACGESCECTETTEGACGESCECTETTEGACGESCECTETCAGESCECTETCAGESCECTETCAGESCECTETCAGESCECTETTEGACGESCECTETCAGESCECTETTEGACGESCECTETCAGESCECTETTEGACGESCECTETCAGESCECTETTEGACGESCECTETTEGACGESCECTETCAGESCECTETCAGESCECTETTEGACGESCECTETTEGACGESCECTETCAGESCECTETTEGACGESCECTETCAGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECTETTEGACGESCECCTEGACGESC	14. BMBL01099827.1_3	GTGGGATAAGA	ACTG CGAG	GTCT	GCAAGTAT	TTGAATCC	ACTGCGTTTCAAAG-		GGAACA-T	IGCAGTACAGGG	GATGGAGGTA	CCTTT <mark>G</mark> GCA-	TTCTAAG	CAAAAC	AGGTGGATTGTT	T-TCC
 16. BMBL01099827.1 MTGGGATAGACTGCGGAGTGCTGCCTTTCAGTTTGGATC 17. BMBL01099827.1 MTGGGATAGACTGGGAGTGCTGCCTTTCGAGTAGTGGCTGGGCTTTCGAGTGGGGAGTGCTTGGATTGGATTGGATTGGATGGGAGGAGTGGTGGGGAGTGTTGGAGTGGGGAGTGGTG	15. BMBL01099827.1_4	GGGGGATGAGA	CTGGAGACA	GCTGCCTT	GCAAGTG	TTGAATC	CCTGGTTTTCAAAC		GGAACTTT	GCATAACAGGA	AATGAAGGTA	TTTAGCA-	CTCAAAG	CCAAAT	AGGCAGATTGTT	T-GCC
17. BMBL01099827.19 8. BMBL01099827.19 8. BMBL01099827.19 8. BMBL01099827.11 8. BMBL0109827.11 8. BMBL0109827.11 8. BMBL0109827.11 8. BMBL0109827.11 8. BMBL0109827.11 8. BMBL0109827.11 8. BGGANAAAGAA TGCAAGTHTCAAGTAGGGANGTHTCAAGTAGGANGTAGGANGTHTCAAGTAGGANGTHTC	16. BMBL01099827.1_7	ATGGGATAAGA	GTGGCAAGT	GTCTGCCTT	TCAATTAC	TTTAATTC-	CCTGGTTTTCAAAG-		GGAGCTTT	IGCATTACAACG	AATGAAGGTA	CTTTCGCA-	TTCTAGG	A	GAAC GATTGTT	TAACC
18. BMBL01099827.19 GGGAATAGGACTGGCUTTCTCAGGATTCGGCTTTCCAGGATTCGGCTTTCCAGGATTGGCTGCUTTCGGGATTGGCTGCCTTCGGGATTGGCGAGTGGCGGGATTGGGGAGGGGAGTGGCGGGATTGGGGAGGGA	17. BMBL01099827.1_8	ATGGGATA	CTGGCGAGT	CTCTCGTTT	TCCAGTAC	TTGATTCC-	CCTGCTTTTCAGAG-		GGAACTTT	IGCATTACAGGG-	AATGAAG	CTTT GCA-	TTCTCAG	AAGAAC	AAGCAGATTGTT	TCACC
19. BMB.01099827	18. BMBL01099827.1_9	GTGGAATAAGA	CTGGCTTGT	GTCTGCTT	TCCAGTA-	TTGAACTC-	CCTGCTTTTCAAAG-		GGAAGTTT	GCATTACAGGG	AAAGAAGGTA	SCTTTCACA-	TTCTA <mark>G</mark> G	AAGAAC	AAGCAGACTGTT	TGGCC
20. BMBL01099827.1 ATGGGAMAAGAATGGCAGGTUTCAGATAGGGAGCUTCGGGTTTCGAGGGACTTGGCTTGCAGGAACTTGGGAGGACGTGGCTTGCAGGAACTTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG	19. BMBL01099827.1	ATTGGAAATGA	ATGTCCAG	тстстстт	TCAAGAAC	TTGTATCCC	CCGCATTTTCAGATG	TTTTCTA	AAGGAGTTTT	TGAGTTATAGGG	AATGTGTGTG	ATTTCACA	TTCTA <mark>G</mark> G		ATGCAGACTGTT	TCCTG
21. BMLD1099827.1	20. BMBL01099827.1	ATGGGAAAAGA	ATGCACGT	TCTCTCTT	TCAAGTA	ATGAAGCC	TTGCGTATTTAAGTG	TTTTCTA	TGA GAACTT		AATGAATGTG	CTTTCACA-	TTCTA <mark>G</mark> G	AAGACCA	AAGCGGACTGTT	TCCCA
22. BMbL01099827.1	21. BMBL01099827.1	ATGAAAAAAAGA	ATGCAACT	TCACACTA	TCAAGTAC	GGGAGTC	CTGGGTTTTCAGATG	ΤΠΤΟΑΑ	IG GGAACTTT	GCGTTACCGGA	AATGAACGTG	CGTTCACA-	TTCTAGG	AAGACAA	AGGCAGATTGTT	CCAG
23. BMBL01099827.1 ATGGAAAGCTGAGATTECTGCTTTCAAGTAIGTAAGCCICTGAGTTTCAAGTTTCAAGTTTCAAGTTTCAAGTTTCAGGTTTCAAGTTTCAGGTTTCAAGTTTCGGTTATCAGG 24. BMBL0109564.1.1 STIGGATAGACTGGGAGTTCCTGCCTTGCAAGGATTTCAAGCCICTGGGTTTCAAG 25. BMBL0105564.1.1 STIGGATAGACTGGGAGTTCCTGCCTTGCAAGGATTTCAAGG CILAICTTCGCTTAACAGG AAGGAGGAACTTTGGGAT 26. BMBL01043776.1.1 GTGGGATAGACTGGGGATTGCTGCCTTCAAGGATTTCAAGG GAGAGTTTCCAAGG AGGAGGACGTTTGGGAT 27. BMBL01043776.1.3 GTGGGATAGACTGGGGATGCTCTGCCTTCAAGTATTGAAGTCCCCCGGGGTTTCAAGG GGGAGCTGCTGTGGATTGCGCCACCCCCCCCCC	22. BMBL01099827.1	CTGGGAAAAGA	ATGCAAGT	TCTTCCTT	TCAACTAC	GTGATGCT	CTGGGTTTTCAGATG	TITICIA	TG GGAACT C T	GTGTTACCAGA	AATTAATGTG	CTTTAAAAA	CTCTAGG	AAGAAGA	AGGTATGTTGTT	TCCTG
24. BMBL01099827.1 ALTERARATARTGIGAGATIGUCAGGUTUCTURACTICURACTIGGATIGGT TICGGT 25. BMBL010105564.1.2 ATGGGATAGACTGGCGATGTCTGCCTTCAGACGUTUGGCTCCCGGTTTTCAAGG 26. BMBL01043776.1.2 GTGGGATAGACTGGCGATGTCTGCCTTCAGTGTTGAATCCCCCGGTTTTCAAGG 27. BMBL01043776.1.2 GTGGGATAGACGGGGAGGTGTCTGCCTTCAGTGTTGAATCCCCCGGGTTTCCAGGG 29. BMBL01043776.1.2 GTGGGATAGACGGGGAGGGCTGCCGCCTTCAGTGTAGTGAATCCCCCGGGTTTCCAGGG 29. BMBL01043776.1.2 GTGGGATAGACGGGGAGGGCTGCCGCCTTCAGTGATGGAATCCCCCGGGTTTCCAGGG 29. BMBL01043776.1.2 GTGGGATAGACGGGGAGGGCTGCCGCCTTCAGTGATGGAATCCCCCGGGTTTCCAGGG 29. BMBL01043776.1.2 GTGGGATAGACGGGGAGGGCTGCCGCCTTGCAGTGATGTGAATCCCCCGGGTTTCCAGGG 29. BMBL01043776.1.2 GTGGGATAGACGGGGGGCGCCGCCCTGCAGGGTTTGAATCCCCCGGGTTTCCAGGG 29. BMBL01043776.1.2 GTGGGATAGACGGGGGGCGCCGCCCTGCAGGGTTTGAATCCCCCGGGTTTCCAAGG 29. BMBL01043776.1.4 GTGGGGATAGCGGGGGGGCGCCGCCCTGCAGGGTGTCGGCCTTCCAGGGGATGCGGCCGCCGCGGGGGGGG	23. BMBL01099827.1	ATGGGAAAAGA	ATGTCAAGT	TCTCCTTT	TCAAGTA	GTGAAGCC	CTGAGTTTTCAAAT	TTTTCAA	TG GAAACTTT	IGGGTTATGGAG	AATGAACATG	TTTTCACA-	TTCTAGG		AGGCAGATTGTT	TCCTG
25. BMBL01105564.1_G GTGGGATAAGACIGGAGATIGECIGCCITTCAGATA GTGGATAAGACIGGAGATIGECIGCCITTCAGATC CCTGGTTACCGGA TTCAAGAAGAAGACGAGATGCIGGAGATGCITGCCCTTCAAGTACITGAATCC 27. BMBL01043776.1_G GTGGGATAAGCIGGAGAGTGCIGCCCTTCAAGTACITGAATCC GCGCGTTCIGGTTACCGGA AAGGAGAGATCCITTCAAGAGAGAGAGAGATCTTTCAAGAG GGGGATAAGTGCGAGAGATCGICCGCCAGGATTGTTTCAAGTCGCCCGGGTTTCAAGGGGGGGG	24. BMBL01099827.1	ATTGAAAATA	ATGTCAAGT	TCTGACTT	TCAGTAT	GTGAGTGT	TTGGTTTTCAGG				AATGAATGTG	CTTTCATA-	GTCTATG	AAGAACA	AGG <mark>G</mark> AGATTGTT	TCCCG
26. BMBL01105564.1_2 all GGGATAAGAC TGGCATGCTTCAAGGAALTCAAGGCAGTTTCAAGG GLAACTTGCATTACAGGA AAGGAAGGTACTTCAAGAGCAGTTTCAAGG 27. BMBL01043776.1_2 GTGGGATAAGATGGTGGCTTTCAAGTACTTCAACGG GGACTTCGTGTACGGCAAGGAAGGCAGTTTTCCAAGGAAGG	25. BMBL01105564.1_1	GTGGGATAAGA	CTGGAGATI	GCTGCCTT	GCAAGCA	TTGAATCC	CCTGGTTATCAAAG			GCCTAACAGGG	AACGAAGGAA	TTTGGGA	TTCAAAG	TC AAAG	AGGTAGATTGTT	T-GC
27. BMBL01043776.1 GTIGGANACARLAGARGTIGTCTCLTCLTTCAAGTACTTGAACCGCCGGGTTTCCAAGTG GGGACT	26. BMBL01105564.1_2	ATGGGATAAGA	CTGGCGAGT	GICIGCCII	TCAGGTAC	TIGAATCC-	CCTGCTTTTCAAAG		GCAACTTT	GCATTACAGGG	AAGGAAGGTA	ACATTCGCA-	TTCTAAG	AAAAAC	AGGCAGATTGTT	T-GCC
28. BMBL01043776.1 2 GIGGATAAAACAGGAGATTAATLCGCCGGATTAALCGCCAGGITTICAAGG CCGGAGCACAGGAGCACAGGAGCACAGGAGCACAGGAGCACAGGAGCACAGGAGCACAGGAGCACAGGAGCAGAGGAG	27. BMBL01043776.1_1	GIIGGAA	- IGALGAGA		ICAAGIAC	TIGAATCO	CAAGTTTTCAGAG-		GGAGCICI	GIGITACCCGG	AATGAAGGTA	LCTTTCACA-	II III IAGG	AAGAACA	AGGCAGATIGIT	
29. BMBL01043776.1_3 GIGGANALALICIT GUGAGIULCI INGCITTICAAGIGATICT CAAGG GIGGANALACARGUAGGIULCI INGCITICAAGIGATICT CAAGG GIGGANALACARGUAGGIULCI INGCITICAAGIATICT GAATLCC CCTGGTTTT CAAAG GIGGANALACARGUAGGIULCI INGCITICAAGIATICT GAATLCC CCTGGTTTT CAAAG GIGGANALACARGUAGGIULCI INGCITICAAGIATICT GAATLCT CCTGGTTTT CAAAG GIGGANALACARGUAGGIULCI INGCITICAAGIATICT GAATLCT CCTGGTTTT CAAAG GGAAACTTTGCATAGACTGACAGUAGGUAGGIULCI INGCITICAAGIATICT GAATLCT CCTGGTTTTCAAAGIATICT GAAGGUAGGUAGGUAGGUAGGUAGGUAGGUAGGUAGGUA	28. BMBL01043776.1_2	GIGGGAIAAGA	AIGGIGAAI	GIIGCCCTT	ICAAGIAC	TIGAATCCG	CCAGGIIIICAGAIG	TICIGIA	ACGGAGCIGI	GIGITAGCCIG	AAIGCAGGIA	LIIICCGCA-	TICCAGG	AAGGACA	AGGCAGATIGIT	ICCAG
30. BMBL0106137/6.14 GIGGABAAALACAGGAGIGGCIGCIGCIGCIGCIGCIGCIGCIGCIGCIGCIGCIG	29. BMBL01043776.1_3	GIGGGAIAA	CIGG GAGI	GICIEGCII	TCAAGIGA	TIGAATCC-	CCTGGTTTTCAAGG-			TAACAGAG	AATGAAAGTA	CCTT GCA-	TIN -AGG	AAAAACA	AGIIAGATICTI	GCC
31. BMBL0100510.1_1 GEGGAGAAGACEGGAGAAGACEGGAGAGETERGEC GEGAGAAGACETEGGACETERGACACEGGAGAGETERGEC CECENTEGGACE 32. BMBL01061436.1_2 ATGGGATAAGACEGGCGAGTERTECACCT GEGAGACAGEGGATTEGCT GGAAGACTEGGACAGEGGATAGCACEGCCAGETERTEGGACE TIGAAGE GAGAACAGECGAGTERTEGCC 33. BMBL01061436.1_2 ATGGGATAAGACEGGCGAGTERTECACTTCCAGTAGETGGATCC CCTGGTTTTCCAAT GGAACTTEGCTTACCAGEG ATGGGATAAGCATGGCGACTGCCGCGGGATTEGTCT GCCAACTTEGCTTACCAGEG AAGGAAGGATAGCT TICTAAGE GAAAAAGCGGCGAGTTEGTT GCCAACTTEGCTTACCAGEG AAGGAAGCATACCGCCAGTTEGTT GCCAACTTEGCTTTCCAGTGGCGATAGCT TICTAAGE GAAAAAGCGGCGAGATTEGTT GCCAACTTEGCTTCCCACTGCGGGAAAGCGCGAGATTEGTT GCCAACTTEGCTTCCAGEG AAGGAAGCATACCGCCAGTTEGTT GCCAACTTEGCTTTCCCACTGCGGGGAAAACAGCGCGAGATTEGTT GCCAACTTEGCTTCCCACGGCGGAAAACAGCGCGAGATTEGTT GCCAACTTEGCTTCCCACGGGGAAAACAGCGCGAGATGGTTEGCCATTGGGGAAACAGCGGGAAAACAGCGCGAAAGGGCGGAAAAGAGGCGAATTGGCCCCTGGGGGTTTCCAGATGGTTCCCACGGGAAAACAGCGCGCTTCGGGAAAAAAAGGCGGACTTGGCCTTCCAGGGAAAAAAAGGCGAAAAGGCGAAAAGGCGAAATGGCGGTTTCCAGAGGGAACTTGCGCTGGGGAAAAAAAGGCGAGATGTTCCCCCGGGAAAAAAAGGCGGACTTGCCCTGGGAAAAAAGGCGGACTTGCGCTTCCAGGGAAACAAGGCGGAATGTTCCCCCGGGGATTCCAGAGGGAACTTGCCCCGGGAAACAAGGGGGGTTCCCCCTTCCAAGGGAAAGGGCGGATTGGACGGGAACAAGGGGGGGTTTCCACGAGGGAACTTGCCCCGGGGAAACAAGGCGAGGTTGTCCCCCGGGGTTTCCAGAGGGAACTTGCCGCGGAAAGGGCGGGTTTCCAGAGGGAACGGGGGTTTCCACGGGAAAGGGGGGGG	30. BMBL01043776.1_4	GIGGGAGAA	CAAGAGAGI	GCTGCCTT	GCAAGTA	TIGAATCC-	CCTGGTTATCAAAG-			IGCUTAACAGGG	AAGGAAGGAA	TTTGGGA-	TTCAAAG	CAAAG	AGGAGAGATIGTT	I-GCI
32. BMBL01061436.1_1 at GGG TAGACG CONTROLOG TO CONTROL AND GGATACG ALCONTROL AND GGAAACG ALCONTROL AND GGAACG ALCONTROL AND G	31. BMBL01006510.1_1	GTGGGAGAGAAGA	CIGGAGAGI	GICIGCCI	GGAAGCA	TIGAATCI -	CCTGATTTTCAAAG-		GGAACTTT	IGCATAACAGGG-	AATG-AGTTA	LITEIGGCA-	TTCAAAG	GAAAC	AGGCAGATIGIT	GGCC
33. BMBL01061436.1_2 All GGAATAGACTGACGTTCACCTTCACTTGACTTGAATCC CCTGGTTTCAAAC GGAACTTGCGTAATAGCAGAACAGTGACTGACGTTTCACATTGAATCC CCTGGTTTCAAAC GGAACTTGCGTAATAGCAGTACTGCGCATTGCCTTTCCGAGTACTGAATAGCCCCTGGTTTCAAAC GGAACTTGCGTAATAGCAGTACTGCGCAATGCGCAGTGCCTTCCGCAGGCAG	32. BIVIBLU1061436.1_1	ATGGG-TGTAA	CTG CTACT	TCT CCTT	TCAAGTAC	TTGGATCC	CTCCTTTTCAAAT		GGAAGTT	TAATTATAGAG	ANTGAAGGAA	CTTTCCCA	TTGAAAG	GAGAAC	AGECAGATIGIT	T GCC
34. BMBL01061436.1_3 35. BMBL01061436.1_4 36. BMBL01061436.1_5 37. BMBL01061436.1_6 37. BMBL01061436.1_6 37. BMBL01061436.1_6 37. BMBL01061436.1_6 37. BMBL01061436.1_6 37. BMBL01061436.1_6 37. BMBL01061436.1_6 38. BMBL01061436.1_7 39. BMBL01082959.1_1 31. GGAAAAGAATATGCCATTCCCTTTCCAATGCCTTCCAGACGTTTCAAGACGACGTTCCAGACGTTCCAGGTTTCCAAGGTACCTCCGAGGTTTCCAAGGTACCTTCGAGGTTTCCAAGGTACCTTCGAGGTTTCCAAGGTACCTTCGAGGTTTCCAAGGTACCTTCGAGGTTTCCAAGGTACCTTCGAGGTTTCCAAGGTACCTTCGAGGTTTCCAAGGTACCTTCGAGGTTTCCAAGGTACCTGCGGTTTCCAAGGTACCTTCGAGGTTTCCAAGGTACCTTCGAGGTTTCCAAGGTACCTTCGAGGTTTCCAAGGTACCTTCGAGGTTTCCAAGGTACCTTCGAGGTTTCCAAGGTACCTTCGAGGTTTCCAAGGTACCTTCGAGGTTTCCAAGGTACCTTCGAGGTTTCCAAGGTACCTTCGAGGTACCTTCGAGGTACCTTCGAGGTAGGT	33. BMBL01061436.1_2	ATGGGATAAGA	CTGACCAGI	TCTACCTT	TCAAGTAC	TTGAGTCC	CIGGITTICAAA		GGAACTTT	IGCOTTACA GC	AAAAAGATA	CTTTCGCA-	TTGTAAG	GAAAA	AGGCAGATIGIT	TACC
35. BMBL01061436.1_6 IndegAtaAcaTeGCAACATEGTCACCTTTCCGTTTCCGTTTCCGGTTTCCAACATEATEGCATTTCCGAACATEATEGCAACAACATEATEGCAACAACATEATEGCAACAACATEATEGCAACAACAACAACAACAACAACAACAACAACAACAACAA	34. BMBL01061436.1_3	ATGGGATAAGA	CTGGGGAG		TCLAGTAC	TTGAATCC-	CCTCCTTTTCAAA		GGAACTTT		AATCAAGGTA		TTAAAG	GAAAA	AGGCAGATIGIT	T
30. BMBL01001435.1_5	35. BMBL01061430.1_4				TCHAGTAC	TTGAATA		CITICIA	GAACTT		AATGAACAT	CTTTC CA-	TTCTACC	AMAMAAC	AGCAGATIGIT	TECCA
37. DWDL010014351_0 artigeActivation concentration of the addition of the additi	30. DIVIDLU1001430.1_3		ATGACAG	TCTCCCTT	TEGGIA	TTGAAGCC		TTTTCTA	GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG		AATGAATGTG	CTTTCAAA	TTCTAGG		AGGCTAATTGTT	TCCCA
 38. BMBL01082959.11 39. BMBL01082959.12 31. GGAAAAGAATATICLACGTTCTACCTTTCAAGTACTTGAATCCCCGAGTTTTCAGAGGAACTTGCGGAAAGAACAAGGCGGACTTTCACAGGAACTTGCGGAAAGAACAAGGCAGACAAGGCAGATGTCTCCCTTTCAAGTACTTGAATCCCCGAGGTTTTCAGAGGAACTTGCGGAAAGAACGAGGGGGACTTTCAGAGGAACAAGGCAGATGTGTCCCCAGGAAGAACAAGGCAGATGTGTCCCCAGGTTTCAGAGGAACAAGGCAGATGTCTCCCTTTCAAGTACTTGAAGTCCCCCGAGGTTTTCAGAGGAACAAGGACGAGGTGCGGCCGCCTTCCAGGAGGAACAAGGCAGATGGTGCCGCCTTCCAGGTGCGGAGACTGCCGCGAGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCG	37. BIVIBLU1001430.1_0		AIGACAG		TCAARCAC	CTGAAGCC		CTTTCAA			CATGAACGTG	CTTACACA	ACANTETETA			TELCC
39. BMBL0101082959.1_1 altigeAdaAcataticCateConstraint Constraint Constrain	30. BIVIBLU1001430.1_/	ATGGAAAAAGA	ATATCOACC	TCTACCTT	TCAAGTAC	TCAAACCC	CONSTITUTO	TTTCCTA	ACTU	CATTAC CCC	ATTAAAGGTG	CTTTCACA-	CTCTACC		AGGCAGATTGTT	TCCAG
40. BMBL01041892.1_2 41. BMBL01041892.1_3 42. BMBL01041892.1_3 43. BMBL01041892.1_3 44. BMBL01041892.1_3 45. BMBL01041892.1_4 45. BMBL01022471.1_1 45. BMBL01022471.1_4 45. BMBL0	40 PMPL 01082959.1_1		ATCICAC	TCTACCTT	TCAAGTAC	TTGAATCO		TTTTCTA	ACCOALCTTT		AATGAATGTG	TTTCACA	TTCTAGG		AGOCAGATTGTT	TCCCA
41. BMBL01041892.1_1 STGGGATAGATGGTGAAGATGGTGGAAGATGTTTTTTTTT	40. DIVIDLU1002959.1_2	GTEGGARAGA	TGATGAG	TCTCCTT	TCAAGTAC	TTGAATCO			GGAGCIT	GTGTTACCCGG	AATGAAGGTA	CTTTCACA	TTTTAGG		AGGCAGATTGTT	TCACT
42. BMBL0104392.1_2 GTGGGATAATTCTGGTGAGTCTTTGATTGATTGATTGATCC GTGGATATTTGCGCAGATAGTAGGAGTGGAGTGTTTGAATGC 43. BMBL01041892.1_3 GTGGGATAATTCTGGTGAGTCTTTGATGATTGATCC GTGGATTTTTGAGG GGGATTTAACAAGATAGGAGTAGTACCTTTGCAAGTAGTTGAATGA 44. BMBL01041892.1_4 GTGGGAGAATACAAGGAGGGAGTGCCTGCCTTGCAAGGATTGATCC CCTGGTTTTCAAGG GTGGGAGAATACAAGGAAGGAAGTAGTGGGGAGGTGCGGCTGCCTTGCAAGGATGATTGAT	41. DIVIDLU1041092.1_1	GTGGGATAAGA	ATGGTGAAT	GTEGECCTT	TCAAGTAC	TTGAATCO	CAGGTTTTCAGAG	TTOTOTA	ACGGAGCTGT	GTGTTAGCCTG	AATGAAGGTA	TTTCCCCA-			AGGCAGATTGTT	TCCAG
43. BMBL01041892.1-3 GTGGGAGAATACCAGGGAAGGTGCTGCCTTGCAAGGAGTTGCAAGCCCTGCAGTATTTGAATCC 44. BMBL01022471.1-1 45. BMBL01022471.1-2 46. BMBL01022471.1-2 46. BMBL01022471.1-2 47. BMBL01022471.1-3 AGGAATAAGCAGGGCGAGTGTCTGCCTTTCAAGGAATTGCAATCC CCTGGTTTTCAAGG 48. BMBL01022471.1-3 AGGGATAAGCAGGGCGAGTGTCTGCCTTTCAAGGACTTGCAAGCCGCGGTTTTCAAGG 48. BMBL01022471.1-3 AGGGATAAGCAGGGCGAGTGTCTGCCTTTCAAGGACTTGCAAGCCGCGGTTTTCAAAG 49. BMBL01022471.1-3 AGGGATAAGCAGGGCGAGTGTCTGCCTTTCAAGGACTTGCAAGCCGCGGTTTTCAAAG 49. BMBL01022471.1-3 AGGGATAAGCAGGCGGGGGGGCGGGCGGGGGTGTGGCCTTTCCAAGGACTTGCAGTTTTCAAGC 48. BMBL01022471.1-4 ATGGGATAAGCAGGCGGAGTGTCTGCCTTTCCAAGGACTTGCAGTTTTCAAGC 49. BMBL01022471.1-4 ATGGGATAAGCAGGCGGGGGGGGGGGGGGGGGGGGGGGG	42. DIVIDLUTU41052.1_2	GTGGGATAAGA	CTGGTGAGT	TCTECTT	TCAAGTCA	TTGAATCC	CCTGGTTTTCAAGG-				AATGAAGTA	COTTOCCA-	TTT-AGG		AGUTAGATTOTT	TECC
45. BMBL01022471.1_1 ATGGAAGAATACTGGCGAATGGTCGCCTTTCAAGGATTGCATCC CCTGGTTTTCAAAG GAATGCAGGTAGGAAGGTAAGATTGCAGTCC-TTCGTG AAAAACAGGCAAGGCAATGAATGCAGTCGCCTTTCAAGGAATGCAGTCGCCTTTCAAGGCAATGCCCTTGCAGTCGCCTTTCAAGGCAATGCCCTTGCAGTCGCCTTTCAAGGCAATGCCCTTGCAGTCGCCTTTCAAGGCAATGCCCTTGCAGTCGCCGAGTGCGGCAATGAAGGTAGGCAGTGCGGCAATGAAGGTAGGCAGTGCGGCAGTGCGGCAGTGCGGCAGTGCGGCAGTGCGGCAGTGCGGCAGTGCGGCAGTGCGGCAGTGCGGCAGTGCGGCAGTGCGGCAGTGCGGCGAGTGCGGCGGGGGGGG	43. DIVIDLU1041052.1_3	GTGGGAGAATA	CAAGAGAG	GCTGCCTT	GCAAGTAT	TTGAATCC	CCTGGTTATCAAG-			IGCT TAACAGGG	AACGAAGCAA	TTTGGCA-	TTCAAAG	TCAAAG	AGGTAGATTGTT	T-GC
46. BMBL01022471.1_1 GGATTAGAGGGGAGGGGGGGGGGGGGGGGGGGGGGGGG	44. DIVIDEU 1041032.1_4	ATGGAAGAATA	CTGGCGA	GICTGCCTT	TCAAGAGA	TTGCATCC	CCTGGTTTTCAAAG-		GGAACTTT	GCAGTACATGG	AATGAAGGTA	GATT GCA-	TTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC	- 44444	AGCAGATI	T-GCC
47. BMBL01022471.1_3 AAGGAATAAGACAGGCGAGTGTCTGCCTTTCAAGTACTTGAATGC ACTGGTTTTCAAGAGTTCTATGAAAAGCAGGCGGG-AATGAAGGTAAGGT	45. BIVIDLUTU22471.1_1	CTGGGATAAGA	CAGGCGAGT	GTCTGCCTT	TCATGTAN	TTGAATCC	CCTGGTTTTCAAAG-		GGAACTTT	GCTGTACATGG	AATGAAGGTA	CATT GCA-	TTCTG	AAAAAA	ACGCAGATGATT	T-GTC
48. BMBL01022471.1_4 ATGGGATAAGACTGGCGAGTGCCTGCCTTTCTAGTACCTCGAGTACCTCGAGTACCTGCGCTTTCCAAGAC TTGCAGTACCAGGGAAGGTAAGACGAGTGCGAGTGCGCGAGTGCCGCCTTCCTAGGAAAAAAAA	40. DIVIDEO 10224/1.1_2	AAGGAATAAGA	CAGGCGAGT	GTCTGCCTT	TCAAGTA	TTGAATAC	ACTGGTTTTCAAAG-		GGAGCACT	GCAGTATGGGG	AATGAAGGTA	CGTTAGCA-			ACCAGATTATT	T-GCC
	47. BIVIDLUTU22471.1_3	ATGGGATAAGA	CTGGCGAGT	GTCTGCCTT	TCTAGTAC	TTGAATCC	TGGTTTTCAAAM-		GGAACTTT	GCAGTACAGGG	AAGGAAGGTA	CATT GCA-	TTCTATG	- 444444	ATGCAGATTGTT	T-GCC
	49 BMBL01022471.1_4	ACGGAATAAGA	CTGGCGAGT	GTCTGCCTT	TCAAGTA	TTGAATAC	TGGTTTTCAAAG-		GGAACTTT	CAGTACATGG	AATGAAGGTA	ACATT GCC	TTCTATG	AAAAA	ACCAGATTGTT	T-GCC
	50 BMBI 01022471.1_5	CTGGGATAAGA	CTGG-GAGT	CAATICCTA	TGAAGTAT	TTGAATCC-	-CCAGCTTTAAAAG-		GGAACTTT	GCGTTACAGGG	AATGAAGGTA	CTTTCACA-	TTCCAGG	AAGAGC	AGCAGATTETT	TOGCC

Supplementary Figure 6. Alignment of 50 monomers manually isolated from the *Dugong dugon* assembled reference genome (accession GCA_015147995.1). Part two of six.

	300	310	320	330	340	350	360	370	380	390	400	410	420	430	
Consensus	TTCAGCTAGAGGA	GG-CAGA	AGCGGTTCAG	TTTTGCTGACT	AAACCTTGT	GTGGTTAAC	ACGAGC	ATGTWTCCTAT	TTCCCTGCAA	GAAAGAGAAAA	CAAAAGACA	GTTGCTCTCAG	TCTGGGGG	ATGAAGTGA	<mark>G</mark>
Identity														In the second	-
								_							
1. BMBL01009774.1_1	TTCAGCTAGAGGA	ACAG-CAGA	AGAGGTTCAG	TTTTGCTGACT	AAAGCTTGT	TGCTTAAC	GGAGC		TTCCCTGC	GAAAGAGAAA	GCAAGAGCH	GTTGCTCTCAG	TCTGGGG-	ATGAACGGA	G
2. BMBL01009774.1_2	- TTCAGCTAGAGGA	AGG-CAG	AGAGGTICAG	GITIGCIGACI	AAAGCTIGT	GIGCITAACA	AGGAGC	ATGTAGC	TICCCIGC	GAAAGAGAAA	GCAAGAGCII	GATGITCTCAG	TCTCG-GG	ACGAAGGGA	G
3. BMBL01009774.1_3	TTCAGCTAGAGGA	AGG-CAGA	AGCGGTTCAG	STITIGCIGACI	AAACCGIGI	GIGCITAAC	ACGAGC	AIGIAGC		GAAAGAGGAA	CAAGAGCH	TIGCICICAG	TCTGG-GG	AGGAAGGGA	G
4. BMBL01009774.1_4	TTC GCTAGAGGA	ACAG-CAGA	AGAGGTICAG	TTTTGCTGACT	AAACCIIGI	GTGGTTAA	AGGATC	GIGIAICGICI	TTCCCTGCA	GAAAG	GCAAGATCH	GTTGCTCTCAG	TCTGGGTA	GGGA	G
5. BMBL01009774.1_5	TICAGCIAGAGGA	ACAG-CAGA	AGAGACICAG	STITIGCIGACI	AAACTIGI	GIGCITAACA	AC AGC			GAAAG	CAAGAGCI	GIIGCICICAG	TCTGG-GG	ACGAAGGGA	G
6. BMBL01009774.1_6		AGG CIGA	ATACGTICAG	GTTTCCTCAG	AAACCIC	GITCTTAAC	AGGAGG		TICCCIG	GACAGAGAAA	GCAAGACCH	GTTGCTATGAC	TCTGG-GG	ATGAAGGGA	G
7. BMBL01009774.1_7	TTCAGETAGAGGA	ACGG-CAGA	AGAGGTTCAG	GITTGCTGAC	AAACCGIGI	CTGCTTAAC/	AGGAGC	ATGTAGC		GAAAGAGAAAA	GAGAGE	GUIGCICICAG	TCTG -GG	ACAAAGGGA	G
8. BIVIBL01002074.1_1		CGG CAGA	AGAGGTTCA	TTTTGTTGACT	AAACCGICI	GTGCTGAAC	ACGAGC	ATGTAC		GAAAGAGAAAA		GTTGCTCTCAG	GATGG-GG	ATCAACCOT	G
9. BMBL01002074.1_2		CGG CAGA	AGAGGTTCAG	TTTCCCCCAN	AAACCITGT	GTGCTTAAC	ACGAGE	AIGIAGC		GANGAGAA	GCAACI GCH	GTTGCTCTCAG	TCTCC-CC	ATGAAGGCT	G
10. BMBL01012666.1_1			GETTCAG	ATTTCCTCACT	AAACCITGT	GTOCTAAC	ACCAGE	ATGTAG	TTCCCTCCT	GAAGGAGAAA		GTTGCTCTCAG	TCTGG-GG	ATGACACCA	G
12 PMPL 01000927 1 1		CAA AGA	AGAGGTTCAG	TTTTCCTGACT	AAAGCTTGT	GTOCTOAAC	ACGAGC	ATGTAGATTAT	TTCCTGCC	GAAAGAGAAAA		GTEGETETEAG	TCTGGGGGG	ATGAGAGGA	G
12. DIVIDLU1099627.1_1	TTCAGCTAGAGGA		AGCAGTICA	TTTTGCTGACT	AAAG	GTOCTTAAC			TTCCTTCCAA	GAAAGAGAAAA		GATGCTCTCAG	TCTGGGGG	ATMAAGTGA	G
13. DIVIDLUTU99627.1_2		CAG-CAGA	AGCAGTTCAG	TTTTGCTGAG	AAACCTTGT	GTOCTTAAC	ACGAGC	AGTATCCTAT	TTCCCTGCAA	GATAGAGAAA	CAAAAGACA	GTTGGTCTCAG	TCTGGGGG	ATGAAGTGA	G
15 PMPL010099627.1_3	TTCAGCTAGAGGA	GG-CAGA	AGAGGTTCAG	TTTTGCTGACT	AAACCAGT	GTOCTTAAC	AGACC	- ATGTAGUTCI	I GCCATGCCA	GAAAGAGGAAA	CAAAAGCCA	GTTGCTCTCAG	TETGGGGG	ATGAAGTGA	G
16 BMBL01099827.1_4	TTCAGATATTTGA	G-CAGA	GCGGTTC	TTTTTCTGACT	AAACCTTGT	GTEGTAACC	ACATGC	- ATGT TCCTAT	TTCACTGCAA			GTTGCTGTCAG	TCTGGTG	ATGAACTGA	G
17 PMPL 01009927.1_7	TTCAGGTAAGGGA	TAG-CAGA	GCGGTTCAG	TTTTACTGACT	AAACCTTGT	GITTIATCE	ACATGC		TTAACTGCAA			GTTGCTCTCA	TCTGGTGG	ATGAAGTGA	G
19 PMPL 010099627.1_6	TTCAGCTAAAGAG	-G-CAGA	GGGGTTCAC	TTTTT	AAAccirtur		TIGC		TTCACTGC			CTTGCTCTCAG	TCTGGTGG	ATTAAGTGA	G
10 BMBL01009827.1_9	TTCAACTAAAAGA	GG-CAAA	ACCGATATAC	TTTTGCTGACT	GAATCATGT	GTGGTAACC	ACGIGI	- GTGT TCCAAT	TTCACTACAA	GAAAGAGAAAA	CAAAATTCA	GTT-TTCTCAT	TCAGATIG	GTTAATTGA	ATCTA
20 BMBI 01099827.1	TTCAACTGCAATA		GAGATGCAG	TTTTGCTGACT	AAACCTTGT	AACC	ACGIGA	- ATATTTCAAAT		GAAAGAGAAAA	CAAAATGCA	GTT-TTCTCAT	TCTAGTIG		-
21 BMBI 01099827.1	TTCAATTAAAGGA	AG-CAAA	GACAATGTT	TTTTGCTGACT	AAACCTTGT	GTTGTAACT	GCATGC		TTCGCTGCAC	GTAGGAGAAA	CAAAATAT-	-TTTCTCTCAT	TCTGATTG	ATTAAGTCA	G
22 BMBI 01099827.1		AGG CAAA	AGCGATGTAG	TTTTGCTGAG	AAACCTTGT	GTTGTAAGC	ACGIGC	ATGTTCCAAT	TTCAGTGCAT	GACAGAGAAA		GTT-TTCTCAT	TCTACT	AT A	
23 BMBI 01099827 1	CTCACCTACAAAA	G-CAAA	ATGATATAC	TTTTGCTGACT	GAAACTTTT	GTTGTAAAC	ACGTGC	ATGTTTCCAAT	TTCACTGCAA	GAAAGAGAAAA		GTTTCTCAT	TCTGGTG	ATTAAGTCA	G
24. BMBI 01099827.1					AAACCTTGT	GTTGTAAGC/	ACG GC		TTCATCACAT	GAAAGATAAA		GTTTCTCTCAT	TCTGGTGG	ATTATGTCA	
25 BMBI 01105564 1 1			AGAAGTTCAG	TTTGGCTGAG	AAAGCTTGA	GTGCTTCAC	GTGAGC	ATGTAGCCTCT	TTCCCTGCCA	GAGAGAGGAA	CAAAAGCA	GTTGCTCTCAG	TCTGGGGG	ATGAAGTGA	G
26. BMBL01105564.1.2	TTCAGCTAGAGGA	G CAGA	AGCAGTTCAG	TTTTGCTGACT	AAAGCTTGC	TGGTTAAC	ACGAGC	ATGTATCCTAT	TTCCCCGTAA	GAAATAGAAA		GTTGCTCTCAT	TATGGGGG	ATGAAGTGA	G
27. BMBL01043776.1 1	TTCAGCTAAAAGA	G-CAGA	ACAGTGCAG	GTATGCTGACT	AAACATTGT	GCTGTAAGC	ACTTC		TTCACCACAA	GAAAGCAAAA	CCAAAAGACA	GTTGCTCTCAG	ACTGGTAG	ATGAAGTCA	A
28. BMBL01043776.1 2	TTCAGCTAAGGGA		AGCAGTGCAG	GTTTGCTGAGT	AAACCTTCT	GTT GTGAGG	ACTTGC	TGTTCCAAT		GAAAGAGAAA		GCTGCTCTCAG	TCTGGTGG	ATGAAGTCA	G
29. BMBL01043776.1 3	TTCAGCTAAGGCA	GG CAGA	AGCAGTTCAG	TTTGCCTGACT	AAACCTAGT	GTTGTAAGC/	ATGTGT	AAGTTTCCTAT	TTAACTGCAA	GAAAGACAAA	GCCAAAGACA	GTTGCTCTCAG	TCTGCTAG	ATGAAGTGG	G
30. BMBL01043776.1 4		GT TAGA	AGAAGTTCAG	TTTGGCTGAG	AAAGCTTGA	GTGCTTCAC	AGGAGC	GTGTAGCCTCT	TTCCCTGACA	GAAAGAGGAA	GC-AAAGCA	GTTGCTGTCAG	TCTGAGGG	ATGAAGTGA	G
31. BMBL01006510.1_1	TTCAGCTAGAGGA	AATGCAGA	AGAGGTTCAG	TG TTGCTGACT	AAACCTTAT	TGCTTAA	GT GACC	- ACGTAGCTTAT	T-CCTTGCCA	GAAGGAGAAA	GC-AAAGCAA	GTTGCTGTGAG	TCTGGAGG	ATGAAGTGA	G
32. BMBL01061436.1_1	TTCACCGAGAGGA	G CACA	AGCAGCTCAG	CTTGTCTGACT	AAAGCTTGT	GCGGTTAA	ATCAGC	- ATGTATCCTAT	TTC TG T T CAA	GAAAGAGAAAA	CCAAAAGACA	GTTGCTCTCAG	GCTGTGGG	ATGAAGTGA	G
33. BMBL01061436.1_2	TTCAGCTAGAGGA	TAG CAGA	AGCAGTTCAG	GTTTGCTAACT	AATGCTTGT	GTGGTTAAC	AGAGC	- ATGT TCCTAT	TTCCCTGCAA	GTAACAGAAA	CCTAAAGACT	C TTGCTGTCAG	TCTGGGTA	ATGAAGTGA	G
34. BMBL01061436.1_3	TTCAGCTAGAGGA	CAG AGA	AGCAATTCAC	GTTTTGCT ACT	AAAGCTTGT	GTGGTTAAC	GAGC	- ATGT TCCTAT	TTCCCTGCAA	GAAAAAGAAAA	CCAAAAGACA	GTTGCTCTCAG	TCTGAGCA	ATGAAGTGA	C
35. BMBL01061436.1_4	TTCTGCTAGAGA	CGA-CAGA	AACAATTAA	TGTTGCTTACT	TAAAGCTCAT	GTGGTTAAC	A GAGC	- ATGT TCCTAT	TTCCCTGCAA	GAAAGACAAAA	CCAAAAGACA	GGTGCTCTCAG	TCTGGGGA	ATGAAGTGA	G
36. BMBL01061436.1_5	TTCAACTACAGAA	GA-CAAA	AGCGATGTG	STTTTGC <mark>AA</mark> ACT	AAAGCTTGT	GTTGTCTAC/	AAGC	- ACGTTTGTAAT	TTCACTGCTC	GAAAGAGAAA	C-AAATATT	GTTCTTA	GCTGCTTT	AGTAAGTCT	G
37. BMBL01061436.1_6	TTCAACTACATAA	ACGG-CAAA	AGCGATCTGA	TTTTGGTGACT	AAATCTTGT	TCATTC	ATCCAC		TTCACTGCTC.	AAAAGAGAAAA	CCAAAATAC	GTT CT TCA	ACTGCTTG	AGTAAGTCA	G
38. BMBL01061436.1_7	GGTTCAAGTACAAGA	ACGA CAAT	AGAGATGCAG	TTTCACTAAGT	AAACACTGT	GTTATCAGA	A G GC	- ATGTTCCCAAT	TTCACTACA	GAAAGAGAAA	GCAAAA ATACA	ATTICTCACAT	TCAGCTAG	AGTAAGTCA	G
39. BMBL01082959.1_1	TTCACCTAAAGGA	GG AAAA	AGTGATGTAG	STTTT <mark>A</mark> CTGACT	AAAGCTTGT	GTTGTAAGC/	A GGC	- ATGT TCCCAT	TTCACTGCAA	GAAAGAGAAAA		GTTCCTCTCA	TGTGGTGG	AATAAGTCA	G
40. BMBL01082959.1_2	TTCAACTAAAGGA	GG-CAAA	AGC AAGGT AA	TTTTGCTGACT	AAACCTTG	GGTTTATGC	CAGGC	ATGTTTCCCAT	TGCACTGAAA	GAAAGAGAAAA		GATTCTCTCAT	TCTGCTGG	AATAAGTCA	G
41. BMBL01041892.1_1	TTCAGCTA A AAGA	ACAG-CAGA	AACAGTGCAG	GTATGCTGACT	AAACATTGT	GCTGTAAGC/	ACTTC		TTCACCACAA	GAAAGCAAAAA	CCAAAAGACA	GTTGCTCTCAG	ACTGGTAG	ATGAAGTCA	A
42. BMBL01041892.1_2	TTCAGCTA <mark>AG</mark> GGA		AGCAGTGCAG	GTTTGCTGAGT	AAACCTTCT	GTTGAGG/	ACTTGC	TGTTTCCAAT	TACACTGAAA	GAAAGAGAAA	CAAAAGA	GCTGCTCTCAG	TCTGG	ATGAAGTCA	G
43. BMBL01041892.1_3	TTCAGCTAAGGCA	GG-CAGA	AGCAGTTCAG	STTT GC CTGACT	AAACCTAGT	GTTGTAAGC/	AGIGI		TTAACTGCAA	GAAAGACAAAA	GCCAAAGACA	GTTGCTCTCAG	TCTGCTAG	ATGAAGTGG	G
44. BMBL01041892.1_4	TTTAGCGAGAGGA	GT TAGA	AGAAGTTCAG	GTTTGGCTGAGC	AAAGCTTGA	GTGCTTCAC	AGGAGC	GTGTAGCCTC	TTCCCTGACA	GAAAGAGGAA	GC-AAAGCA	GTTGCTGTCAG	TCTGAGGG	ATGAAGTGA	G
45. BMBL01022471.1_1	TTCAGCTAGAGAA	GGG-CGGA	AGTGGATCAG	TTTTG TGAC1	ATATCCTGT	GTGGTTAAC/	ACGAGC	- AGGTATCITAT	TTCC	GACAGAGAAA	CCAAAAGACA	GTTGCTCTCAT	TCTGGGGG	ATGAAGTGA	G
46. BMBL01022471.1_2		GGG AAGA	AGCGGTTCAC	1 I I I TGCTGACT	AAACCTTGT	GIGGTTAAC		AGGTATC	TICCTTGCAA	GAAAGAGAAA	LCAAAAGACA	GITGCTCTCAT	I C T G G G G G	AIGAAGTGA	
47. BMBL010224/1.1_3		GGG-CAGA	GGTTCAG	TITIGCIGACT	AAAGCIIGT	GIGGIIAAC	GCGAGC	AGGIAICCTAI	TICCIIGCAA	GAAAGAGAAA	LCAAAAGACA	GIIGCIGICAT	TCTGGGGGG	ATGAATIGA	A
48. BMBL010224/1.1_4		GGG-CAGA	GGTTCAG	TITIGCIGACI	AAAGCTTGT	GIGGITAAC	ACGAGC	AGGIAICCTAI	TTCCTTGCAA	GAAAGAGAAAA		GITGCICIAGI	TCTGGGGGG	ATGAAGIGA	G
49. BMBL010224/1.1_5		GGG-CAGA	GGTTCAG	TTTTGCTGACT	AAAGCIIGT	GIGGIIAAC/		ATCTATO	TTCACTCCAA	GAAAGAGAAAA		GTTGCTCTCAT	TETEEGGGG	ATGAAGIGA	G
50. BMBL01022471.1_6	I I CAGCIAR GG	-CAGA	GUAGIICAG	I I I I GA I GACI	MAAGCIIGI	GIAGC	GCGGGC	AIGINIC	IICAGIGCAA	MAAGAGAAAA	JCAAAC GACA	GIIGCICICAG	ICIGGG G	AIGCAGINA	G

Supplementary Figure 6. Alignment of 50 monomers manually isolated from the *Dugong dugon* assembled reference genome (accession GCA_015147995.1). Part three of six.

	440	450	460		470 4	80	490	500	510	520	530	540	550	560	570	580
Consensus	AAGGATTC	ACTTAC-	-TGTGTTTG-	AAAAT	CCCTCTGTGT	TACATGCCC		AGAAAGTTGCCT	GTGCATTGT	CTACCTGGGT-	AAAGAGT	TTGACTTTT	CTCCCTTTCAAC	TACGTGAAGC	CTGCGGGG	TTACAGCT
Identity						_										The Party of the P
1. BMBL01009774.1_1	CACGATIC	ACTINC-		AAAGI	CCCTCTGTGT	TACATGCC	A	AGAAAGTTGCCT	GIGCATIGI		AAAGCGI	AIGACITIT	CICCCITICAA	TACTIGAAGC		TICCAGCA
2. BMBL01009774.1_2	AATATIC	ACTIC	- IGIGIIIG	AAAAT	CCCIGIGIGI	TACATGAC	A	AGAAAG	GGCATT	CTACCTGGAT-	AAAGGGI	AIGACIIII	CTCAATTICAAC	CAGIIGAAGC	GICAGGII	TICCAGCA
3. BMBL01009774.1_3	AACAATTO	CTTTC-	- IGIGIIIG	AAAAI	CCCTCTGTGT	TACATGCC	A	AGAAAGTTGCCT	GIGCATIGI	CTACCTGGAT-	GAAG	AIGACITTI	CICCCITICAA	JIAGI IGAAGO	GICAGGGI	TICCAGCA
4. BMBL01009774.1_4	AACAGTIC	ACTINC-	- IGGGIIIG	AAAA	CCCTCTGTGT	TACATGCO	p	AGAAACTIGCCT	GIGCATIGI		AAAG GI	AIGACIIII	CICCCITICAAC	AGGIIGAAGC		TICCAGCA
5. BMBL01009774.1_5	AAGGATTC	ACTT C	TGIGITIG	AAAAT	CCCTCTGTGT	TACATGCC	p	AGAAAGAIGCAI	GIGCATTOT	CTACCTGGAT-	AAAG	ATGACTITI	CTCCCTTTCAAC	AGGITGAAGC		TTCCAGCA
6. BIVIBL01009774.1_6	AAGGATTC	ACTT	TGIGITIG	AAAAT	CCCTCTGTGT	TACATGCO		CAAAGTTGCCT	GIGCATIGI	CTACCTGGAT	AAAGCGT	ATGACTITT	CTCCCTTTCAAC	TAGETGAAGC		TICCAGCA
7. BIVIBL01009774.1_7	AATGATTC	ACTIC	TETETTE	AAAAT	CCCTCGGIGI	TACATGCO		GAAAGTTGCCT	GIGCATGGI	CTACCTGGAT-	AAAGGGT	ATGACGTTT	CTCCCTTTCAAC	TAGETGAAGC		TTECACCA
8. BIVIBLU1002074.1_1	AANGATTC	ACTT	TCTCTCTC	AAAAT	CCCTCTCTCTCT	TACATGCC		AGAAAGTTGCCT	GIGCATIGI		AAAG	ATGACGTTT	CTCCCTTTCAL	TAGETGAAGC		TTECAGCA
9. BIVIBLUTUU2U74.1_2	CACCACTO	ACTTAC	TGTGTGTG	AAAAT	CCCTCTCTGTGT	TACATGCO		AGAAAGTTGCCT	GTGCATTGT		AAAG	ATGACUTTT	CTCCCTTTCAA	TAGETGAAGC		TTCCAGCA
10. BIVIBLUTUT2000.1_1	GAGGAGTC	ACTINC	TGTGTTTG	AAAAT.	CCCTCTCTGT	TACATO		AGAAAGTTGCCT	GTGCATTGT	TACE TOGAT	AAAGGGT	ACGACCTTT	CTCCCTTTCAA	TAGETGAAGC	GTTCAGGGT	TTCCAGCA
12 PMPL 010092711	AAGGATT	ACTAC-	TGTGTTTG	AAAAT	CCCTGTGTGTGT	TACALGCCC			GTGCATTGT	CTACCAGAGT-	AAAGAGT	TTGACTTTT	CTCCCAGTCGA	TCGTAAAGO	CTACAGO	TTACAGCT
12. BIVIDL01099627.1_1	AAGAAGT	ACTOAC-	- TGTGTTTG	AAAAT	CCCTCTCTGTGT	TACEGCCC		GAGAGGTGCCT	GTGCATTGT	ATACCTGGAT-		TTGAATTTT	CTCCCATTGGA	TCGTGAAGC	CTGCGGGG	TTACAGCT
14 PMPI 01009271 2	GAGGACTC	AATTAC-	- TGTGTTTG	AAT	CCCTTGGTGT	GANATGCO		AGAAAGTTGCCT	GAGCATTAT	CTACCTGGGT-	CAAGAGT	TTGACTTTT	CTCCCGTT	COTGAAGE		TTACAGCT
15 BMBL01009827.1_5	AAAGATTC	ACTTAC-	- TGTGTTTG	-AAAT	CCCTCTGTGT	TACATACO	A	ATAAAGTTGCCG	GGCATTGT	CTACCTGGGT-	AAAGGGT	TTGACTTTT	CTCCTGTTTGA	TAGGTGAGC	TACAGGGT	TTACAGCT
16 BMBL01099827.1_4	AAAGATTC	ACTTAA-	TATGTT	ATAGT	CCCTTTGTGT	TACATO	A	AGAATGTTGCCT	GTGAATTET		AACAAAGCAT	TTCACTTTT	CTCTCCTTAGAC	TACACGAAGC	TACGGAGT	TTGCAGCT
17 BMBL01099827.1_7	AAGGATT	ACTTAA	TGTGCTTG	AAAAT	CCCTCTGTGT	TACATGCC	A	AGAATGTOCCT	ATGCATTGT		AAAGCAT	TTCACTTTT	CTACCETTGAA	TACATGAAGC	CTGCGGGGT	TTACAGCT
18 BMBL01099827.1_0	AAGGATTC	ATTAC-	TGTGTGTGTG	AAAAT	CTCTGTAT	CACATGCO	A	GGAATGTTGCCT	ATGLATGET	CTACCAGGGT-	AAAGAGA	TTGACTTTT	CTCCCCTTCGAG	TACGTGAAGC	CTGCAGGAT	TTACAG
19 BMBI 01099827 1	GTTGATTA	AGTTGTC	AGTTTGA	AAAAC		AGTATCOAC		AACCTTCCTC	ATACGTTCT	CTGAATGGGA-	AATGAAT	GTCATGTTT	CTCCCTTTCCAC	TACTIGAATC	TCCAGGTT	TCAGACTT
20 BMBI 01099827 1	ATTA	ACTTAC-	TCTAGTTG	AAG	TGGTCTGTTT	TG-ATA	A	AACACTCTTCCC	ACACATTCT	GT GAAT GGGA	AAAGAAT	GTACAGTTT	CTCCCTTTCAA	TAAATGAGGA	TTCAGGTAT	TCTGATGT
21 BMBI 01099827 1	AAGGATTA	ACTTA	TGTGTTTG	AAAAT	AACTCTGTGT	TACTTGCT		AGAATGTTGCCT	ATACATTCT	CTACCAGGGT-	GAAGAAT	TTAACTTTT	CTCCCATTCAA	TACATGGAGC	ATATGAGGT	TTAGATGT
22. BMBL01099827.1	ATT	AAT	CTTGTTTT					AGAAAACACTGT	TCTCGTTCT	CT GGATGGGA	AAAGAAT	ACCAAATTT	CTCCCTTTCAAG	TACATGAAGC		TCAGATAT
23. BMBL01099827.1	AAAGATTA	AATTAC-	TGTGTTTG	AAAAC	TCTGTGT	TACTTGCA	0	AGAAGGTTGCCT		CTACCAAGGT-	AAAGAAT	TTAACTTTT	TCCCAATCTA	GAGAAGC	ACACGGGAT	TTAGATGT
24. BMBL01099827.1	GAGGGCTG	ACTTAC-	TATGTTTG	AAAA-	CACTCTGTGT	TCATGCAT	C	AGAATGGTTTGT	ATGCATT	CACCTGGGT-	AAATAAT	TTAAATTTT	GTCCCTTTCAA	TAGATGAAAC		TTAGATGT
25. BMBL01105564.1 1	AAGGATCC	ACATAC -	TGTGTTTG	AAAAT	CCCTCTGTGT	TACATGCCC	A	AGAATGTTGCCT	GTGCATTGT	GACCTGGAT-	AAAGAGT	TTGACTTTT	CTCCCTTTCGAG	GTACATGAAGC	CCTGCTATGT	TTACAGCT
26. BMBL01105564.1 2	AAAGATTG	ACTTAC-	TGTGTTAG	AAAAT	CCCTCTTAT	TATATGO	A	GAAATGTTGTCT	GTGCATTGT	TACCTGGGT-	AAAGACT	TTGACTTTT	CTCCCTTTCCAC	GTAC ATGAAGC	CCTGCGGGGT	TTACAGCT
27. BMBL01043776.1 1	AAGGATTC	ACTTAC-	TATGTTTC	AAAAT	CCCTCTGTGT	TATGAGCCC	A	AGAATGTTGCCT	ATGCATTCT	CACCTGGAT-	AAAGAGT	TCGGATTTT	CTCTCCTTCGAC	TACGTGAAGA	CCTGTGGGGCT	TAAGAGGT
28. BMBL01043776.1_2	AAGGATTC	ACTTAC-	TGTGTTTG	AAAAT	CTCTCTGTGT	TCATGCCC	A	AGACTGTTGCCT	GTGCATT	CACCTGGGT-	AAAGAG	TCGACTTTT	CTCCCCTTTGA	STACGTGAAGC	CCTGCAGGGT	TCAAGGCT
29. BMBL01043776.1_3	GATTO	GCTTAC	TGTGTTGG	AAAGT	CCCTCTGTGT	TACACACCO	A	GAATGTTGTCC	AGGAATTCT	CTACCTGGGT-	AAAGACT	TTGATTTT	CTCCCCTTCCA	TACGTGAAGO	ACCGCAGAGT	TTACAGCT
30. BMBL01043776.1_4	AAGGATTC	AC-TAC-	TGCGTTTG	AAAAT	CCCTCTGTGT	TCATGCO	AAGGA	AGGAAGTTGCCT	GTGCATTGT	GTACCTGGGT-	AA <mark>T</mark> G <mark>G</mark> GA	TCCACTTTT	CTCCCTTTCAAC	CAGGTGAAGC	CAT C GGGT	TTAAAGCA
31. BMBL01006510.1_1	AAGGAGTC	ACCTAC-	TGTGTTTG	AAAAT	T CCT A TGTGT	TACATGCC	ββ	AGAAAGTTGCCT	CTGCATTGT	CTCCCTCGGT-	AAAGGGA	TCCACGTTT	CTCCCTTTCAAC	GTATGAAGCO		TTACAGCT
32. BMBL01061436.1_1	AAGGATTC	ACTTAC-	TGTGTTAG	AAATT	CCTCTGTGT	TACATGCCC	A	AGAACTTTCCCT	GTGC	IGCCCTGGGT-	AAAG T GT	TTGACTTTT	CTCCCTTTCGA	TACGT	CCTGCGGGGT	TTACCGCT
33. BMBL01061436.1_2	AGGAGTC	ACTTA	TGTGTTTG	AAAAT	CCCTCTCTGT	TACAGGCCC	2A	AAAATGTAGCAT	GTACATTGT	CTACCTG T GT-	AAAGAGT	TTGCCTTTT	CTCCCCTTCGAC	STACGTGAAGC	CCTGCGGGGT	TTACAGCT
34. BMBL01061436.1_3	AAGGTTTC	ACTTAC-	TGTTTG	AAAAT	CCCTCTGT	TACATGCCC	2A	AGAAAGTTGCC	CTGCGTTGT	TACTGGGGT-	AAAGAGT	TTGACTTTT	CTCCCATTCGA	CACGTGAAGA	CCTGTGGGGGT	TTAC
35. BMBL01061436.1_4	AAAGATTC	AC	- TIT TGTTTG-	AAAAT	CCCTCTATGT	TACATGCCC	A	AGAAAGTTGCCT	GTGCGTTGT	CTACCTGGGT-	AAAGAGT	TTGACTITT	CTCCCCTTTGA	STACT TGAAAC	CCTGAGGGGT	TTACAGCT
36. BMBL01061436.1_5	AAAGATTA	ACTICC-	IGIGITIA	AAAAA	CCCTCTGTGT	TACATGCA	6	AGAGIGTIGICT		CTACCAGGGT-	AAAG <mark>G</mark> GT	GAACIIII	CICCCITICAA	TACATGAAAC		TTAGATGT
37. BMBL01061436.1_6	AAGGATTA	ACTTAC-	-TGCGTTTG	AAAAC		TAAATGCAC	A	AGAGIGTTGCCT	ATACATTCT	CTACCAAGGT-	AAAG GC	I GAACTITT	CTCCCATTCAA	TACGTGAAGC	GGGGG	TTAGATGT
38. BMBL01061436.1_7	AAGAATTA	AATTAC-	- IGIGIIIG	AGAAA	CACIGIGI	TACAGGCA	6	AGAACATTGCCT		ATAIGIGGGI-	AAAGAAT	GAACITIT	CTCICACICAA	TACGIGAAGO	AIGCAGACI	ICALATGI
39. BMBL01082959.1_1	AAGGATTA	AGITA -	- IGIGCIIG	AAAAI		TACATGCAC	P	AGAALGIIGC	AIGCATICI	CACCIGGGI-	AGAGI	TAACTIT	CICACATICGAG	TACGIGAAGC		TIAGA
40. BMBL01082959.1_2	AAGGATTA	ACTTAC-	TGIGGIIG	AAAAT	COCTOTOTOT	TACTIGCA		AAAAGTTGCCT	ATACATTET	CACCIGGGI-	AAAGAGA	TAAGCIIII	CTCCCATTGAC	TACGIGAAG		TAGGIGT
41. BMBL01041892.1_1	AAGGATTC	ACTTAC-	TATGTTTC	AAAAT	CECTETETET	TATGAGECE		AGAATGTTGCCT	ATGCATTCI	ACCIGGAT	AAAGAGT	ICGGATTTT	CTCCCTTCCAC	TACGIGAAGA		TAAGAGGT
42. BMBL01041892.1_2	AAGGATTC	CTTAC		AAAAT	COCTOTOTOT				GIGCATICI	CTACCTGGGT	AAAGAG	TTCATTTT	CTCCCCTTCCA	TACGTGAAGC		TTACACCT
43. BIVIBL01041892.1_3	AACCATTC	AC-TAC	TOTOTTE	AAAGT	CCCTCTGTGT	TACATCCC	AACCA	GAAGTTGCC	CTCCATTCT	TACCTGGGT-	AAAGAC	TIGALITIT	CTCCCTTTCAM			TTACAGET
44. DIVIBLUTU41892.1_4	ARGETTC	ACTTAC-	TGTGGTTG	AAAAT	CCCTCTATGT	TACATGOOD	AAGGA	GAAAGTTGCCT	GIGCATIGI		AAAGECT	TTGANTTTT	CTCCCTTCCA			TACAGCT
45. DIVIBLUTUZZ4/1.1_1	GGGATEC	ACTTAC-	TGTGCATC	AAAGT	CCCTCTGTGT	TACATAC	-	GCAATGTTCCT	GTACATTOT		AAAGAGT	TCLATTT	CTCCCGTTCGA	TACGTGAAGC		TTACAGET
40. BIVIDL01022471.1_2	AAGGATTC	ACTTAC-	-TGTGTTTG	AAAAT	CCCTCTATC	TACATO		AGAAAGTTGCCT	GIGCATIG	ATCCCTGGGT-		TTGAATTAT	ATCCAGTTCAA	STACGTGAAG	CTGCGTGGT	TTACAGCT
47. BIVIDLUTU22471.1_3	AAGGTTTC	ACCTAC-	-TGTGCTT	AATAT	CCCTCTGTGT	TACATGOOD		AGAAAGTTGC	GTGCATTGT		BAAGAGT	TTGAACTTT	CTCCTGTTCGA	TACGTCAAGC		TTACAGCT
40. BIVIDL01022471.1_4	AAGGATCO	ACTTAC-	TGTGCTTG	AAAAT	CICICICICIC	TACATATCO		AGAAAGTTGCCT	GTGCATTGT	ATCCCTGGGT-	AAAG	TTGAATTAT	CTCCCGTTCAA	TACTIGATICO	CTGCGGGG	TTACAGCT
50 BMBI 01022471 1 6	AAGGATTC	AC	TGAGTTTG	AAAAT	CCCTCTGTGT	GACATGCCC	A	ACAATTCTGCCT	GTGCGTTGT	CTACCTGGGT-	AAAGAGT	TTGACTTTA	CTCCCCTTCCA	TAAGTGAAG	CCTTC-GGGT	TTACAGCT

Supplementary Figure 6. Alignment of 50 monomers manually isolated from the *Dugong dugon* assembled reference genome (accession GCA_015147995.1). Part four of six.

	590	600	610	620	630	640	650	660	670	680	690	700	710	720	73
Consensus	TTCCGGGAGGCA	ACTCTGTGC	TGCGCGGCCG	AGAATG	-CCGCCTTACAC	GACGGTATCTC	GAAGGAGAAG-	AGAAAGCG	GTTTCTCCCA	TTCTGGTATA	CGACAGCGG	CAGAGMTGTA	GTTTTG	TAAGTAAACA	CTATG-TT
Identity							- In- stall-								
1 BMBI 01009774 1 1	TTCCGGGAGGCA	ATCTGTGC	TGCGTAGCTG	AGAATA	GGGCTTTACAG	GACGCTATCGC	TAAAGGAGAAG -		ATCTCTCCCA	TTCTGGAAGG	CGAAGGAGG		GTTCTC	TGAGTAAAAA	
2 BMBI 01009774.1_1	TTCCGGGAGTCA	ATTCTGGTC	TGCGCGGTTG	AGAATA	CCGCTTCACAG	GCGGTATCCG	TAAAGGAGAAG-	AGAAACAGO	GTTTCTCCCA	TTCTGGAACG	CGAGGGAGT	CAAAACTGTA	GTTGTGC	TGAGTAAAAC	CTTTG-TT
3. BMBL01009774.1 3	TTCCGGGAGGC	ATTCTGTGC	TGCGCGGC	AGAATA	CCGCTTTACAC	GACGGTATCTC	TAAAGGAGAAG	AGAAACCG	GTTTCTCCCA	TTCTGGAACC	CGAGGGAGG	CAAAACCATA	GCTGTG	TGAGTAAAAG	CTTTG-TT
4. BMBL01009774.1 4	TTCCGGGAGGA	ATCTTCC	TGCGCGGC	AAATA	CCGCCATACAG	GCAGTATCGC	TAAAGGAGAAG -	AGAAACCC	GTTTCTCCCA	TTCTGGAACG	CGAGGGAG	CAAAACCGT	GTTGTG		CTTTG-TT
5. BMBL01009774.1 5	TTCTGGGAGGCA	ATTCTGTGC	TGTGTAGATG	AGAATA	GGGCTTTACA	GCGGTATCGC	TAAAGGGGAAG -	AGAAACCGO	GTTTCTCCCA	TTCTGGAACG	CGAGGGAGT	CGAAACTGTA	TTGTGC	TGAGTAAAAAA	CTTTG-TT
6. BMBL01009774.1_6	TTCCGGGAGGC	ATCTGTGC	TGCGCGGC	AGAATA	-CC-CTTTACAG	GACGGTATC GC	TAAAAGAGAAG	AGAAACCG	GTTTCTCCC	TTCTGGAAGG	CGAAAGAGG	CTTAACCGCA	GTTGTGC	GGAGTAAAGA	CTG-TT
7. BMBL01009774.1_7	TTCCTGAGGC	TTTTTGGGC	TGCGCGGC	AGAATA	-CCGCTTTACAC	GACGGTATC <mark>G</mark> C	TAAAGGAGAAG	AGAAACCAC	GTTTCTCCCA	- CTGGAACT	GAAGGAGG	CAAAATCGTA	GTT <mark>G</mark> TGC	TGAGTAGAAA	CTTTG-TT
8. BMBL01002074.1_1	TTCCATAAGGC/	ATCTGTGC	TGCGTGGCCC	AGAAT	-CGGCTTTACAC	GACGGTAACGC	TAAAGGAGAAG-	AGAAAACGO	GTTTCTCCCA	TTCTGGAACG	CGA <mark>AG</mark> GGGG	- AAAACCGTA	GTT <mark>G</mark> TGC	TGAGTAAAAAA	CTTTG-TT
9. BMBL01002074.1_2	TTCCGGGAGGCA	ATCTGTGC	TGCGCGGC	AGACTA	-CCCGTATA-		GAAAGAACTGG-		TCTGCCA	ATCTGCTTC-	CTT	TAAGATTACA	AC	CAAGAAAAC	CTATG
10. BMBL01012666.1_1	TTCC GGAGGCA	ATCCTGTGC	TGCGAAGCTG	AGAATA	GG GC T TTACAG	GACGGTA CCG C	TAAAGGAGAAG	GAAACCGO	GTTTATCCCA	TTCTGGAACG	CGA <mark>AG</mark> G A GG	CAAAACCGTA	GTTGTGC	TGAGTAGAAA	TTTG-TT
11. BMBL01012666.1_2	TTCCGGGAGGC/	ATCTGTGC	TGCGCTGCTG	AGAATA	-CCGCTTTACAC	GACGGTATC G C	TAAAGGAGAAG	AGAAACGGG	GTTTCTCCCA	TTCTGCAAGG	CTAAGGAGG	CAAAACCGTA	CTTCTCC	TGAGTAAAAC	ATT TG-TT
12. BMBL01099827.1_1	TTCTGGGGGGGC/	ACTCTGTGA	TGCATGGA-G	AGAATG	CCGCCTTACAG	GACGGAATCCC	TGAAGAAGAA-	AGAAAG	GTTTCTC	TTCTGGTATA	CGACAGCAG	CAGCGCGGTA	GTTTTC	TAAGTAAACA	CTATG-TT
13. BMBL01099827.1_2	TTTCAGGAGGC/	ACTCTGTGC	TGCGAGGCCG	AGAAT	-CCGCCTTACAG	GAGGTATCCC	TAAAGGAGAAG -	-AGAGCGC	TITCICCCG	TCCTGGTATA	CGACAG	AGCGCTGTA	GTTTTGC	TGAGTAAAGA	TATG-T1
14. BMBL01099827.1_3	TTCCAGGAGGC	TGTGC	TGCGCAGACG	AGAAT		SACGGTATCAC	TAAAGGAGAAG	AGAAA	GTTTCTCCCA	TTCTGGTATA	GAAAGCGG	CAGCALIGIA	GTTTTGG	TGAGTGAAGA	CTATG-TI
15. BMBL01099827.1_4	TTCCGGGGAACCA	AGTETETEE	TGGGCAGCCG	AGAATG		ACGGTATCCG	AAAAGGAGAAAG-	AGALACCE					GTTTTC	TAAGTAAAGA	CTACG-TT
16. BMBL01099827.1_7	TTCCCCGGGGGA	ACTOTOTO		AGAATG						TTCTCCTATA	GATEGOAA	AGAGACATA	CTTTTC	TAAGGAAACA	CTACA GI
17. BIVIBLU1099827.1_8	TTCAGGGAGGC	ACTCTGT		AGAAAG	COCCTOACAC	ACAGTATOTO	TGAATGAGAAA			TTCTGGTATA	CGACAGTAG		GTTTTA	TARGUAAACA	CTATE-TT
10 PMPL 01009271	TTCHIGGGA	ACTUTOCAT		GAATGGA	GCCTURACAC		GAAGAAGAAGG			TTCATCTATA	AGATGGCAA	AGAGATGTA	-CTTTGO	TGACTGAACC	CTGTG-TG
20 BMBL01099827.1	TTTCTATGAGA	ACTITOCA	TAC-CAGGAA	GAATG	TGT CTTT		GGAAGAAGAAGG	CAGA	TGTTCCCG	TTCAATTACA	AGACAGCAA	-AGGTATGTA	GTTTTG	TGACTAAACC	TATG-TT
21 BMBL01099827.1	TTCCAAGAAGCA	ACTCTGTGC	TGCCTGGCCA	TAAATG	- GCTTTACAG	ACAAAATATC	TGAAGGAGAAA		GTTTCTCCCG	TTCAAGTATA	GGACGGTAA	TGGAGAAGTA	ATTTTG	TAACTAAACA	CTGTC
22 BMBI 01099827 1	TTTCTATGGGA	ACTTTGGG	TATGGGGA	ATAACT	TGGCCTTTCAC		AGAAGGACAAGG	CAGA	TGTTCCTG	TTCAACTACA	GGTCAGCGA	AAGCCATGTA	GTTTTG	TAAGCAAACA	CTGTG-TT
23 BMBI 01099827.1	TTTGAAGCAGT/	ATTCTGTG	TCCCAGGC	GTATG	- GCCTTAGG	ACAAAGTCTC	TGAACAAGAAA	ACAATGCG	GTTTCT	TTCAAGTATA	GGATGGCAT	TAATGATGTA	GTTTTG	TCAGCAAACA	TATG-TT
24. BMBI 01099827.1	TTTCAAGAAGT/	ACTCTGTGC	TGCCTGACCA	GAATG	GGCCTTACAG		TGAACCAGAAA	ACAATGCAA	GTTTCTCC	TTCAAGTATA	GGA		GTTTTG	TAAGTAAATA	CTCTG-TT
25. BMBL01105564.1 1	TTCTGGGAGGCA	ACTCTGTGC	TGCCCAGATG	AGAATG	-CCGCCTTGCC	GAGGTATCC	TGAAGGAG <mark>GG</mark> G	AGAAAGCGC	GT-TCTCCC	TTCTGCTATA	CAACAGTGG	CAGCACTGTA	GTTTTG	TAATTAAGCA	CTATA-TT
26. BMBL01105564.1 2	TTCCGGGAGGCA	ACTCTGTGC	TGCGCCACCG	AAATG	TAGCT	ACTGAGTCT	TAAGAAAAGAA		C <mark>TG</mark>	тт		GAGTATTATC	GTATTA	AAAATGATGT	ATATAATT
27. BMBL01043776.1_1	TTCTGGGAGGC	CTCTGTGC	TGCCCAGCCA	AGAACG	CAGCCTTATA	GATGGTATCTC	TAAAAGAGAAA	ACAATGIG	GTTACACTTA	TACAGGTATA	AGAGAGCAG	AGACATGTG	ATT-TGO	TAAGTAAACA	CTATG-TT
28. BMBL01043776.1_2	TTC T GG TG GGC/	ACTCTATGC	TGCACAGCCA	AGAAAC	-CAGCCTTACA	AAAATGTCTC	TGAAGG <mark>G</mark> GAA <mark>A</mark> -	AGAAAGCAA	ATTTCTCCCA	TTCTGGTATA	CCCCAGCAG	AGAGATGAA	GTTTTG	TAAATAAACA	CTATG-TT
29. BMBL01043776.1_3	TTCC-GGAGGA	ACTCTATGC	TGCCCGGC	AGAGAG	TG GCCTTACA	AGCGAGTCTC	TGAAGG <mark>G</mark> GAA <mark>A</mark>	AGAAAGCGC	GTTTCTCCCA	TTCTGGTATA	CCACAGCAG	TG GAG A TATA	GTTTGG	GAAGTAAACA	CTACG-TT
30. BMBL01043776.1_4	TTCC GATGACA	ATCTGTGC	TGCTGGGCCG	GAATG	CAGCTTTACAG	GACGCTATCCC	TAAAGGAGAAG	-AGAAACC	- TTTCTCCAG	TTCTGGTATA	CCACCGTGG	CAGAGCCGCC	GTTTTG	TGTGTCAAAA	GTATG-TT
31. BMBL01006510.1_1	TTC GA GGAGGC/	ATCTGTCC	TGCGCGGCCG	AGAATG		SACGGTATCAC	AGAAGGGAG-	GAAACCG	GTTTCTC	TTCTGGTATA	GGACAGCGG	TGGAGCCACA	GTTTTG	TGAGTAAAGA	CTATG-GT
32. BMBL01061436.1_1	TTCCGTGTGGC	CTCTTTGC	TGCGCTGC <mark>G</mark> G	ATAATG	-CCGCCTTGCA	CCGTATCCC	TGAAGG-GAAG-	AAAAGCG	ATTTCTCCCA	TTCCCATATA	CGGCAGCAG	CAGTGCTGTA	CITTIGO	TAACTAAACA	CTATG-TT
33. BMBL01061436.1_2	TTCAGGGAGGC		TGCACGGCGG	GAATG		ACATCITCAC	TGAAGAAG	AGAAAAC	GITTCICCAG	TICIGGIAIA	GAGAGCGG	CAGCAGIGIA	GITTGO		CIGIG GA
34. BMBL01061436.1_3	TTACAGGAGGCA	ACTOTOTOC	TGCTAGGGAG	AAATG	CCACITACAC	JACICIAIGIG	TG GGGGGGGAGAG	AGAAAACGC		TICAGGIAIA	GACAGCAA	CAGCIGIGI	GIIIIGO	TAACTAAAGA	CIAIG-II
35. BIVIBLU1061436.1_4		ACTUCIGIC	TUCCOUL	CAATG			TCAACCACAAA		GTTTCACCT		CONCERCINC		CTTTCT	TAAGAAAAG	CTATC-TT
30. DIVIDLU 1001430.1_3	TTTCCAGCAG	ACTOCTOTOC	TITCIATICG	GAATA		ACAGAATCT	TGAAGGAGAAA		GTTTCTCCCA	TTCAAGTACA	GGATGGCGA	CAGAGATGTA	GTTTCA	TAAGTAAACA	
20 PMPL 01061430.1_0	TTTCAAGTTGC	ACTCTGTGC	Techceer	GAATG	- GCCTTACA		TGARGGAGAGAA		GTTTCTCCCC	TTCAAGTATA	AGACAGTAA	AGEGAGGT	GTTCTG	TAAGTAAACA	CTGTG-TT
20 PMPL 01092050 1 1	TTCCAGTAGGC	CTCTGTGC	Techcoecc	CAAATG	GGCCTTACAC	ACAGAATETC	AAAGGGGAAA-	- ACAATGCG	GTTTCTCCCC	TTCAAGTATA	AGACAGCAA	CAAAGATGT	ATTTCG	TAAGGAAACA	
40 BMBL01082959.1_1	TTTCAGTAGGC	ACTOGTOC		GAATG		GACAGTATCTC	TAAAGCAGAAA		GTTTCTCCCA	TTCAAGTATA	GACAGCAA	AGAGATGTA	GTTTTG	TAAGGAAACA	CTATG-TT
41 BMBI 01041892 1 1	TTCTGGGGAGGC	CTCTGTGC	TGCCCAGCCA	AGAACG		GATGGTATCTC	TAAAAGAGAAA	ACAATGIG	GTTACACTTA	TACAGGTATA	AGAGAGCAG		ATT-TGO	TAAGTAAACA	CTATG-TT
42. BMBL01041892.1.2	TTCTGGTGGGC/	ACTCTATGC	TGCACAGCCA	AGAAAC	CAGCCTTACA	AAAATGTCTC	TGAAGGGGAAA-	AGAAAGCAA	ATTTCTCCCA	TTCTGGTATA	CCGCAGCAG	AGAGATGAA	GTTTTG	TAAATAAACA	CTATG-TT
43. BMBL01041892.1 3	TTCC-GGAGGA	ACTCTATGC	TGCCCGGCTG	AGAGAG	GCCTTACA	AGCGAGTCTC	TGAAGGGGAAA-	AGAAAGCG	GTTTCTCCCA	TTCTGGTATA	CCACAGCAG	GAGATATA	GTTTGG	GAAGTAAACA	CTACG-TT
44. BMBL01041892.1 4	TTCCTGATGAC	ATCTGTGC	TGCTGGGCCG	GAATG	CAGCTTTACAG	GACGCTATCCC	TAAAGGAGAAG -	-AGAAACC	- TTTCTCCAG	TTCTGGTATA	CCACCGGGG	CAGAGCCGCC	GTTTTG	TGTGTCAAAA	GTATG-TT
45. BMBL01022471.1 1	TTCCCG-AACCA	ACTAGTTGC	TGCGCGGCC	AGAACA	CCGCCTTATAC	GACTGTATGCC	TGAAGGAGAAG-	AGAAAGCG	GTTTTTCCTG	TTCTGGTA	AGACAG	CAATGCTGTA	GTTTTG	TAACTAAAGA	CTATA-TT
46. BMBL01022471.1 2	TTCTGGGAGGCA	ATCTGTGC	TGTGTGGCCG	AGAACG	CCCTCTTACAG	GATGGTATCCC	TGAAGCAGAAG	AGAAAGCG	GTTTCTCCC	TTCTGGTATA	AGACAG	CAATGCTCTA	GTTTTG	TAACTAAACA	CTATG-TT
47. BMBL01022471.1_3	TTCCGGGAGGCA	ACTCTGTGC	TGCGCGG <mark>GT</mark> G	AGAA <mark>G</mark> G	CCTCCTTACAG	GACCTTATGCC	TGAAG <mark>A</mark> AGAAG	AGAAAGCG	GTTCCTCCCA	TTCTGGTATA	GCCCAGTGG	CAATGCG	GTTTTG	CAAGTAACCA	
48. BMBL01022471.1_4	TTCCTGTAACCA	ACT <mark>GGT</mark> TGC	TGCGCGGCCA	AGAACG	-CCACCTTATAG	GAAGGTATCCC	TGAAGGAGAAG	GGTAAGCG	GTTTTTCCCA	TTCTGGTA-A	AGACAGCGG	CAATGCTGTA	GTTTTG	TAACTAAACA	GTAT TT
49. BMBL01022471.1_5	TTCAGGGAGGCA	ACTCTGTGC	TGTGCGGCCG	AGATCG	-CCGCCTTACAG	GACGGTATCC	TGAAGCAGAAG	AGAAAGCG	GTTTCTCCCG	TTCTGGTATA	CAACAGCAG	CAACGCTGTA	GTT <mark>G</mark> TGC	TAACTAAACA	CTATG-TT
50. BMBL01022471.1_6	TACCAGGTGTC	ACTCAGTGC	TGTGCGGCCC	AGAATG	-CGGCCTTACG	GACAGTATGTC	TGAAGGAGAA	ACAAAGCG	GTTTCTCCC	TTCTGGTCTA	GACCGCTT	AGAGATGTA	GTTTTG	TAAGTAAACG	CTACT-T1

Supplementary Figure 6. Alignment of 50 monomers manually isolated from the *Dugong dugon* assembled reference genome (accession GCA_015147995.1). Part five of six.

	740	751
Consensus	AGCAGCACGTGCATGTTC	CCA
Identity		
1. BMBI 01009774.1 1	AGCAGTAAGTTAAGTTO	CC
2. BMBL01009774.1 2	AGGAGCAAGTGGAAGTTC	CCC
3. BMBL01009774.1 3	AGGAGCAAGTGGAAGTTC	CCC
4. BMBL01009774.1 4	AGCAGGAAGTGGAAGTTO	
5. BMBL01009774.1_5	AG <mark>G</mark> AGCA <mark>A</mark> GTG <mark>G</mark> AAGTT	CCC
6. BMBL01009774.1_6	AGCAGCAAGTGCCAGTT	CCC
7. BMBL01009774.1_7	AGCAGCAAGTGTAAGTTC	CCC
8. BMBL01002074.1_1	AGCAGCAAGTGCAAGTT	CCC
9. BMBL01002074.1_2	GAGCAC	
10. BMBL01012666.1_1	AGCAGCAATTCTAAGTTC	CCC
11. BMBL01012666.1_2	AGCAGTAAGT	CC
12. BMBL01099827.1_1	AGCAATACTTGCATGTTC	CGA
13. BMBL01099827.1_2	AGCAGCAAGTGCAAGTTC	CCA
14. BMBL01099827.1_3	AGCAGCAAGTGAAGTTC	CA
15. BMBL01099827.1_4	AGCAGCAAGTGCAAGTTC	.CCA
16. BMBL01099827.1_7	AGCAGAGC	
17. BMBL01099827.1_8	AGCAGGAC	CAA
10. DMDL 010099827.1_9	AGCAGAACTIGCATGTT	CCA
19. BIVIBLU1099827.1	GTAAGCACTIGTATGTT	CCA
20. BMBL01099827.1		CCA
27. BMBL01099827.1	ACCAGAAAGTGCATGTTC	CCA
23 BMBI 01099827.1	ACCAGACTGAGCACTTT	TCA
24 BMBI 01099827.1	ACCAGAATGTGCAAGTTC	CCA
25. BMBL01105564.1 1	AACAGCACGTACATGTT	ATA
26. BMBL01105564.1 2	CACACCATGATTTT	
27. BMBL01043776.1 1	AACAGGGCATGCATGTTC	CCA
28. BMBL01043776.1_2	AGCAGGACGTGCATGTT	CA
29. BMBL01043776.1_3	AG <mark>GG</mark> G <mark>G</mark> A <mark>A</mark> GTGCATGTTC	CCA
30. BMBL01043776.1_4	AGCAGCAC	CCA
31. BMBL01006510.1_1	ACCAGCAAGTGTAAGTTC	CCA
32. BMBL01061436.1_1	AGCAACATTTGCATGCTC	CCA
33. BMBL01061436.1_2	AGAAACACGTGCATGTT	CAC
34. BMBL01061436.1_3	AGTAGCACATGTATCTTC	CCA
35. BMBL01061436.1_4		ICCG
36. BMBL01061436.1_5	ACCAGAACGIGCGIGIIC	.CCA
37. BMBL01061436.1_6	ACCAGAAAGIGCAIGIIG	CCA
38. BMBL01061436.1_/	ACCAGAAAGGGCATGTTC	.CCA
39. BMBL01082959.1_1	AGCAGAACTIGCACGIAC	.CCA
40. BMBL01082959.1_2	ACCAGCATGTGCATGTTC	CCA
41. BIVIBLU1041892.1_1	AGCAGGACGTGCATGTT	
42. DIVIDLU1041092.1_2	AGCAGGACGIGCATGIT	CCA
43. BMBL01041892.1_3	AGCAGCACTIATAACTTC	CCA
45 BMBL01022471 1 1	AGCAGCAC	CTA
46 BMBL01022471.1_1	GCAGCACATGCATGTTC	CCA
47. BMBI 01022471 1 3	AGCAGCAC	CCA
48. BMBL01022471 1 4	AGCAGCACGTGCAAGTT	CCA
49. BMBL01022471.1 5	AGCAGCACGTGCAAGTTC	CCA
50. BMBL01022471.1 6	AGCAGGACCTGCATGTT	CCA

Supplementary Figure 6. Alignment of 50 monomers manually isolated from the *Dugong dugon* assembled reference genome (accession GCA_015147995.1). Part six of six.

	1	10	20	30	40	50	60	70	80	90	10	00	110	120	130	140
Consensus	MATTCACT	-GCATGAA	AGAGAAA	CAAAATACAGT	TTCTCAT	CAAGCTG	-GTTAACTTAA	TCAAC		TTCAGTTTTGA	-TAAAACA-	CAGTCC	CACACCAT	CAACAA	CTITCTCAAACAT	TCTC
Identity									-					-		
1 10 600 2010 1000067 1	-						CTT I I CTT I I		-	TTOLOTTO	TICLICI				CTTTCTCLALCAT	-
1. JACANZ010423367.1	GITCACT	GLATALA	AGAGAG	CAAAATACAGT		ICAAGCIG	GITAACTTAA	TCAAC	AAGTT	TICAGITIIG	TAGAACA		GACACCA	CAGCIA	CTTTCTGAAACAT	TOTO
2. JACANZ010389660.1	ATTCACT	GCATGAA	AAGAAAG	CAAAATGCAGT	TTCICAT	CAGCCA	GTTAACTAA	CAAC	TGAAGTT	TTCAGITTIG	TACAACAA	ACAGICA	GACAGCAC			TCTC
3. JACANZ010311924.1	ATTCCCA	GLATGAA	AGGAAGG		STICACATI	I CAAGC I G	GITAAGITAA	GCAAC	TAAGTT	TTCAGININTGG=	TAGAACA	AAGGICC		CAGCIL	ATTICICAAACAT	TCTC
4. JACANZ010229584.1	ATTCACT	GAIGAA	AGAAAG		TTCTCAT		GTTAACTTAA	TGAAG	TTAAGTT		TAGAACA	GCAGTCC	CACCAC			TCTC
5. JACANZUTUT59291	ATTRACE	CONTOPA	ACACACAC		TTCTCAT		GTTAACTTAA	TCAAC	TTAAGTT	ATCACTTETCA	TAGAACA	AGAGICC			CTITCTCAAACAT	TCTC
6. JACANZUTUT59291	ATTCACE	GCATGGA	AGAGAGAG	CAAAATAGAGT	TTCTCAT	AACTG	ATTAACTTAA	TCAAG			TAAAACA	CAGICG	CACACITAT			TCTC
7. JACANZUTUTZZU94	ATTCACT	ACACGAA			TTCTCAT	CAAGETG	GTTAACTTAA					CAGI		CACCEA		TCTC
0. JACANZ010122094	ATTCACT	CATGAA	ACACAAAA		TTCCCT		GTTAACTCAA	TCTAC	TGAAGTT	TCAGTETTC	TAGAACA		AGGACCAC			TATC
10 IACANZ010120054	ATTORCE	ACATGAA	AGAGAAAA	CALATCOACT			CTCAACTCAA	CCAC	TCAAGTT		TAGAAGA	AATGTCC			CTTTCTCAAACCT	TCTC
11 JACANZ01012005	ATTCACE	ACOTGAA	AGAGAAAA			ICAAGCTG	CTTAAC AGAA	TCTAC	GAAGTT	TTCAGTTTTG	TAGAAGAA	AGAGTCC			CTTTCTCAAACGT	TOT
12 JACANZ01007857	ATTCAC	CATGAA	AGAGAAAAG			CAAGCTG	CTTTTCTCAA	TCTAC	TGAAGTT	TTCAUTTTTG	TAGAGGA	AGAGTEC				TCTC
12 JACANZ01007657	ATTCAC	GTCTGAA	ACCAAAAA			CAAGCTG		TACAC	TGAAGTT	TTCAGTATTC	TAGAAGA	ACAGTEC			CTTTCCCAAATGT	тстс
14 JACANZ010000488.1	ATTCACT	ACCTGGA	AGAAAAA			ICAAGCTG	CTTAACTGAA	TCTAC	TGAAGGT	TTCAGT	TAGAAGAA	ACAGTCC	CACACCAC	AGCIG	CATTACAAATGT	TCAC
15 IACANZ010507883 1	ATTCAC	GTATGAA	ATAGAAAG	CAAAATACAGO	TTCC-T		CTTTACTGAA	TCTA			-CTGAAGA	AGAGT	CCCACCAC		TITECTCAAATGT	TETETE
16 JACANZ010255388 1	ATTCAGT	GCATGAA	AAGAAAG	CAATATGCAGTO		CAAGCTG		ACAAC	TTAAGTT	TTCAGTGTTG	TAGAACA	AGCGTCC		GGCTG	CTITCTCCACCT	TCT
17 IACAN7010245123 1	ATTCACT	GCAAGAA	AGAGAAAG	CAAAATACAGTO		CAAGCTG	GTTAACTTAA	TCAG-	TTAAGTT	TTCAGTTTTGA	TAGAACA	ACAGTOC	CACACCA	GTACTA	ATTTCTCAAAC	TCTC
18 JACAN701012927	ATTCAAT	GCTTGAA	GAGAGAG	AAAATACAGTO	TTGICCT	CAAGCTG	CTTAAC		TTAAG	TCACTTTTGA	TCAAACA	GGTCC	CACACCAT	ACAA	ACTITICTCAAGGGT	TCTC
19 IACANZ01012927	ATTCACT	GCATGAA	AGAGAGAG	CAAACTGCACTO	TTCTCAT	ICATGCTA	-GTTAACTT		TC	ATCAGTTTTGA	TAAAATA	CAGT	GCTCCAT	TACAA	CTITCTCAAGCAT	TCTC
20. JACANZ010754534.1	ATTCACT	GCATGAA	AT-GAGAG	CAAAATACAGTO	TTCT	CAAGCTG			GTTAAGTT	TCAGTTTTGA	TAAAA	TGGTCC	CACAGCAT		CTTCCTCAAGCAT	TCTC
21 JACANZ010175637.1	ACTCACT	GCACAGA	AGAGAAG	CAAAGTAGAGTO	TTCTCAT	CAATCTG	GTTATCTTAA	TCAAC	ATTAAGAT	ATCAATTTTGA	TAAAACA	GAGTCC	TCACACCAG		CTTTCTCAAACAT	TCAC
22. IACANZ010271070.1	ATTCACT	GCATGAA	AGAGAAA	CAGAATACAAT	TTCTCAT	CAAGTTG-	ATTAACTTAA	TCATGTT	ATTATGTT	GTCAATCGTGA	CACAGCA	CAGTCC	CACACCAT	CACAA		TATC
23. JACANZ01003562	ATTCAGT	GCAGAAA	AGAGAAA	CAAAATACGGT	TTC-CTT	GAAGAAG	GTTAACTTAA	GTTGA	TTAGGT	ATTAGTATTGA	TAAAACA	CAGTAG	TCACACCAT	CACAA	CTTTGTCAAACAT	TCTC
24. JACANZ01003562	ATTOCT	GAATGAA	AGAGAAAA	CAAAACACAAT	STTCTCATT	CAAGTTG-	GTTAACTTAA	TCAAGTT	GAATAAGTT	ATGAGTTTTGA	TAAACCA	CAGTCCT	TCAAACCAT		CTTTCTCAGACAC	TCTC
25. JACANZ01003562	ATTCACT	GTATTAA	AGAGAAAA	CAAAACACAGTO	STTCTCATI	FCAGGACG	GTTAACTTAA	TCAAATCO	GAATAAGTT	ATCAGTTTTGA	TAAACCA	CAGTCC	TCACACCAT	CACAA	SCTTTCTAAAACAT	TTC
26. JACANZ01003562	ATTCATT	CACATGGG	AAGA	-AAAAGACCGTO	STTCTCATI	TCAACCCT	GTATACTTAA	TCAAC	TGAGGTT	TTCCGTTTTGG	TAGAGCA	GCAGCCC	GACACCAC	CAGCTA	CTTTCT G AAACAT	ТСТС
27. JACANZ01010798	ACTCACT	GCAAGAA	AGAGAAAG	CAAAATACAGTO	STTCCCATT	FCAAGTTG-	GTTAACATAA	TCAAGTA	ATTAACTT	ATGAGGTTTGA	TAAAACA	CACTCCT	FCACACCAT		SCTTTCTCAAACAT	TCAT
28. JACANZ01010798	ATCTCCT	GCAGGAA	AGAGAAAA	CAAAATACAAT	T CTCATT	FCAAGTTG	GCTAAG	TCAGTTO	ATCAAGTT	ATCAGTTTTGA	TAAAACA	CAGTCC	FCACACCAT		CTCTTTAAGACAT	TCTC
29. JACANZ01006138	AATCACT	GTAAGAA	GAGAAAG	CAATATAGAGTO	STTCTCCTT	FCAAGCTG	GTTAACATAC	GITCA	TCAGATT	CTCAGTTTTGA	TCAAACA	CAGTCC	TCACATCAC		ACTATCTCAAACAT	TGTC
30. JACANZ01006138	ATTCAC	GTATGAA	GAGAAAG	CAATATACAAT(STTCTCCT	FCAAGCTG	GTTAAGTTCA	TC	AAGTT	CACCGTTTTGA	ACAAACA	CAGTCC	FCACACCAT		CTTTCTGAAACAT	TCTC
31. JACANZ010515044.1	ATCCCCT	GCAAGA	ATAGAAAG	CAAAACAATO	ATCTCATI	FCAAGCTG	GTTAACGTAA	TCGAC	TTAAGAT	ATG AGTTTTGA	TAACACA	CAGTCC	ICACAC <mark>TG</mark> T		SCTTCCTCCAACAT	TCTC
32. JACANZ01019032	ACTTACT	GTATGAA	AGAGAAA	CAGAATACAGG	STTCTCATI	ICAAATTG	GTTAAGCTAA	TCAAGTT	ATTAAGTT	ATCAGTCTTTA	TACAACA	CTGTAC	FCACACCAT	CACAA	SCTTTC-CAAACAT	TATC
33. JACANZ01019032	ATTCACT	GTATCAA	GAGAAAG	CAATATACAGT	STTCTCATI	FCAAGCTG	GTTAACTTAA	TCAAGTA	ATTAAGTT	CTCAGTTTTGA	TAAAACA	CACTCC	ACACCAT	CTATAG	GCTTTCTCAAACAT	TCTC
34. JACANZ010402190.1	ATTCACT	GCATGAA	AGAGAAA	CAAACTACAGT	TTCTCCT	FCAAGCTG	GTTAACTTAA	TACAGTT	GATTAAGTT	ATAAGTTTTCA	TCAAACA	CAGT	ICACAC <mark>G</mark> AT	CACAA	ACTTCTCATACAT	TCTC
35. JACANZ010202422.1	AATTAGT	GCCTTAA	AGAGAAA	CAGAATACAG	NNNNN	INNNN TG	GGTAACTTAA	TCTACT	GAGTAAGTT	ATCAGTTTGCA	TAAAACA	CACTGC	CACAGGAT	CACAA	GTTCCTCCTACAT	TCTC
36. JACANZ010316682.1	ATTCACT	GCATGAA	AGAAAAA	ICAAAATACAGA	STTCTCATT	FCAAGTCG	GTTAACTTAA	TCAAT T	ATTAAGCT	ATCAGTTTTGA	GAAAACA	GITTC	ICACAC GAG	CCACAA	ACTGTCTCATACAT	TCTC
37. JACANZ010418236.1	TTACT	GCATGAA	AGAGAAAA	CAAAATACAG	TTCTCATT	FCAAGTTG	GTTAACTTAA	TCAAGTG	AATGAAGTA	ACGAGTTTTGA	TAAAAC	CCTTCC	ICACACC <mark>G</mark> T	CCACAA	STITTCTCAA	TCTC
38. JACANZ010596973.1	ATTCAC	ACACGCA	AGAGAAAA	CAAAATACAG		FCAGGC-G	CTTAACTGTA	TGTAG	TGAAGTT	TTCAGTTTTG	TAGAAGA	ACAGTCC	CACACCA	AAGC	CTITCACAAATGT	TCTC
39. JACANZ010334892.1	сссспст	GCTATAAG	ACAGCGGC	AGGGCTGTAGT	GIGCTAAC		GTTCATCTAG		GTAAATT	TTCAGTTTTG	TAGAACA	ACAGTCC	TCACACCAT	CAGCTG	CTITCTCAAACAT	TCTC
40. IACANZ010280849.1	ATTCACT	GAATGAA	AGAGAAAG	CAAAATNININ								NNNNN	ICACACCAC	CAGCTG		CATC

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

		150	160	170	180	190	200	210	220	230	240	250	260	270	280
Consensus	TCAA	-TGGGATAA		GTGTCTGCO	TTTCAAGTAC	TTGAATCCCC	TG	GTTTTC	AAAGGGAACT	TGCATTACAG	GGAATGAAGG	-TACCTTT	GCATTCTAG	GAAAAACA	<mark>A</mark> -
Identity			Designed and the second se		Concernance of the second		-		State of the local division of the local div	Constraints of the local division of the loc		-1	- I- III		_
		_				-									
1. JACANZ010423367.1	TCAA	TGGGATAA	AGICTAGCGA	GTGTCTGC	TTTCAAGTAC	TICAATCCCC	:TG	GTTTTC	AAAGGAAA	TGCATTACAG	GGAACGAAGG	AACTIT	CGCATTCTAA	GGGAAACA	G
2. JACANZ010389660.1	TCAA	TGGGAT	A-ACTGICAA	GGICIGCO		TIGAATCACC	.TG	GIGITC	AAAIGGAAAIC	TGAATTACAG	GGAATGAAGG	AAGCTTT	ACATICAAA	GAAACA	G
3. JACANZ010311924.1	AAG	TGGAATAA	AACIGGCAA	GIGICIGCO		TIGAATCC	.16	CITIC	AAA-GGAACII	IGCATTACAG	GGAC I AAAGG	TAGCITI	GCATICIAA	GCAAAACA	G
4. JACANZ010229584.1	TGAA	TGGGATA	ACTGG		TTTCAAGTAC	TTI GATCCI C	TG	GTTTTC	AAAGGGAAGTT	TGTATTACAG	GGAATGAAGG	TAACTIT	CGCATTCTAA	AGAAAAGA	G
5. JACANZ010159291	TCAA	TGGGATAA	AGATTGGCGA	GTGTCTGC	TTTCAAGTAC	TTGAATCCCC	TC	G <mark>G</mark> TTTC	AAAGGGAACTT	TGCATTACA	GGAATGAAGG	TACATTI	GCATTCTAA	GAGAAACA	G
6. JACANZ010159291	TCAG	TGGAATAC	ACTGG	GTCTCTGC	TTTCAAGTAC	CTGAATCCTC	AG	GTTTTC	AAAGGGAAATT	TCATTACAG	AATGAAG-	TAACTTT	GCATTCTAG	GAAGAACA	A
7. JACANZ010122094	TGAA	TGGCAAT	AGAATATCAA	GT		GTGAAGTCCA	GAGIII	TCAGATATITIC	TATGCGAAT II	TGTGTTACCA	GAAATGAATG	AGACTIT		GCATAACA	AG
8. JACANZ010122094	TCAA	TAGGATAA	4		AGTAC	TTGAATCIGC	TG	GTTTTC	AAATGGAAATT	TGAATTATGG	GGAATGA	AAGTIT	CATATICAAA	AAAAAACAAA	AAACAAAACAG
9. JACANZ010120854	AGAG	TGGGAGAA	AG <mark>G</mark> CTGG <mark>A</mark> GA	GTG <mark>CGG</mark> GCO	AAGCACGTAT	TTGAATCCCC	TG		AAAGGGATCTT	TGCGTACCAG	GGAATGAAGG	GATTACT	GCTTTCAAA	GCCACACC	
10. JACANZ01012085	TCAG	TGGGCAAA	AG <mark>G</mark> CTGG <mark>A</mark> GA	GTGCCTGC	CTT <mark>GG</mark> AAGTA T	TTGAATCC	TG	GTTTTC	AACGGAAATTT	GGAATAAGAG	GGAAT AGAGG	TACTACT	GGCTTTCAAA	GCAAAACG	C
11. JACANZ01007857	TGAG	- GGGGA <mark>G</mark> AA	AG <mark>G</mark> CTGG <mark>A</mark> G <mark>T</mark>	GTG <mark>GAG</mark> GCO	ET <mark>AG</mark> CAA <mark>AG</mark> AT	TTGAAT <mark>G</mark> CCC	TG	GTTTTC	AAAGGGAACTT	TGCATTCCAG	GGAATGAAGG	TACTACT	GATTTCAAA	GCAAAACA	G
12. JACANZ01007857	TGAG	TGGGAGAA	AG <mark>G</mark> CTGG <mark>A</mark> GA	GAGCAGGC	CTAGCAAGGAT	TTGAATCCC	TG	GTTTC	AAAGGGAACTT	TGCATACTAG	GGAATGAATA	TACTACC	GGCCATTAAA	CCCAAGCA	G
13. JACANZ010060488.1	CCAG	TGGGATAA	AGACTGGAGA	GTGCCTGC	CTTACAAGTAC	TTGAATCCC	TG	GTTTC	AAAGGGAACTT	TGCATAACAG	GGAATGAATG	TACCTT	GCATTCAAA	GCAAAACA	G
14. JACANZ010031279.1	TOTG	TGGGAGAA	AGAGTGGAGA	GTGTCTCC	CTT <mark>GGG</mark> AG A A <mark>T</mark>	TTGAATCCCC	TG	GTTTTC	AAAGGGAAAATT	TGCATCACAG	AGTAGGAAGG	TACATT	GGCTTTCGAT	GCAAAACA	G
15. JACANZ010507883.1	TCAG	AGGGAGAA	AGACTGGAGA	GTGTCTGC	CTT <mark>GG</mark> AAG C AT	TTGAATCICC	TG	ATTTTC	AAAGGGAACTT	TGCATAACAG	GGAATGAGT-	TAC	GCATTCAAA	GCAAAACA	G
16. JACANZ010255388.1	GCAATG	GTGGGAGAA	AGGCTGGAGA	GTGCAGGCO		TTGCATCCCC	TG	GTTTTC	AAAGGGAACTT	TGCAGTACA	GGAATGAAGG	TAAGATT	GCATTCTG	GAAAAACA	
17. JACANZ010245123.1	TCAA	TGGGATAA	AGACTGGCGA	GTTTCTGT	TTTTATCTAC	TTGAATACCC	TG	GTTTTC	AAAGGGAACTT	TACATTACAG	GGAATGAAG	TACCTTT	GCATTCTAG	GAAAAAGA	G
18. JACANZ01012927	TCAA	TGGGAAAA	AGGCAGGCGA	GCGTCTCC	CTTTCAAGTGC	CCGAATCCC	TG	GGTTGC	AAATGGAACTT	TGCTTACAG	GGAATGAAGG	TACCTTT		GAAGAACA	A
19. JACANZ01012927	TCAG	TIGGATI	AGAACAGCGA	GTGTCTCC	TTTGAAGTAC	GTGAATCCCC	TG	GCTTTC	CAAGTGAACTT	TACCTTCTAG	AGAATGAAAG	TACCTTT	CGCACTCTAG	GAAGAACA	A-
20. JACANZ010754534.1	TCAA	TGGGACAA	AGACTGGTGA	GTGTTCC	TTTCAAGT	TTGAAT	TG	GTTTTC	AAAGGGAACTT	TACATTATAG	GGAATGACAG	TACCTTT	CATTCTAG	GAAGAACA	A-
21. JACANZ010175637.1	TTGG	TGGAATAA	AGACTGGAGA	TCCCTCCC	CTTTGAAGTAC	TAGAATCCTA	CG	GTTTTC	AAAGGGAACTT	TGCATTACAG	GTATGGAAGT	TACCTCT	GCATTGTAG	GAAGAACA	AG
22. JACANZ010271070.1	TCCA	TGGGATAA	AGAAAGGAGA	GTCTCTCC	TTTCAAGTAC		TGAGTIT	TAAGAT GTTTTC	AAGGGATCTT	TGTGTT-CCT	GGAATGAAGG	TATCTTT	ACAGTCTAG	GAAGAACA	AG
23. IACANZ01003562	TCAG	TGGGATA	AGAATGGAAA	GTCTCTCC	TTTCAAGTAC	TTGAAT	GGATGT	TCAGAGGTTTTC	AGGAGAACTT	TGCATTACCC	GGTATGAAAG	TACCTTT	CTCTCCAG	GAAGAACA	AG
24. IACANZ01003562	TCAG	AGGGATAA	AGAATGGCGA	GTCTCTCC		TTGAATCCCC	CAGGTIT	TCTGTTGTTTTC	AAGAGAACTT	TGCATTACCC	AGAATGAGGG	TAACTTT	GCATTCTAG	GACGAACA	AG
25. IACANZ01003562	TCAG	TGGGATAA	AGAATGGCGA	GTCTCTCC	CTTTCAAGTCC	TTGAATCC	GGGGTTT	TCTGTTGTTTC	AAAAGAACTT	TGCGTTACCC	AGAATGGAGG	-ACCTTT	GCATTCTAG	GAAGAACA	AG
26. JACANZ01003562	TCAA	GGGGTTA		GTTTCTGC	CTTT G AAGTAC	TTGAGTCCCC	TG	GTTTTC	AAATGGAACTT	TGOCTTACAA	GCAAAAATGA	TAACTTT		GGAAAATA	G
27. IACANZ01010798	TCAG	TGGGATA		ATGTTGCC	CTTTCAAGTAC	TTGAATCCG	CAGGTIT	TCAGATGTTCTG	TAACGGAGCTC	TGTGTTAGCC	GAATGCAGG	TACCTT	CGCATTCCAG	AAAGGACA	AG
28. IACANZ01010798	TCAA	TGGGATA		GTGTTCC		TTGAATCTCC	TG	GTTTTC	AAA-GGAAGTT	TGCATTACAG	GGAAGGAAGG	GGCCTTT	GCATTGTA-	GAAAATCA	AG
29. IACANZ01006138	TCAC	TGAAATAA	AGATTG	GTGTCTCC			CA	GTTCTG	AAAAGGAACTT	TGCAGTACAG	GAATGAAGG	TACGTTT	GTATTCTAG	GAAGAAC	
30. JACANZ01006138	TCAC	AGAGATAA		ACGTCTGC	TTTCAAGTAC	TTCAAGTTCC	CA	GTTCTG	AAAAGGAACTT		GAATGAAGG	TACGTTT	GTGTTCTAG	GAAGAACA	AT
31. JACAN7010515044.1	TCAA	AGGGAGAA	AGGCTGGGGA	GTGTCTCC	TTTCAAG <mark>G</mark> AC	TTGAATGCCC	TT			TGCATTACAG	GGAAACAGGG	TACCTT	GCCTACTTG	GAAGAAAA	AG
32 IACAN701019032	TCAA	TGGGATAA		GTCTTTCC	ATTTCAAATGC	TTGAATCCCC	GGATT	TCAGACGTCTTC	AAGGGATCTT	TGTGTTACCA	AGAATGAAGG	TACCTTT	CCAGTCTAG	GAAGAGCA	AG
33 JACAN701019032	TCAA	TGGGATAA		GTGTCTGC	TTTCAATTAC		TG	GTTTTC	AAAGGGAGCTT	TGCATTACA	GAATGAAGG	TAACTTT	GCATTCTAG	GAGGAA	
34 IACANZ010402190 1	TGAA	TGGGAGA	AGAATGTCAA	GCTTCTCC	TTTCAAATA-	TTGAATCCCC	GIGGIGI	TCAGATGTTTTC	GGGGGAGCTT	TGCGTTACCA	GGAATGAATG	-ICCTTT	ACATTCTAG	GAAGAACA	AG
35 IACAN7010202422 1	TGTA	AGGAAGAA	AGAATGTCAA	GTGTCTGC	TTTAAAGTAC	TTGAAT	CGGGCTT	CAGAGGTGTTC	GGGGAAAATG	TGTGCTACCA	GCAGTGAATG	TGCATTT	GCATTCTA	GAACAAGA	AG
36 IACANZ010316682 1	TGAA	TGGGAAAA	GAACGTCAA		TTTCCAGAAC	TTGAGTCCCC	TGAGILL		AAGGGAACA-			ACGITI	GCATTCTA	GAAGA	AG
37 IACAN7010418236 1	TTAG	TGGGATA	GAATGACAA	GTCTCTCC	TTTCAAGTAC	TTGAATCCC	GGATTT	TGAGAT GTTTTC	TAAGGGCACT	TGCAGTAC	AGAATGAAGG	TACCTTT	CATTCTAG	GAAGAACA	AG
38 IACAN7010596973 1		TGGGATGA	GAGTOCAGA	GTGCTGCC	TAGCAAGTAT	TTGAATCECC	TG	CTTTTC	AAAGGAATCTT	TGCATACCAG	GGAATGAAGG	TACTACT	GCTTTCAA	GCAAAACA	G
39 IACAN701033/802 1	TOTA	TGGGARA	GACTGGGGA	GTGTCTCA	TTTCAAGT	TTGAATTACC	ТА	GTTTTC	AAAGGGAACTT	TGCCTTAMAG	GGAATGAAGG	TGACATT	GCATTETAA	GAAAAAA	G
40 IACANIZ0102808491	TCAA	TGGGATA	ACTGGGGA	STGTATGC	TTTCAAGTAC		TG	GTTTTC	AAACTTAACTT	TGCTTTACAG	GGAATGAAGG	TAACATT	GCATTCTA	GAAAAACA	6
TO. JACANZO 10200049.1	1200	, JUUA I AA	- ACTOUDDA		- I I CAAUTAC	AILCEL		UTITC	www.	I SCALLINCAG	JOHN UNAGO		SCALICIA	SI VINACA	9

Supplementary Figure 7. Alignment of 40 monomers manually isolated from the *Hydrodamalis gigas* assembled reference genome (accession GCA_013391785.1). Part two of six.
	290	300	310	320	330	340	350	360	370	380	390	400	410	420
Consensus	GCAGATTGTTT	-GCCTTCAC	CTAGAGGACG	GCAGAAGC	GGTTCAG	TTTTGCTG	ACTAAACCTT	GTGTGGTTAA	CACGTGCATGT	TCCTATTO	CACTGCAAGA	AGAGAAACC	AAAAGACAG	TEGCTCTCAGTCT
Identity												_		
1. JACANZ010423367.1	GCAGATT	GCCTTA	ICTAGAGGACG	GCAGAAAC	AGTTCAG	TTTTG GAT	ACGAAATCTT	GTGTGGTTAA		TCCTATI	CTGCAACA	AGAGAAACC	AAAAAG	TTGCTCTC
2. JACANZ010389660.1	GCAGATTGTTT	GCCTTCAG	CTAGAGGACA	GCAGAAGC	GTTCAG	TTTTGCTG/	ACTAAAGCTT	GTGTGGTTAA	CATGAGCAGGT	TCCTATTO	CTGCAAGA	AGAGAAACC	AAAAGATAG	TTGCTCTCACTCC
3. JACANZ010311924.1	GCAGACTGTTT	GCCTTCAG	CTGGAGGATG		GGTTCAG	TTCTGCTG	GCTAATCCTT	GTGTGTGTITAA	LACAAGCAT			AGAGAAA <mark>G</mark> C	AGAAGACAG	ATGCTCTCAATCT
4. JACANZ010229584.1	GCAGATTGTTT	GCCTTA	CTAGAGGACA	GCAGCAGC	AGTTCAG	TTTTGCTG/	ACTAAAGCTT	GTGTGGTTAA	AGCATGT	TATTO	CTGCAAGA	AGAGAAACC	AAAAGAC	TTGCTCTCAGTCT
5. JACANZ010159291	GCAGATTGTTT	GCCTCCAG	ICTAGA GATG	GCAGAAGC	GGTTCAG	TTTCCCTG/	AGTAAATCTT	GTCTTGTCAG	ACATGAATGT	TCCTATTO	CACTGCAGAAA	GAGAAACC	AAAAGATAG	TTGCTCTCAGTC
6. JACANZ010159291	GCAGATTATTT	CCTTCAG	CTA T A A GA T G	GCAGAAG	GGTTCAG	TTTTGCTG		GTGTTGCAAG	CACGTGCATGT	TTCCTATTO	CACTGCAAGA	AAG <mark>G</mark> GAAACC		TTGCTCTCATTCT
7. JACANZ010122094	ACAGACTCTTT	CCATTCA	CTACAGGACC	ACAAAAA	AATGTAC	TTTTGCTG/	ACTAAACCTT	GTGT	CACGTGCGTGT	TTCCAATCO		AGAGAAACC		TTTCTCATATTCT
8. JACANZ010122094	GCAGATTGTCT	TO	ICTAGAGGAC	GAAGAAG	AGTTCAG	TTTTGCTG/	ACTAAAGCTT	GCATGGTTAA	CATGAGCATGT	TGCTATTO	CTGCAAGG	AGAGAAACC	AAAAGAC	TTGCCCTCAGTCT
9. JACANZ010120854	GCCCATTGTCT	GCCTTCAG	ICTAGAGGATG	GTAGAAGA	GG <mark>G</mark> TCAG	GTTTCCTA	AGTAAACCTT	GTGTGCTTAA	CACGATCCTGT	GOTICITO	CCTGCCTGA	AGACAAAGC	AAGAGCCTG	TTGCTCTCAGTC
10. JACANZ01012085	AGAGATTGTTT	GACTTCAC	CTAGAAGATG	AAAGAAGA	GGTTCAG	TTTTACTG/	ACTAAAAACTT	GAGTGCTTA-(GCGAGTGTGT	GACTCTT	CCTGCCTGA	AGAGAAA <mark>G</mark> C	AAGAGCTTG	TTGCTGTCAGTCT
11. JACANZ01007857	CAGATT	GCGTTCAG	AGAGAGACG	GCAGACGA	AGTTCAG	TTTTGCTG/	ACTAAACCTT	GTGTGCTTAA	ACACATGT	GOTITIC		AGAGAAA <mark>G</mark> C	AAGAGTTTG	TTGCTCTCAGATT
12. JACANZ01007857	GCAGATTGT	ACCTTCAG	ICTAGAGGATG	GCCGAAGA	GGTTCAG	TTT <mark>G</mark> GCTG/	ACTAAACCTT	G <mark>G</mark> GTGCTTAA(CAGGAGCATGT	GOTTOTTO	CTGCTGA	AGAGAAAGC	AAGAG	TTGCTCTCAGTCT
13. JACANZ010060488.1	GCAGATTGTTT	GCCTTCAG	CTAGAGGATG	GCAGAAGA	GGTTCAG	TTTTGCTG/	ACTAAACT	GTGTGCTTAA	CACGAGCATGT	GTTCTTC	GTGCCAGA	AGAG <mark>G</mark> AA <mark>G</mark> C	AAAAGCAG	TTGCTCTCAGT <mark>G</mark> T
14. JACANZ010031279.1	GCAGATTGTTT	GCCTTCAG	CTAGAGGATG	GCAGAAGA	GGTTCAG	TTTTGCTG/	ACTAAACCTT	GTGTGCTTAA	GGAGCGT-T		CCTGCCAGC	agagaaa <mark>g</mark> c	AAGAGACTG	TTGCTCTCAGTCT
15. JACANZ010507883.1	GCAGATTGTTG	GCCTTCAG	ICTAGAGGAAA	GCAGAAGA	GTTCAG	TGTTGCTG/	ACTAAACCTT	ATTTGCTTAA	GTGAGCATGT	GOTTATA	C-CTGCCAGA/	AGAGAAA <mark>G</mark> C	AAAGCA-AG	TTGCTGTGAGTCT
16. JACANZ010255388.1	GCAGGTTTTT	GCCTTCAG	ICTAGAGAAGG	GTGGAAGC	GGATCAG	TTTTGCTG/	ACTATATCCT	GTGTGGTTAA	ACGAGCAGGT	TATTO	CTTCTAAGA	AGAGAAACC	AAAAGACAG	TTGCTCTCATTCT
17. JACANZ010245123.1	GGAGATTATTT	GCCTTCAG	CTAGAGGACC	GCAGAAG	GGTTCAG	TTTTGCTG/	ATTAAACCTT	ATGTGGTTAG	ACGTGTATGT	TCCTACTO		GAGAAAGC	AAAAGACAG	TTGCTCT
18. JACANZ01012927	GCAGACTGTTT	GGCTTCAG	ICTAGAGGAGA	GCAGAAGC	GGTTCAT	TTTTGCTG	GATAAACCTT	GTGTGGTAAA	GCGTTCATGT	TTCCTATTO	CACTGCAAGA	AGGAAACC	AAAAGACCC	TTGCTCTCCGTCT
19. JACANZ01012927	GCAGATTGTTT	GCCTTCAG	ICTAGAAGACA	GCAGAAGC	GGTTCAG	TTTTACTG/	AGTAAA-CTT	GTGTGGTAAC	GCGCACGTAT	TTCCTATTO	CACTGCAAGA	AGACAAACC		TTGTTCTCCGTCT
20. JACANZ010754534.1	GCCAATTGTTT	GCCTTCAG	ICTAGAGGA	GCAGAAG	GGTTTAG	TTTTGCTG/	AGTGAACCTT	GTGTG TTA A T (GCC TGC G TGT	TTCCTATTO	CAGCAAGA	GAGAAACC	AAAAGACAG	TTGCTCTCCGTCT
21. JACANZ010175637.1	GCACATTGTTT	GCCATCAG	CTAAAGGACG	GCAGAACC	GGTTCAG	TTTTGCTG/	ATTAAACCTT	GGGTTGTAAC	CACCTGCATGT	ITGCTATT	CACTGCAAGA	AGAGAACCC	AAAAGACAG	TTGCTCTAAGTCT
22. JACANZ010271070.1	GCAGATTGT	GCCTTCAG	ICT GA AAGACG	GCAGAAGC	TCTGCAG	TTTTGCTG/	ACTGAAACTT	GTGT	ACATGCAT-T	TTTCTATC	CTGCAAGA	AGAGAAACC	AAAAGACAG	TTGCTCTCAGTCT
23. JACANZ01003562	GCAGATTGCTT	AGCATTCA	GTGAAGGATG	GAAGAAG	GGTGCAG	TTTTGCTG/	ACTAAGCCTT	GTGTTGTAG	ACGTGCATGT	TTCC AGTTC	CACTGCAAGA	AGAAAACC	AAAAGGCAG	TTGCTCTCAGTCT
24. JACANZ01003562	GCAGATTGCTT	GCAATCAA	GAGACGTACA	GCAGAAGC	AGTACGGTTTT		ACTAACCCTT	GTGTTGTAAA	CACTTGCATGT	TCCAGTTO	CACTGCAAGA	AGAGAAACC	AAAACAGAG	TTGCTCTCAGTCT
25. JACANZ01003562	GCAGATTGTTT	GCCACTCAG	CTAAAGGACG	GCAGAAGC	GTGCAG		ACTAACCCTT	GAGTTGTAAG	ACATGCATGT	TTCCAATT	ACTGAAAGA	AGGAAACC	AAAAGGCAG	TTGCTCTCAGTCT
26. JACANZ01003562	ACAGATTGTTT	ACCTTCAG	CTAGAGGATA	GGAGAAGC	AGTTCAG	GTTTGCTA	ACTAATGCTT	GTGTGGTTAA	ATGAGCGTGT	TTCCTATTO	CTGCAAG	AGAAAACC	TAAAGGCTC	TTGCTGTCAGTCT
27. JACANZ01010798	GCAGATTGTTT	CCAGTTCAG	CTAAGGGACG	ATAAAAGC	AGTGCAG	GTTTGCTG/	AGTAAACCTT	CTGTTGTGAG	ACTTGCATGT	TTCCAATA	CACTGAAAGA	AAAGAAATC	AAAAGATGG	TGCTCTCAGTCT
28. JACANZ01010798	GCAGATTGTTT	CAGCTTCA	CTGAAGCATG	GCAGAAGC	AGTTCAG	TTTTGCTG/	ATAAATCTC	GTGTTGTAG	CACGTGCATGT	CGTTATTO	CACTGA/	AGAGAAAC	AAAAGTCAA	TT-TGTCACTCT
29. JACANZ01006138	GCAAACTGTCT	GCCTTCC	TTTAAGCACG	GCAGAAGC	GGTTCAG	TTCTGCTG/	ACAAAACATT	GTGTTGTTAC	CATGTGCATGT	TTTTTTTT	CACTGCAAAAA	AGGAAACC	AAAAGACAG	TTGCTCTCACTCT
30. JACANZ01006138	GCAAACTGTCT	GCCTT-AG	ICTA AGCACG	CTAGAAGC	AGTTCAG	TTATGCTG/	ACAGAACATT	GTGT	ATGTGCATGT	TATCTTO	CACTGCAAGA	AGAGAAACC	GAAAGACAG	GG GCTGTCACTCG
31. JACANZ010515044.1	CAGATT	GCCTTCAG	ICTA AAGCACG	GCAGAAGG	GTTCAG	TTTTGATG/	ACTAAACCTT	GTGTTGTAG	ACGTGTATGT	TTCCTGTT	AGTGCAAGA	AGAGAAAC	AAAAGACAG	TTGCTCTCAGTCT
32. JACANZ01019032	GCAGATTGTCC	CATCTCAAG		GCAGAAGC	GGT <mark>G</mark> CAG	TTTTG <mark>AG</mark> G/	AGTGAACCTT	CTGTTGTAAG	CACAGGCATGT	TTCCTATTO	CACTGTAAGA	AGAGATACC	AAAAGACAG	TT CATGC CAATCT
33. JACANZ01019032	- GATTGTTT	AACCTTCAG	ATA		TTOTG	TTTTTCTG/	ACTAAACCTT	CTGTTGTAAC	CACGTGCATGT	TTCCTATTO	CACTGCAACA	AGAGAAAC	AAACAG	TTGCTGTCAGTCT
34. JACANZ010402190.1	GCAGACTGTTT	CCTGTTCAC	GTAAAGGACA	GCAATAGT	GATGTAG	TTTAGCTG/		GTGTTGCAAG	ACGTGTATGT	TTCCATTO	CACTGCAAGA	AGAGAAACA	AAAGTACTG	TTTCTCTCATTAG
35. JACANZ010202422.1	GTAGATTCTTT	CCCGTTCAG	CTACAGGACG	GCAGAAGC	GTGGTTG	TTTTGCTG/	ACTAAACCTT	GTGTTGTAAT	ACGTGCGTGT	TCTTA		AGAGAAACC	AAAGACAG	TTGTTCTCAGTCT
36. JACANZ010316682.1	GCAGATTGTTT	CCCGTTCA	TGAAAGGACT	GCAAAAGC	GATCTAA	TTTTGCTG/	ACTAGAACTT	GTGCTTTAAG	ACGTGCGCGT	TCCTATTO	CACTGGAAGA	AGGAAACC	GAAA ACAG	TTTCTCTCCTTCG
37. JACANZ010418236.1	GCAGATTGCTT	GCATTCAG	CTAAAGGAAG	GCAGAAGC	GGTGCAG	GTTTGCTG		GTAT GTAG -	ATGT			AGAGAAACC	AAAGACAG	TTACTCTCAGTCT
38. IACANZ010596973.1	GCAGATT	GCCTTA		GCAGAAGA	AGTTCAG	TTTTGCTG/		GTGTGCTTAA		GCGTCTTO			AAGAGCTTG	TTGCTCTCAGTCT
39. JACANZ010334892.1	GCAGATTGTTT	GCCTTCAG	CTAGIGGAIG	GCAGAAGC	GATTCAG	TTTTGCTG/	ACTAAAGCTT	GTGTGTTGA	CCGAACATGT	TCCGATT	CTGCAGGA	AGAGAAACC	CAATGACAG	TTGCACTCCTCT
40. JACANZ010280849.1	GCAGATT	GCCTTCAG	CTAGAGGACG	GCAAAAAC	GATTCAG	TTTTACTG	ACTAAACCTT	GTGTGGTTGA	AAGACCATGT	TCCTATTO	TGCAAGA	AGAGAAACC	AAG	TTGCTCTCATACT

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

	430	440	450	460	470	480	490	500	510	520	530	540	550	560
Consensus	GGTGGA	- GAAGT GAGAA	GGATTCACT	ACTGTGTTT	GAAAATCCCTCT	GTGTTACAT	GCCCAAGAATG	TGCCTGTGCA	TTGTCTACC	GGGTAAAGAG	GACTT	тстессетт	CGAGTACGTG	AAGCCCTGCG
Identity												iniiin		
4. 11. 6. 11. 704.04000067.4								a company of						
1. JACANZ010423367.1	CAGGAA	TGAAGTGAGAA	GGATTCATT	TACTGTGTTT	GAAAATCCCT	AIGITACII	CCCAAGAAAG	GGCIGGTGCA	TIGICIACC	TGGGTAAAAAA	STITGACCTT	TCTCCCATT	CCGGTACTC	AAGCCCTGCG
2. JACANZ010389660.1	GGGGAA	TGAAGAGAGAA	AGGITICACT		GAAAATCCCTC	GILIIALAI	SCCCAAGAAAC	IGCCIGIGCA	TIGICIACC	I GGG I CAAGAG	JIIIGACIII	TCTCCCCTT	AGAATACGIG	AAGCCCTGCA
3. JACANZ010311924.1	GGGGGA	AAGIGAGAA	AGAAGI ACI	ACIGIGIII	GAAAAICCCICI		CCCAGGACAG	IGCCIGIGCA	TIGIATACC	IGGAI-AAGAO		TCTCCCATT	GAGICGIG	AAGCCCTGCG
4. JACANZ010229584.1	G	TCAATCACAA	GGATTCACT	ACCGIGITI	GATAATACTCG	GIGITACAT			TIGICIACC	I GGG I AAAGAG	CIIGACIII	TCTCCCTT		AGCCCTGCG
5. JACANZ010159291	GGIGCA	- IGAACIGAGAA	GGATICACT		GGAAATCCCTCT	GIGIIAC	SCCCAGGAC IG	TICCGIGIGCA	TICINIACC	IGGGIAAACA	TTTCACTTT	TCTCCATT	CGACTACATG	AGCICIGC-
6. JACANZ010159291	AGAGIA	GAANTGAGAA	AGGATICACT	ACIGIGITI	GAAAATCCCT	GIGITACAT	SCCCAAGAATGI		GICICCACC	IGGGIAAA-AC	JIIIGACIII	TCTCCCCTT	CGAGCACATG	AAIGCCIGC
7. JACANZ010122094	GGI		AGGATTAACT		AAAAATCCCTCT	GIGITACAT	SCCCGAGAATGT	TGCCTATACA	TICICCACC	I GGA I AAAGAA	ATTIAACTIT	TCTCCCATT	AGACTACATG	AAGCCATGCG
8. JACANZ010122094	GGGGAA	GAGGI GAGAA	GGATICACT		GAAAATCCCT	CTCTTACAT	CCCAAGAAAGI	TGCCTGTGCA	TIGICIACC	I GGG I AAAGG		TCTCC	GAAAGIG	AAGCCITGIG
9. JACANZ010120854	AG-GG	TGAAGGGGGAAA	GACICACI	CIGIGITI		GIGITACAT	SCCTAAGGAAGI	TGCCTGTGCA	TIGICIACO	GGATACAG	AIGACGII	TETEE		AAGCCATCA
10. JACANZ01012085	GG-GGA	TGAAGGGAGAA	GATTCACT		GAAAATCCCTCT	GIGIICCAT		TGCCTGTGCA	TIGICIACC	I GGAAAAAGI (TATGACTT	TCTCCC	CAAGTAGTIG	AAGCCGCTCA
12 JACANZ01007857	GG-GGA	TGAAGGGAGAA	GATTCACE		GAAAATCCCTCT	GIGITACAT		TGCCTGTGCA	TTGTCTACC	I GGA I AAAG	TATGACGTT	TETEE	CAAGTAATTG	AAGCOTTA
12. JACANZUTUU/85/	GG-GGA	TGAACGGAGCA	GATTCACT	CIGIGITI	GAAAATCCCTCT	GIGITACAT		TGCCTGTGCA	TIGICIACC	I GGA I AAAG	AIGAIGII	TCTCCC	GAAGTAG TG-	AAGCC
13. JACANZ010060488.1	GGGGGA		GGATICACT	ACIGIGITI	GAA-ATCICTC	GIGITACAT		TGCCGGGGGCC	TIGICIACC	I GGG I AGAGG	STITGACTIT		GAGTAGGIG	AGCCOTIGCA
14. JACANZ010031279.1	GG=GGA	TEAL CTEACA	GATICACT	ACIGIGIII	GAAAATCCCTCT	AIGIICCAI			TIGICIACC	I GGAAAAGI (AIGACIAI	AICCOLI	AGIAGIIG	AAGCCAT
15. JACANZ010507883.1	GGAGGA	TGAAGTGAGAA	GGAGICAC		GAAAAI	GGTTACAT		TIGCCICIACA	TIGICICCC		AICACGII		CAAGIA	AAGCCATCA
16. JACANZ010255388.1	GGGGGA	TGAAGTGAGA	GGATTCACT		GAATATCCCTCT	AIGITACAT	SCCCAGGAAA	TECTETEA		I GAGTAAAG	TITGAATTI	TETEEE	CGAGTACAAG	AAGCCCTG-G
17. JACANZ010245123.1	GGGGGA	TGAACTAAGAA	GGATTCACT		GAAGATCCCCCT	GIGITACAT		TGCCTGTGCA	TIGICIAIC	I GGG I AAAGA	TTTGACTTT	TCTCCCGTT	-GGGTA GTG	AAGCCCTG
18. JACANZ01012927	GGTGGA	TGAAGTGAGAA	GGATTCACT	TACTGTGGTT	GAAAACCCCTCT	GGTTATAT	GCACAAGAATGI	TGCCTGTGCA	TTETETACE			TETEE	CGAGTACGTG	AAACCCTGCG
19. JACANZUTUT2927	AGTGGA	TGAAGTAGGAA	GGATTCACT	ACTGIGATI	GAAAATCCCTCT	GIGITACAT	GCCCAAGAATGI	TIGCCTGTGCA	TIGICIACC	I GAGTAAAGG			AGAGTACGTG	AAGCC-TGIG
20. JACANZ010/54534.1	GCTGGA	TGAAGTGAGAA	GGATICAT	ACTGIGGII	GAAAATCCCTCT	GIGITACAT	SCCCAAGAA I GI		TIGICIACC	TGGGTGAAGAG	TTTGACTT	TCTCCACT	CAGTACATG	AGGCCCTGCA
21. JACANZ010175637.1	GGTGAA	TGAAGTGGGAAA	AGGATTACT	ACIGIGITI	GAAAAI	GIGITACAT		TOCCTO	TICIACACC	I GGG I AAGGAC	STITGACTI	TCTCCCCT		AAGCCCTGCG
22. JACANZ0102/10/0.1	GGGGAA	TGAAGTGAGAA	GATTCACT	ACIGIGITI	GAAATCCCTCT	GAGIGAAAT	SCCCICGAACGI	I I GCC I GGGCA	TTOTOGCCC		THIGAGITT	TCTCCCGTT	GAGTAGGTG	AAGCCCTGCG
23. JACANZ01003562	GGTGGA	TGAAATCAGAA	GGATTCACT	ACIGIII	GGAAATCCCT	GIGITACAT		AGCC-AIGC	TTOTOCAG	I GGG I AAAGAG	TITGACTI	TETECECT	GAGCAAGTG	AAGCCCTGCG
24. JACANZ01003562	GATGGA	TGAAGTCAGAA	GATICACI	ACIGIGITI	GGAAATCCCTCT	GIGITACAT	SCCCAALAATGI	TECCATAC	TTTTTCACC	TGAATAAAGA	TTTGACTTT	TETEE	CGAGTACGTG	AAGCICTGCA
25. JACANZ01003562	GCTGGA	TGAAGTGAG	GATGCACT	ACGIGITI	GAAAATCCCTCT	GIGITACAT		I I GCC I G I GCA	TTCTCTATC		JIIIGACITI	TCTCCCATT	CGAGTACGTG	ACCCTGCG
26. JACANZ01003562	GGGTAA	TGAAGTGAGA	GGAGTCACT		GAAAATCCCTCT	TGTTACAG	SCCCAAAAA TGT	AGCATGTACA	TTGTCTAC			TETEEEET	CGAGTACCTG	AAGCCCTGCG
27. JACANZ01010798	GGTGGA		GGATTCACT		GAAAAT	CIGIICAI	SCCCAAGAATGI	TGCCTGTGCA	TTCTCCACC			TCTCCCCT		AAGEECTGCA
28. JACANZ01010/98	CCCCCCA	TGAAATGAGAA	GAGTCACT		GAAAATCIGICI	GIGITACAT		TECCTATERA	TTGTCCACC	TGGATAAAAG	TIGGATATI		CCCCTATCTC	AAGTACTGCG
29. JACANZ01006138	GGGGGA	TGAGGICAGAA	GGACTCACT		GAAAATCICTCT	GIGIAACAT		TCCCTATCCA	TTETCEACC			TETECC	GGCGTAGTG	AAGCCCTGCA
30. JACANZ01006138	GGGGGA	TGAAGTAAGAA			GAAAATCCCACT	GIGITACAT		TGCCTATGCA	TOTOTOTO			TETEEC		AAGCCCTGTG
31. JACANZ010515044.1	GGTGGA	TOTAGTOTOAGAA	AGATTCACT	ACTGTGTTT		GIGITACAT		TGCTATCCA	TTGTCTACC		TITGOGTTT	TCTCC	CCACTACATC	AACCCTAG
32. JACANZO1019032	GGTGGA	TGAGTCAGAA	GGATTCAGT		CAAAATCCCTCT	GIGITACAT	SCCC AGAATGI	TGCCTATGCA	TTGTCLACC		TTTCACTT	TCTCCCT	CGAGTACATG	AAGECCTACA
33. JACANZ01019032	GGTGGA	TGAACTGAGAA	AGGATICACT	ACTOTOT	GALLGICCULL	GIGITACAT		TECCTOTOAA			TTCACTT	TCTCCCAT	AGAGTACGTG	AAGCONACTA
34. JACANZ010402190.1	GGTGGA		ATTACT		GAAAATCICTCT	GIGITACAT		CCCTATACA	THETCHALC		JII GAACIII	TCTCCCAT	CGAGTACATG	AAGCCAAGTA
35. JACANZ010202422.1	GCTGAA	AGAAATGAGAA	GGATICACT	ACCATGIAI	GAAAATCCCTCT	GIGITAAGIO		GCCTATGA		I AGG I AAAGAG	ACTI	TCTC	CGARTAGGIG	AAACCIGCA
36. JACANZ010316682.1	GGGGGA	AICAGIAAGAA	AGATTAACT		AAAAATCCCT	GIGITACAC	GCCAGAATGI	TGCATATACA	TTETE	I GGG I AAAGAG		TETEEE	GGAT TACATG	AAGCCATGCG
57. JACANZUTU418236.1	GGIGGA	TCAAGACAGAA			GAAAAA	GIGITACAT		TGC-TATGCA	TTCTCTAC		TATGACITI	TCTCCCGI	AGAG ACGIG	AAGCCCTGCA
36. JACANZU 105969/3.1	GGGGGGA		CONTROLOT		GAAAATCCC CT	GTATTACAT			TTGTCTACA		TTTCACUT		CAAGIAGI IG	AGCCCTCC
39. JACANZU10334892.1	GGGGGA	TGAAGIGAGIA	GGATTGACT		GAAAATCCC-CT	GIATTACAT			TTGTGTACC	GGGTAAAGA	TIGACITI		AGAGIOCGIG	AAGCCCTGC
40. JACANZ010280849.1	GGGGAA	IGAAGIGAGAA	AGGATICACT	ACIGIGCI	GAAAACCICTCT	GILLICAT	GAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG	IGCGIGIACA	IIGICIACC	I GGG I AAAGAO	GAATTA		CGAGIACCIG	AAGGCCIGIG

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

	570	580	590	600	610	620	630	640	650	660	670	680	690	700 710
Consensus	GGGTTTAC	AGCTTTC	CGGGAGGCACT	CTGTGCTGCG	GGCTGAGAAT	GCEG-CCTT	ACAGACGGTA	CTCTGAAGGAG	AAAAGAAAG	G-TTTCTCC	GTTCTGGT	ATACGACAGCA	GCAGAGMTGTA	GTTTTGCTAA
Identity	I-III-				and a little state					_		La serie	and the second	
1. JACANZ010423367.1	GGTTTAC	CAGCTTTC	CGGGAGGCATT	CTGTGCTGC	CGGCTGAGAAT	GC <mark>G</mark> G-CCTT/	ACGGACGGTA	TCCCTGAAGGAG	GAAGACAAAGCG	G-TTTCCCC	ATTCTGGT	ATATGACAG	GCAGGGCTGTA	G-TTTTGCTAA
2. JACANZ010389660.1	GAGTITA	AGCTTTC	C-GGAGGCACT	CTGTGCTGCG	GGCAGAAAAT	GCTG-CCTT/	ACAGACGGTA	ICGCTGAAGGAG	AATAGAAAGC	CG-TTTCTCO	GTTTCTGGT	ATAGGACAGC	GCAGCATTGTA	
3. JACANZ010311924.1	GGGTTTAC	AGCTTT	CAGGAGGCACT	CTGTGCTGTG	GGCCAAGAAT	G G CCTT	ACAGAGGGTA		AAGAGAGCA		ATCCTGGT	ATACAACAG	GTAGCGCTGTA	G-TTTTGCTGA
4. JACANZ010229584.1	GGGTTTAC	AGCTTTC	TAGGAGGCACT	CTGAGCTGCA	GGC GGAGAAT	GCCT CCTT/	ACAGACCTTA	CGCTGAAGG-G	AAGAAAAGCA			ATACGACAGCA		G-TTTTTCTAA
5. JACANZ010159291	GGGTTTC		GGGAGGCACT	CTGTGCTGTG	AGCGAAGAAT	GCAG-CCTT/	ACAGCGATA	ICGCTGAGGGAG	AAAACAAAGCG	CA-TITCICC	ATTCTGGC	ATACGACAGCA	AGAGAGATGTA	G-TTTTG TA AA
6. JACANZ010159291	AAATTTAC	AGCITIC	CAGGAGGCIAT	CTGTTCTGAA	GGCTGAGAAT	GCAG-CCTT/	ACAGATGGTA	ICTGTGAAGCAG	AAAACAAAACG	G-TTTCTGC		ATGCIGCAGCA	ACAGAAATGTA	G-TTTTGCAAA
7. JACANZ010122094	GAGIIIAC		CAGLALGCALT	CIGICIIACC	AGCCAAAAAIG	G-CCII	ACAGACCGAA		AAAACAAIGIG		TICIAGE	ACTIGAATOO	CCAGGIIIICA	AAIGIIIICIAI
8. JACANZ010122094	GGGTTTAC	AGCTITC	GAGGCATT	CTGTGCTG	GGC GGAGAAT	GC G-CATG	ACAGAC	ICCCTAAAGGAG	AAGACAAACCG	CA-TTTCTCC	ATTCTGGT	ATATGATAGC	GCAGCGTTGTA	G-TTTTCCTGA
9. JACANZ010120854	GAGTITAC	AGCATTO	GGAAGGGATT	CTGTGCTGCG	GGCTGAGAAT		ACAGACTA	TCTGAAAGGAG	AAGAGAAACCG	CG-TACCTCC	CGTTCTGGA	ACACGAACGAC	GCAAAACTGTA	G-TGGTGGAGA
10. JACANZ01012085	GGGTTTC	AACATT	GGLAGGCALT	GINIGCIGCA	LGGC TGAGAAT		ACAGACGGTA	GCTAAAGGAG	AAGAGAAACTG		ATTCTGGA	ACACGAATGA	ACAAAACCGTA	IG TGCTGA
11. JACANZ01007857	GGGTTICC	AGCATIC	CGGGAGGCA	CIGIGCIGC	_GGC I GAGCA I		ACACACGGTA	I CGC TAAAGGAG	AAAAGAAACCG	CG-TTTCTCC	ATTCIGGA	ACGIGAAGIAC	GCAAAACIGI	G-TICIGCILA
12. JACANZ01007857	GGGTTTC	AGCATTC	CGCGAGGCATT	CTGTGCTGCG	GGCTGAGAAT	CG-CTTT	ACAGA	CACTAAAGGAG	AAGAGGAACCG	G-TTTCTCC	ATTCTGGA	AT CGAAGGAC	GAAAACCGTA	G-TTGTGCTGA
13. JACANZ010060488.1	GGGTTTAC	AGCTITO	GGGAACCATT	CTGTGCTGGG	AGCTGAGAATO	GCCG-CCTT/	ACAGACGGTA	CCTGAAGGAG	GAGATAAAGAG	G-TTTC-CC	ATTCTGGT	ATACAACAGAC	GCAGCGCTGTA	G-TTTGGCTAA
14. JACANZ010031279.1	GGGTTTAC	AGCATIC	CGGGAGGGATT	CTGTGCTGTG	_GGTTGAGAAT		ACAGATGGTA	I CCCTAAAGGAG	AAGAGAAACGG	ATTICICO	CITCIGAA	ACACGAC	GCAGAGCCGC	G TITIGOOGA
15. JACANZ010507883.1	GAGIIIAC	AGCITIC	GAGGAGGCA	CIGICCIGCG	_GGCCAAGAAT	GCCG-CCII	CAGACGGTA	CACAGAAGG	GAGIGAAACCG	G-THCICC	CGITAIGGI	ATAGGACAGC	GIGGAGCCACA	G-TITIGCIGA
16. JACANZ010255388.1	GGGTTCAC	AGCITIC	C-CGAGCACT	AGTIGCTGCG	GGCCTAGAAC	CCG-CCTT	AGACCGTA	GCCTGAAGGAG	AAGAGAAAGCG	G-TTT TCC	GITCIGGT	ATAAGACAGC	GCAATGTTGTA	G-TTTTGCTAA
17. JACANZ010245123.1	GGGCTTAC	AGCTTT	GGGAGGCACT	CTGTGCTGTG	GGCCAAGAAT	GCTG-CCTT/	ACAGACC-TA		AGAAAG			ATATGATAGIC	GCAGIGCIGTA	GTTTTGCTAA
18. JACANZ01012927	GGITIAAC	GCTTTC	CAGGAGGCACT	CIGIGCIGCG	AGCIGAGAAI	GG CCTT	ACAGALGGTA	GICIGAAGICA	GAAAGAAAGCG		CGITCI	AAGACAGCA	AGAGGATGTA	G-TITIGCTAA
19. JACANZ01012927	GGGTTTAC	AGCITIC	AGGGAGACAGI	IGIGCIGCG	AGICAAGAGG		AGAGCIGGCA	GICIGAAGGAG	GAAAGAAAGCA	CG-TTTCTCC	CGIICIGGI	ATAGGIGGG	GAGAGAIGAA	GTIGIGCIAA
20. JACANZ010754534.1	GGCTTTCC	AGCITIC	CAAGAGGCACT	GICIGCIGCA	GGTTGAAAAT	CGG-CCTT	AGACGGIG	GICIGAAGGAA	GAAAGAAAGCA	G-TICCICC	AGTICIGGT.	ATACAAIGGCA	AGAGAIGIA	
21. JACANZ010175637.1	GGGTTTAC	AGCITIG	CGGGAGGCACT	CTAGGCTGTT	GGCTGAGAAT	GCCA-CCTT/	ACAGACGGTA	GTCTGAAGGAG	AAAAAAGGCA	CA-TITCICC	ATTCIGGT	ATACGACAGCA	AGAGACITA	GTTTTGCCAA
22. JACANZ0102/10/0.1	GGGTTTAC	AGCITIC	GGGAGGCACT		AACGGACAAG	GCCA GCTT	ACAGACGGTA	I CGC I GAAACI O	AGGAGAAAGC		I I AIGGI	ATAGGACAGCA	AGCAG GC IGIA	
23. JACANZ01003562	AGGITTAC	AGGITTC		GIGIGCIGC	GGCTGAGAATO	GG-CCTT	GGAAGGTA	I C I C I GAAGGAG	AA-ACAALGCA		ATTCAGGT	ATACGACAGCA	AGAGAIGIC	
24. JACAN201003562	GGGTTTAC	AAGIIIC	CAGGAGGCACT	CIGIGCIGC	GGCCGACAAT		ACALAGGIA	ICICIGAGGAAG	AAAACAA		LGIICAAGI	ATACAACAG	AGAGAIGI	A TITIGCIAA
25. JACANZ01003562	GAGTITAC	GCTTTC			GCAGALAAT	JCG-CCTT	GCALACCITA	I CGC I GAAGG=0	AAGAAAAGCG	ATTICICO	ALICICAL	ATACGACGGCA	GCAG GCIGIA	
26. JACAN201003562	GGGTTTAC	AGCITIC		(GIICIGGI	ATACGAGAGC	GCAGAGETGTA	G TITIGCIAA
27. JACAN201010798	GGGTTTA	GCTTTC	GGGGGGGCACT	CIAIGCIGCA		CAG CCTT	ACACAAAGIG		AAAAAAAGCG	AITTCICC	ATTCIGGT		AGAGAGATGAA	G-TITIGCTAA
28. JACANZUTUTU798	GGATTIAC	AGCTITC			GGC-GAGAATC	SCAG-CCTT	ACAGACAGTA	ATCTGAGGGG	AAAAGAAACCA		ALICIGGI	ATACCACAGC	G ACAGA IGGA	G-TGTTGCTAA
29. JACANZUTUU6138	GGGTTTAC	AGCTITO	GCCAGGCAGT	CIGIACIGC-	GIGGAGAAC	JCAG CIGI	ACAAACGGTA				GITCIGGI		ACAGAGITGTA	G-TTTGCTAA
30. JACANZ01006138	GGGTTTAC	ACCICIC	CG GAGGCAGT	CCGIGCIGC=0	AGGGGAGAAT		ACAGATGGTA				GITCIGGI	AAGCACAGCA	ACAGAGE IG	
31. JACANZ010515044.1	GCTTTGC	ACCITIC	GGGAGGCGCT	CICCCCCCC	GGCCAAGAAA	CAGAL TI		I C I C I GAAGGAG				ATACGACAGCA	AGAGATATA	G TTTTCCTAA
32. JACANZ01019032	GCATTIAC	AGCTITC	GUGAGCACT	CTGTGCTGCA	GCTGAGAATO		ACAGGCAGTA					A TAGGACAGCA	AGAGAGGTA	G-TTTTGCTAA
33. JACANZ01019032	GAGITIAC	AGCITIC	GGGGAACT		GCTCAAGAAT		AGAGATCITA				ATTCACCT	ACATGATGGCA	AGAGACATA	G TTTTCCTAA
34. JACANZ010402190.1	GGGTTTAC		CACTAGGCACT		LGGC I GAGAATO		ACAGACAGIA			G-THCICO	ATTCAGGT	GTAGGACGGCC	ALAGIGATGI	G TITIGCIAA
35. JACANZ010202422.1	GGGTTTAC	AACTITC			GGAGGAGAAT		ACAGACAGIA		AAAAGAAAGCA		GITCIGGI	ATACCACAGCA	AGAGAGATGTA	G TTTTCCTAA
36. JACANZ010316682.1	GGGTTTA		CAGEAGGCACT	CIGIGCIGIC	GGCTAACACT		AGGACCGIG	I I AAAGGGAG	AAAACAATICG	AG-TITCAGO	GIICAAGI	ATACGGGAGCA	ACAGAGAIGG	GTITIGCTAA
37. JACANZ010418236.1	GAGITIAC		GGAGGCACT		AGATGAAAT		ACAGACGGTA				GACAL		AGGIIGIA	ITTTGCGAA
38. JACANZU 10596973.1			GAGGCA				ACAGACGGTA						CCACTOTOT	C TTTTCCTAA
59. JACANZU10334892.1	GITTAC	AGCTITC		CIGIGCIGIG	GGCCGAGAAT			CECTGAAGGAG					GCAGIGEIGI	G TTTTGCTAA
40. JACANZU10280849.1	GGGTTTAC	AGCITTC	CGGGAAIGAII	CIGIGCIGIG	JUGC I GAGAAT(JCG CCTT	ACAAACGCTA	ICICICAAGGAG	AAGAGAGAG		GILLIGGT	ATACAGCA	AGC AGC GC IGTA	G IIIICAGA

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

		720	730	740	746
Consensus	GTAAACA	CTATGTT	AGCAGCACGT	GCATGITCO	CAT
Identity					
1. JACANZ010423367.1	GTAAACA	CTGTT	AGCAGCAAGT	ACATGTTCO	CA
2. JACANZ010389660.1	CTAAACA	TGTT	AGCAGCTCAT	GCATGTTCO	AA
3. JACANZ010311924.1	GTAAAGA	TATGTT	AGCAGCAAAT	GCAAGTTCO	CAC
4. JACANZ010229584.1	CTAAACA	CTATGTT	AGCAGCACGT	GCATGTTCO	CA
5. JACANZ010159291	GTAAACA	CTACGTT	AACAGGATGT	GCATGTTCO	CA
6. JACANZ010159291	G-AAACA	CTATGTT	ACCACAACAT	ACATGTTCO	C
7. JACANZ010122094	GGGAACT	T GCAAT	ACCGGGA	ATGAATTG	CTT)
8. JACANZ010122094	GTAAAGA	ATCTT	GGCAGCAAGT	GTAAGTAC	CAT
9. JACANZ010120854	GTAAACA	CTTTGTT	AGCAGCAAGT	GTAAGTTCO	CA
10. JACANZ01012085	TAAAAA	CTTTGTT	AGCAGCAATT	CTTAGTTCO	CA
11. JACANZ01007857	GTAAACG	T-TGTT	AGCAGCAAG	GCAAGATCO	
12. JACANZ01007857	GTAAAAA	GTTTGTT	AGCAGCAAG	GCAAGTTCO	
13. JACANZ010060488.1	CT	CTAT	ACCAGCATGT	CATGTTCO	CA
14. JACANZ010031279.1	TAAAAA	CTACGTT	AGCAACAAG	AGACGTTCO	
15. JACANZ010507883.1	GTAAAGA	CTATGGT	ACCAGCAAGT	GTAAGTTCO	CA
16. JACANZ010255388.1	CTAAAGA	CTATATT	AGCAGCAC	GCAAGTTCO	CAN
17. JACANZ010245123.1	CTAAACA	CT-TGTT	AGCAGCACAT	ACAT	CAG
18. JACANZ01012927	GCAAACA	CTATGT-		GCACCTTCO	CA
19. JACANZ01012927	GTAAACA	CTATGTT	AGCAGGACTT	GCAAGTTCO	CA
20. JACANZ010754534.1	GGAAACA	CTATGTT	AGCAGGA		
21. JACANZ010175637.1	GGAAACA	CTATGTT	AGCAGGACCT	GCATGTTCO	CA
22. JACANZ010271070.1	GGAAACA	CTATGTT	AGCAGCAC	CG TGTTCC	CA
23. JACANZ01003562	GTAAACA	CTGTATT	AGCAGAACGT	GCATGTTCO	CAG
24. JACANZ01003562	GTAAAGA	CTATGTT	AGCAAGATAT	GCGTATTCO	CA
25. JACANZ01003562	CTAAACA	CTATGTT	AGCAACATT	GCATGCTCC	CA
26. JACANZ01003562	CTAATCA	CTGTGGA	AGAAGCACGT	GCATGTTCO	CAT
27. JACANZ01010798	ATAAACA	CTATGTT	AGC <mark>G</mark> G <mark>G</mark> ACGT	GCATGTTO	CA
28. JACANZ01010798	GTAAAC	CTATGTT	AG <mark>GCAG</mark> ACGT	GCATGTTCO	
29. JACANZ01006138	GTGAACA	CTATCCT	AACATTACCT	GCAAGTTCO	TDD
30. JACANZ01006138	GCCAACA	CTATGCT		GCAAGTTCO	
31. JACANZ010515044.1	GCAAACT	CTATGTT	AGCAGGACAT	GTATGTTCO	CAC
32. JACANZ01019032	GTAAACA	CTATGTT	AGCATTCCAT	GCATGTTCO	CA
33. JACANZ01019032	GGAAACA	CTACGGT	AGCAGAGCCT	GCATGTT	
34. JACANZ010402190.1	GTTGAAA	CTGTGTT	ACCAGAAAGT	GCATGTTCO	CA
35. JACANZ010202422.1	GTAATCO	CTATGTT	ACCAGGACAT	GCATGTT	CAC
36. JACANZ010316682.1	GTGAACA	I GCT TT	AGCAGAAAGT	GCATACTCO	CAT
37. JACANZ010418236.1	GCAAACA	CTATATT	AGCAG <mark>G</mark> ACGT	GCATGTTCO	CAN
38. JACANZ010596973.1					
39. JACANZ010334892.1	CTAAACA	CTATGTT	AGCAACACAT	ATGTTC	CAA
40. JACANZ010280849.1	CTAAACA	CTATGTT	AGCAGCAGG	GCAAGATCO	AA

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

Substitutions	per site) over all s	sequence pairs wi		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	

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

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	whole genome shotgun sequence	60	68.51	3,00E-08	JABTCH010005241.1
Elephas maximus	solate icky scattold_o5, whole genome shotgun sequence	35	70.22	1,00E-07	JABTCH010002495.1
Elephas maximus	isolate Icky scattold_/51, whole genome shotgun sequence	58	65.02	1,00E-07	JABTCH010005233.1
Loxodonta africana	solate 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	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

ATTCACCCCATGAAAGAGAAAGCAAAATACAGCGTTCTCCTTCAAGCTGTTTAACTGAATCTACTGAAGTTTTCAGTT TTGTTAGAAGAAAGAGTCCGCACACTACAAGCTGCTTTCTCAAATGTTGTCTCAGTGGGAGAAGGCAGCAGAGTGA AGGCCTAGCAAGAATTTGAATGCCCTGGTTTTCAAAGGGAACTTTGCATTACAGGAATGAAGGTATTGCTAGCTTAG ATAGCCAAACAGGCAGATTGTCTGCCTTCAGCTAGAGAACAGCAGAAGAAGGTTCAGTTTGCTGACTAAACCTTGT GTGCTTAACACGAGCATGTACCTTCTTCCCAGCCTGAAAGAGAAACAAGAAGGTTCAGTTTGCTGCTCTCAGTTTGGGGATGA AGGGAAAAAGATTCCCTTTCTGTGTTTGAAAATCCCTCAGTGTTACATGCCTAAGAAGTTGTCTGTGCATTGCTG CCTTGATAAAGTGTTTGACTTTCTCCATTTGAAGTAGTTGAAGCCGTTCAGGGTTTGCAGCATTCCGGGAGGCATT CTGTGCTGCGCGGCTGAGAATACCGCTTTACAGACGGTATCGCTAAAGAAGAAGAGAAACCGCGTTTCTGCCTTC TGGAACACGAATGAGGCAAAAGCGTAGTTGTGCTGCGCTAAAAAACTTTGTTAGCAGGAAGCGCAAGTGTAAGTTCCCA >NW004444102.1 4

ATTCACCAAATGAAAAGAAAGAAAGCAAAATACAGCGTTCTCCTTCAAGCTGCTTACTGAATCTACTGAAGTTTTCAGTTT TGTTAGAAGAAAGAGTCAGCCCACCCAAGCTGCTTTCTCAAATGTTCTCTCAGTGGGAGAAGGCTGGTGAGTGCA GGCCTTGCAAGTATTTGAATCTCCTGATTTTCAAAGGGAACTTTACATACCAGGGAATGAAGGTTCTAATGGCTTTC AAAGCCAAACAGGCAGATTGTCTGCCTTCAGCTAGAGGACAGCAGCAGAAGAGGGTACAGTTTGGCTGACTAAACCTTGT GTGCTTAACAGGAGCATGTAGCTTCTTCCCTGCCTGAAAGAGAAAGCAAGAGGCTTGTTGGTCTCAGTCTGGGGATG AAGGGAGAACGGTTCACTTTCTGTGTTTGAAAATCCCTCTGTGTTACTTGCCTAAGAATGTTGCCTGTGCATTGTCT ACCTCTGTAAAGTGTATGACTCTTCTGCGTTCAAATAGTTGATACCTTTCAGTGCTTACAGAGTTCCGGTAGGGAG TGTCTGCCCAGCTGAGAATACTGCTTTACAGACGGTATCGCTAAAGGAGAAGAGAAACTGTGTTTCTGCTGTT CTGGAACACGAATAAGCCAGGCCACAGTTTGGCTGAGTAAAAAGTATGTTTGTAGCAAGTATAAGTTACC >NW004444102.1 5

>NW04444102.1 6

ATTCACTGCATGAAAGAGAAAGCAAAATACAGCATGCTCCTTCAAGCTGCTTAACTTCTTCCACTGATGTTTTCAGTT TTGGTAGAAGAAACAGTCCGTACACCACCAGATGACTTCTCCAAATGTTCTCTCAGTGGGATAAGACTGGAGAGTGC CTGCCTTGCAAACACTTGAATCCCCTGGTTTTTAAAGGGAACTTTGCATAACAGGGAATCAAGGTACTTTTGGCATT

GTTCACCACGTGAAAGAGAAAGCAAACTACAGCGTTCTCCTTCATGCTGCTTAACTGAATCTATTGAAGTTGTCAGT TTTGTTCGAAGAAAGAGTCCCCCCAACACAAGCTGCTTTCTCTAATGTTCTCTCAGTGGGAGAAGGCTGGAGAGGG CGGGCCTTGCAAGGATTACAATCCCCTGGATTTCAAAGGGAACTTTGCATTCCAGGGAATGAAGATACTTCTGTCTT TCAAAGCCAACCATGGAGATTGTTTCCCTTCAGCGAGAGGACAGCAGAAGAGGGTTCAGTTTTGCTGACTAAACCTT GTGTTCTTAACACGAGCATGTAGGTTCCTCCCTGCCTGAAAGAGAAAGCAAGAGGCTTGTTGCTCTCAGTCTGGGGT TGAAGGGAAAAAGATTCACTTTCTGGGTTTAAAAATCCCTCTGTGTTCCATGCCTAAGAAAGTTGCCTGTGCATTGT CTACCTGGATAAAGTGTATAACTTTTCTCCCTTTCAAGTAGTTGAAGCCGTTCAGGGTTTCCAGGCATTCCGGGAGGC ATTGTGTGCTGCGCGCTCTGAGAATAGCGCTTTACAGGCAGTATCGCTAAAGGAGAAGAGAAATCGCGTTTCTGCCA TTCTGGAGCACGAATGAGGCAAAACCGTATTTGTGCTGAGGGAAACCTTTGTTAGCAGCAAGTGGAAGTTCCCA >NW004441411.1 2

AATCACTGCATGAAATAGAGAGCAAAAAAACAGCGTTTTCCTTCAAGCTGCATAACTTCATCCACTGAAGTTTTCAGTT TTGCTGGAAAAAAACTTCCACACACCACCAGCTGACCTCTCAAATGTTCTCTCAGTGGGATAAGACTGGTGAGTGT GTGCCGTGCAAGTACTTGAATCCCGTGGTTTTCAAAGGGATCTTTGCATTACAGGGGATGAAATTACCTTTGCCATT GTAAGCAGAACGGTTAGATTGTTTGCCCTCAGCTAGAGGATGGCAGAAGTGGTTCACTTTTCCTGACTACATCTTGT GTGCTTAAAGCGAGCATGTAGCCTCTTCCCTGCCTGAAAGAGGAAGAGAAAAGGCAGTTGCTGTCAGTCTGGAGGGG AAGTGAGAAGGATTCAATTACTGTGTTTGAAAATCCCTCTGAGTTACATGCCTAAGAAAGTTGCCTGTGCATTGTCT ACCGGGGTAAGGGGTTTACTTTTCTCCTGTTCAAGTAGCTGAAGCCCTGCGCGTTTTACCGCTTTCCTGGAAGCAT TCTGTTCTGCGCGGCCGAGAATGCCGCCTTACAGAAGCGGTATCCCTGAAGAAGAGAAAACTGTTTTCTCCCGTT CTGGTATAGGACAGCGGCAGCGTTGTAGTTTTGCTGAGTAAAACTGTTTTAGCAGCAAGTGAAAGTTCCCA >NW00444141.1 4

ATTCACTGCATGAAAAGAAAGAAAAATACAGGGTTCTCATTCAGGCTGGTTAAGTGAATCAACTGAAGTTTTCAGTTT TGGTTGAACAAACAGTCCTGACACCACCTGCTACTTTCAGAAACATTCTCTCAATGGGGTAATACTGGTGAATGTCT GCCTTTCAAGTACTTGAATCCCCTCGTTTTCAAAGGGAACTTTGCATTACAGGGGAATGAAGTTAACTTTCGCATTCTA AGGAAAACAGGCAGATTGTCTTCAGCTAGACAACGACAGAAGCAGGTTCAGTTTTGCTGACTAAAGCTTGGTGGTTA ACACGAGTATGTATCCTATTCCCTGCAAGAAAGAGGGAAGAAAAAGACAGTTGCTCTCAGTCTGGAAAATGAAGTGA GAAAATTCACTTACTGTGCTTGAAAATCTCTCTGTGTTAGATGCCCAAGAAAGTGGCCTGTGCATTGCCTACCTCCGG TAAAGGGTTTGACTTTCCCCATTCGAGTATGTGAAGCCCTGCGGGGGTCTACGGCTTTCCAGGAGGAGGAGGCAGTGTGG CTGCGCACCCGAGAATGCCGCCTTACAGACGGTATCCCTGAAGAAGAAGAAGACAGTACAGCACGTTTCTCCCGTTCTGGT GTATGACAGCAGTAGAGATATAGTTTTGCTAAGAAAACACTATGTTAGCAGGACCTGCATGTTCCT >NW00444141116

ATTCACTGCATGAAAGAGAAAACAAAAATACAGTGTTCTCATTCAAGCTGGTTAACTTAATCAATTTGATTATCAATTTT GATAAAGCGCAGTCCTCACACCATCTACAAGCGGTCTCAAACATTCTCTCCATGGGATAAGATTTGTGAGTTTCTCC GTTTCAAGTACTTGAATCCCCCCGGTTTTCCAAGGAACTTTGCATTACAGGGAATGACTGTTCCTTTCGCATTCTAGG AAGAAAAAGTGCAGATTTGTTCCCCTTCAGCTAAAGAACGGCAGAAGATTTTCAGTTTTGCTGACGAAAACTGGTGT TGTAACCACGTGCGTGTTGCCTATTCACTGCAAGAAAGGGAATCCAAAAGACAGTTGCTCTCAGTCTGGTACATGA AATGAGAATGATTTACTTACTGTGTTTGAGAATCTCTCTGTTTTTCATGCCCAGGAAGGTCGCCTATGCATTCTCCAC CTGGGTAAAGAGTTTGATTTTTCTCCTATCCGAGTGCGTGAGCTCTGCGGGGGTTTACAGTTTTCTGGGAGGCACTC CGTGTTGCACAGCCAAGAATACTGCCTTACAGGTGGTATGTCTGAAGGAAAAAGCAAAGCACGTTTCTCCCCTTCT CCCCAGCAGTAGAGATGTAGTTTTGCTAAGGAAACACTATGTTAGCAGAACCTGCATGTTCCCA >NW00444141.1 7

ATTCACCACATGAAAGAGAAAACAAGCCACAGCGTTCTCGTTCAAGCTGCTTAACTGAATTTACTGAAGTTTTCAGT TTTGTTAAAAGAAAGAGTCCGCCCAACACAAGCTGCTTTCTCCAATGTTCTCTCAGAGGGAGAAGGCTGGAGAGGG CAGGCCTAGCAAGTATTTGACTACCCTGGTTTTCAAAGCGAACTTGGCATAGCAGGGAATGAAGGTACCACTGGCT TTCAAAGCCCAACAGGCAGATTCTCTGTCTTCAGCTAGAGGACAGCAGAAGAGGTTCAGTTTGGCCGATTAAACCT TGTGTGCTTCACAGGAGCATGTAGCTTCTTCCCTGCCTGACAGAGAAAGCAAGAGGTTCAGTTGCTCTCAGACTGGGG ATGAAGGGAGAACGATTCACTTTCTGTGTTTGAAAATCCCTCTGTGTTACATGCCTTAGAAAGTTGCCTGTGCATTG TCTACCTGGATAAAGTGTATGACTTATGCCATTTCAAGTAGTTGAAGCCGTTCAGGGGTTTGCAGCATTCGGGAGG CATTCTGTGCTGCGTGGCTGAGAATACAGCTTTAAAGATGGTATCGCTAAAGGAGAAAACCGCGGTTTCCCC ATTCTGGAACACGAATGAGGCAAAACCGTAGTTGTGCTGCGGTAAAAACATTGAGCAGCAGCAAGTAAATTCCCA >NW00444013.1 2

ATTCACCACATGAAAGAGAAAGCAAAATACAGCGTTCTGCTTCAAGCTGCTTAACTAAATCTACTGAACTTTTCAGTT TTGTGAGAAGAAAGAGTCCGCCCACCACAAGCTGCTTTCTCAAATGTTCTCTCAGTGGGAGAAGGCTGGAGAGTGC AGGCCTAGCAAGTATTTGAATCCCCTGGTTTTTAAAGCAAACTATGCATAGCAGGGAATGAAGGTACTACTGGCTTT CAAAGCCAAACAGGCAGATTGTCTGCCTTCAGCGAGAGTATAGCAGAAGAGGTTCAGTTTTGCTAATTAAACCTAGT GTGCTTCAAAGGAGCATGTAGCTTCTTCCCTGCCTGAAAGAGAAAGCAAGAGGCTTGTTGCTTTCAGTTTGTGGATG AAGGGAGAATGATTCACTTTCTGTGTTTGAAAATCCTTCTGTGTTACATGCCTAAGAGAGGTTGCCTGTGCATTGTCT ACCTGGATACAGCATATGACTTTTCAACCTTTCAAGTCGTCGAAGCAGGCTTCAGGGCTTCCAGCATTCCGGGAGGCA TTCTGTGCTGCGCGGCTGAGAATACTGCTTTACAGACGGTATCGCTAAAGGAGAAAAGAAAACCGCGTTTCCCCAT TCTGGAACACGAGTATGAAGCTTCTTCACTGCCTGAAAGAGAAAGCAAGGAGCTTGCTGCCGCGCGTTTCCCCAT >NW00444013.1 3

ATTCACTGTATCGGAGAGAAAGCAAAATACAGCGTTCTCCTTCAAGCTGCTTAACTGAATCTACTGAAGTTTTCACTT TTGGTAGAGAAACAGTCCGCCAAACACCAGCTGCTTTCTCAGAAGTTCTCTCAGTGGGAGAAGACTGGGAGTGCCT GCCTTGCAAGTATTTGAATCCCCTGGTTTTCAAAGGGAACGTTGCCTAACAGGGAATGAAGGTAATTTTGGCATTCA AAGCAAAACAGACAGATTGTTTGCGTTCAGCTAGAGGAGGGCAGAAGATGTTCAGTTTGCTGACTAAACTTTGTG GCTTAACACGAGTGTGTAGCTTCTCCAGTGCCAGAAGGAGGACAAGAGCCAGTTGCTCTCAGTCTGGGGAAGA AGTGAGAAGGATTCACTTACTTTGTTTGAAAATTTCTCTGTGTTACTTGCCTAAGAAAGTTGTCTGTGCATTGCTTAC CTGGATAAAGTGTATGACTTTTCCCCTTTCAAATAGTTGAAGCCATTCGGGGTTTACAGCATTACGGGAGGGCATTC TGTGCTTCGCGGCTGAGAATACCGCTTTACAGACGGTATCCCTATAGGAGAAGAGAAACCGCGTTTCTCCCGTTCT GGAACATGAACCAGGCAGAGGCGCAGTTTTGCTGAGTAAAGACTGTGTTAGCAGCAAGTGTTCCCCA >NW004444013.1_6

ATTCGTCGCATGAAAGAGAAAGCAACCTGCAGCATTCTCCTTCCAGCTGCTTAACTGAATCTACTGATGATTTCAGT TTTGTTAGATGAAACACTCCGCCGACCACCAGCAGCTTTCTGAATCGTTCTCTCAGTGGGTAAGACTGGATAGTGC CTGCTTTGCCAGTATTTGAATCCCCTGGTTTTCAAAGGGAACTTTGCTTTACAGGGAAAGGGGTAACATTCGCATTC

>NW004443969.1_1

ATTCAGCGCATGAGGGAGAAAGGAATATACAGCGTTCTCCTTCAAGCTGCTTAACTGAATCTACTGAAGTTTTCAGT TTTGTTAGAAGAAACTGTCCACCCACCACCAGCTGCTTTCTCAAACATTCTCTCAGTGGGAGAAGACTGCAGAGTG CCTGTCTTGCAAGTATTTGCATCCCTTGCTTTTCAAAGGGCACTTTGCATAAGAGGGAAAGAAGGTAGTTTTCCCAT TCAAAGACACAGGCAGATTGTTTGCCTTCAGCTAGAAGAAGAGGGAGAGGGGTTCTGTTTTGCTGACTAAACCTTGTG TGTGTAACGCGAGCATGTAGCTTCTTCCCTGCCAGAAAGAGAGAAAGCAAGAGCCAGTTGCTCTCAGTCTGGTGATGA ATTGAGAAGGATTCACTTACTGTCTTTGAAAATCCCTCTGTGTTACATGCCTAAGAAAGTTGCCTGTGCATTGTCTAC CTGTATAAAGTGTATGACTTTTCCCCTTTCAAGTAGTTGAAGCTGTTCAGGGTTTGCAGCATTCTGGGAGGCATTC TGGGCTGCGCGGCTGAGAATACCGCTTTACAGACGGTATCACTAAGGAGAAGAGAACTGCGTTTCTCCCATTCTG GAACACGAATGAGGCAAAAGCGTAGTTGGGCTGAGTAAAAACTTTGTTAACAGCAAGTGTAAGTTCCCA >NW004443969.1 3

>NW004443969.1 5

ATTCACCGCATGAAAGAGAAAGCAAAATACAGCGTTCTCCTTCAAGCTGCTTAACATAGTCCATTGAACTTTTCAGTT ACATTAGAAGAAACAGTCCGCCCACCACCAGCTGCTTACTCATAGGTTCTCTGTGATGGATAAGACTGGGGAGTGC TTGCCTTGGAAGTATTTAAATCCCCTGATTTTTGAAGGGAACTTTGCAAAACAGGAAATGAAAGTACATTTGGTATTG AAAGCAAAACAGATTGTTTGCCTTCAGCTAAAGGATGGCAGAAGAGGTTCAGTTTTGCTGACTAAATCTTGTGTGCT TAATGTGAGCCTGTAGCCTCTTCCCTGCCAGAAAGAGGAAGCAAAGCCAGTTGCTCTCAGTCTGGGGGACGAAGT GAGAAGGATTCACTTACTGTGTTTGAAAATCGCTCTGTGTTACATGCCTAAGAAGTTGCCTGGGGAAGGACTCACTTG GGGTAAAGGGATCCACTTTTCCCCTTCCAGTCGGGTGAAACCATTCAGTGTTTACAACATTCCGGGGAGGACTTCTG TGTTGCGCGGCCGAGAAAGCCGCTCTGCAGATGGTATCTCTAAGGGAGAAACCGCTTTTCCCGGGAGGCATTCTG GTATAGGACAGTGGCAGAGAGCAGCAGCAGTTTTGAGTGTTAGCAGCAAGTGTACCCA

>NW004443969.1 6

CTTCACTGCATGAAAGAGAAAAACAAAATACAGTGTTTTCCTTCAAGCTGCTTAACTCAATCTACTGAAGTTTTCAGTT TTGTTAGAAGAAACAGTCCGCCCGCCACCAGCAGCTTTCTCAAAGGTTGTCTCAGTGGGATAAGACTGGAGAGTGC CTGCCTTGCACATATTTGAATCCCTGGTTTTCAAAGGGAACTTTGCATAACAGACAATGAAATTACCTTTGGCATTCA AAGCAAAACAGGCACATTGTTTGTCTTCAGCGGGAGGATGGTGGACGAGGTTCAGTTTTCCTGACTAAACCTTGTG TGCTTAACGCGAGCATGTAGCTTCTTCCGTGCCAGAAAGAGAAAGCAAGAGTCAGTTGCTCTCATTCTGGGGGATG CAGTGAGAAGGATTCACTTACTGTGTTTGAAAATCGCTCTGTGTTATATGCCTAAGAAAGTTGCCTGGGGATTTTCT ACCTGGGTAAAGGGATCCAGTTTTCTCCCTTTCAAGTAGGTGAAGCTGTTCAGCATTTACAGCTTTCCGGGAGGCA TTCTGTGCTGCGTGGCCAGGAATGCCGCCTTGCAGACATTATCCCTGAAGGAGAAAAAACTGCATTTCTCAGGT TCTGGTATATGACAGCGGCACAGCAGCAGTTTTGCTGAGTAAAGACTATGTTAGCAGCAAGCGTAAGTTCCCA >NW004443969.1 7

CTTCACCGCATGAAAGAGAAAGCAAACTACAGCATTCTCCTTAAGCTGCTTAACTTCATCCACTGAAGTTTTCAGTTT TGTTAGAAGAAACAATCCGCTCAGCACCAGCTGCTTACTCAAAGTTTCTCTCAGTGGTATAAGACTAGAGAGTGCCT GCCCTGCAATTATTTGAATCCCCTGGTTTTCAAAGGGAACTTTGCCTAACAGGGAATGAAGGTACCTTTGGCATTCA AAGCAAAACAGGCAGATTGTTTGCCTTCAGCTAGAGGATGGAAGAAAAGGTTCAGTTTTGCTGAGTAAAGCTTGTG TGCTTAACCTGAGCCTGTAGCCTCTTCCCTGCCAGAAAGAGGAAGCAAAGCCAGTTGCTCTCAGTCTGAGGGATGA AGTGAGAGGGATTCACTTACTGTGTTTGAGAATCGCTCTGTGTTACATGCCCAGGAATGCTGCTTGTGCATTCTT CCTGGAGAAAGATTTTGAGTTTTCCCCTTGGACTTCTTGAATCCCTGTGGCTTTTGCAGGCTTGTGCATTCTCTT CCTGGAGAAAGATTTTGAGTTTTCTCCCTTGGACTTCTTGAATCCCTGTGGCTTTTGCAGCTTTCTGGGAGGCGCTC TTCGCTGCATGGGCAAGAAAGCAGACTTTTCAGGCGGCGCATCTCTGAAGGAGAAAACAAAGTGCCCTTTCTCCCGTTC TGGTATACGACAGCAGTAGAGATATAGTATCTCTAAGCAAACTCTATGTTAGCAGGACTTGTATGTTCCCA >NW004443969.1 11

>NW004443969.1 12

>NW004443969.1 13

ATTCACTGCATGAAAAAGAAAGCAAAATACAGTGTTCTCATTCAAGCTGGCTAACTTAATCCACTTAAACTATCAATT TTGATAAAAGATAGACCTCAAACCATCTACAAGCTCTCTCAAACATTCCTCAATGGGATAAGACTGGTGAGTTGCTC CCTTTCAAATGCTTCAATCTCCCGCTTTTCAAAGGGAACTTTGCATTATATGAACTGAAGGTACTTTTCACATTCAAA GAAAACAAGGCAGATTGGTTCATCTTCAGATTAAAGACAACAGAAGCGGTTCAGTTTTGCTGATTAAACCGTATGTT GTAACCATGTGCATATTTCCTATGCACTGAAAGAAAGAGAAAGAGAAACCTAAAGAACCTGCTCTTCATCTGGTGGATGAAG TGAAAAGGATTCATTTACTGTGGTTGAAACTCCCTCTGTGTTACATGCCCAAGAATGCATGGTCTACCTGGGTAAGA GTTTGACTTTTCCCCCTCAGAGATTGTGAAGCTCTGCAGGGTTTACAGCTTTCCAGGAGTCGTCTGTGCTGCACA GCCGAGAATGCAGCCTTACAGACGGTCTCCCTAAAGGAAGAGAAAGTGTGTTTCTCCCGTTCTGGTATATGACA GTGGCAGTGCCGCAGTTTTGCAGAGGTAAAGACTATGTTAGCAGCAAGTGTAAGTTCCCA >NW004443980.1 4

ATTCACCACATGCAAGAGAAAGCAAAATACAGAGTTCTCCTTCAAGCTGCTTAACTCATTCTACTGAAGTTTTTAGTT CTGTTACCAGAAACAGTCCGCGCACCACAAGCTGCTTTCTCAAAGGTCTCTCCAGAGGGAGAAGACTGGAGAGTGC GGGCCTAGCAAGTATTTGAATCTCCTCCTTTTCAAAGGAAACTTTGCATAACAGGGAATGAAAGTACTACTGGCTTT CAAGGCAAAACAGGCAGATTGTTTGCCTTCAGCCGGAGGAGGGGCAGAAGAGGGTTCAGTCTCGCTGACTAAACCTT GTGTGCTTAACGCTAGCATGTAGCTTCCTTCTTGCCTGAGGGAGAAAGCAAGATCTTGTTGCTCTCAGGCTGGGGA TGAAGGGAGAACGATTCACTTACTGTGTTTGAAAATCCCTTGGTGTTACATGCCTAAGAATGCTGCCTGTGCATTTT CTACCTGGATAAAGTGTATGACTTTTCTCCCTTTCAAGTAGTTGAAGCCGTTCAGGGTTTCCAGGATTTGGTAGGCA TTCTTTGCTGCACGGCTGAGAATACCGCTTTACAGAAGGTATCACTAAGAAGAGAAAACTGCATTTCTCCCCATT CTGGAACACGAATGAGGCGAAACCGTAGTTGTGCTGAGTAAAACTTTGTTAGCAGCAAGTGTAAGTTGCCA >NW004444509.1 2

ATTCACTGCCTGAAAGAGAAAGCAAAATACGGCGTTCTCCTTCAGCTGCTTAACTGAATCTACTGAAGTTTTCAGTT TTGGTAGAAGAAACACTCTGCCCACCACCAGTTGCTTTCTCAAAGATCTCTCAGTGGGAGAAGACTGGAGAGAGTGCC TGCCTTGCTGGTATTTGAATCCCCTGGTTTTCAAAGGGAACTTTGCATTACAGGGAATGAAGGTACTTTTGGCATTC AAAGCTAAACAGGTACATTGTTTCCCTTCTGCTAGAGGACAGCAGAAGAAGTTCAGTTTTAATGAATAAAATATGTGT GGTTAACGGGAGCATGTAGTTTCTTCCATATCAGAAAACAGAAAGCTAGAGCCAGTTTCTCTCGGTGTGGGGATGAA GTGAAAAGGATTCACTTATTGAGTTTGAAAATCCCTCTGTGTTACTTGACTAAGAAAGTTGCCCGTGCATTGTCTAC CTGGAGAAAGGGTTCACTTATGAGCTTTCCCCTTCAAGCAGTTGTAGCCTTTCAGGGTTCATAGCCTTCCGGGGAGGACTT CTGTGCTGCGCGGCTGAGAATACCGCTTTACAGACGTTATCCCTAAAGGAGAAAACGGAGCAAGCTGAGGCCAGTTTCTCCCCTTTC AGGAACATGAATGAGGCAGAGCCGCAGTTTTGCTGGGTAAACACTATGTTAGCCGCAAGTGTTTAAGTTCCCA >NW004444509.1 3

ATTCACTGCATGAAAGAGAAAGCAAAATACAGCGTTCTCCTTCAAGCTGCTTAACTGAATCTACTGAAGTTTTCAGTT TTGTTGGAAGATAGAGCCCCCCCACCACAAGCTGCTTTCTCAAATGTTCTCTCAGTGGGAGAAGGCAGTAGAGTCA AGGCCTAGCAAGAATTTGAATCCCCTGGTTTTCAAAGGGAACTTTGCATAACAGGGAATGAAGGTACTCTTGGCATT CAAAGCAAAACAGGCACATTTTTTCCCTTCTGCTAGAGGATGGAAGAAGCTGTTTAGTTTTCCTGATGAATTTTCAGT GCTTAACACAAGCATGGAGTTTCTTCCATATCAGAAAGAGAAAGCGAGAGCCAGTTTCTCTTGGTGTGGGGATGAA GTGAAAAGGATTCACTTATTGAGTTTGAAAATCCTTCTGTGTTAAATGCCTAAGAAAGTTGCCTGTGCATTGTCTACG TGGACAAAGTGTATGACTTTTCTCCCTTTCAAGTAGTTGAAGCCATTCAGGGTTCATAGCCTTCCGGGAGGCCATTCT GTGCTGCGCAGCTGAGAATGCCCCTTTACAGACGTTATCCCTAAAGGAGAAAGCGAGAAACCGAGTTTCTCCCGGGAGGCATTCT GTGCTGCGCAGCTGAGAATGCCCCTTTACAGACGTTATCCCTAAAGGAGAAAACGGAGAAACCGAGTTTCTCCCATTCAG GAACATGAATGAGGCAGAGCCGCAGTTTTGCTGGGTAAACACTATGTTAGCAGCAGGTGTAAGTTCCCA >NW004444509.1 4

ATTCACTGCATGAAAGAAAGCAAAATACAGCGTTCTCCTTCAAGCTGCTTAACTGAATCTACTGACGTTTTCGGT TTTGGTAGAAGAAACACACCCGCCCACTACCAGCTGCTTTCTCAAAGGTTTTCTAAGTGGAAGTAGCTTCTAGAGTGC CTGCCTTGCCAGTATTTGAATCCCCTGGTTTTCAAAGGGAACGTAGCATAACAGGGAATGAAGGTACTTTTGGCATT CAAAGAAAAACAAGCAGATTGTTTCCCTTCTGCTAGGGGATGGCAGAAGAGGGTCATTTTTGATGAATAAAGTTTGA GTGCTTAACGCGAGCATGTAGCTTCTTCCATGTCAGAAAGAGAAAGCGAGAGCCAGCTTGTCTCGGTCTGGGTATG ATGTGAAAAGGATTAACTTACTGTGTTTGACAATCCCTCTGTGTTACTTGCCTAAGAAAGTTGCCCGTGCATTGTCTA CCTGGACAATGTGTATGACTTTCTCCCTTTCAAGTAGTTGTAGCCATTCAGGGTTCATAGCCTTCCGAGAGGCATT CTATGCTGCGTGGAGGAGAATACTGCTTTACAGAAGGTATCCCTAAAAAGAGAAAACCACGTTTCTCCCGTTC TGGAACACGAATGAGGCAGAGACCGCAGTTTTCCTGAGTAAAGGTAACCTATGTTAGCAGCAAGTGCCA >NW004444509.1_6

ATTCGCCGCATGAGAGAGAGAAAACGAAACACAGCGTTCTCCTTCAAGCTGCTTAACTGAATCTACTGAAGTTTACAGT TTTGTGAGAGGAAACAATCAGCCCACCACCAGCTGCTTTCTCCAAGGTTCTCTCAGTGGGAGAAGACTGGAGAGAGTG CCTGCCTTGCAAGTATTTCAATCCCCTGGTTTTCAAAGGGAACATTACATGACAGGAATGACGGTATTACTAGCTTA CAAAGCCAAAGAGGCAGATTGTTTGCCTTCAGGTAGAGGACAGCAGAAGAGGGTTTAGTTTTGCTGACTAAACCTTG TGTGCTTAACACGAGCATGTAGCTTCTTCCCTGCCTAAAAGAGAAAGCAAGAGGCTTGCTGCTGCTCTCAGTCTGGGGAT GAAGGGAGAACGATTCACTTCCTGTGTTTGAAAATCCCTCTGTGTTCCATGCCTAAGAAAGTTGCCTGTGCATTGTC TACCTGGATAAAGTGTATGACTTTTCTCCCTTCAAGTAGTTGAAGCCGTTCAGTGTTTCCAGCATTCCGGGAGACA TTCTGTGCTGCGCGGCTGAGAATACTGCTTTAAAGAACGTATCCCTAAAGGAGAAGAGAAACCGCGTTTCCCGGGAGACA TTCTGTGCTGCGCGGCTGAGAATACTGCTTTAAAGAACGTATCCCTAAGGAGAAAGCAAGAGAAAACCGCGTTTCCCCGT ACTAGAACACAGCACTGCAGAGCCGCCGTTTTGATGACTAAGAACTATGTTAACAGCAAGTGTAAGTTCCCA >NW004444309.1_4

ATTCGCCGCATGAGAGAGAGAGAGCGAAACACAGCTTTCTCGTTCAAGCTGCTTAACTGAATCTACTGAAGTTTTCAGT TTTGTGAGAGGAGACAATCTGCCCACCACCAGCTGCTTTCTCAAAGGTTCTCTCAGTGGTTTTCAAAGGGAACTTTG CATAACAGACAGTGAAGGTACTACTGGATTTCAAAGCAAAACAGGCAGATTGTTTGCCTTCAGCTAGAGGATGGCA GAAGAGGTTCAGTTTTGCTGACTACACCTTATGTGGTTAACGTGAGCATGTAGATTCTTCACGGGCAGAAAGAGAA AGTAAGAGCCTGTTGTTCTCAGTCTGGGGATGAAGTGAGAAGGATTCAGTCCCAGTGTTGGAAAATCCCTCTATGT TCCATGCCTAAGAAAGTTGCCTGTGCAGTGTCTACCTGGATAAAATGTATGACTTTTCTCCCTCTCAAGTAGTGGAA GCCGTGCAGGGTTTACAGCATTCCGGGAGGCCTTCTGTGCTGCGCGGCTGAGAATACCGCTTTACTGACGCTATC CCTAAAGGAGAAAACCGCGTTTCTCCCCTTTCTGGAACACGGATGAGGCAAAACTGTAGTTGTGCTGAGTAAA AGTTTGTTACCAGCAAGTGTAAGTTCCCAT

>NW004444309.1 5

ATTCACCACTTAAAAGAGAAAGGAAAAGGAAAATACAGCGTTCTCCTTTAAGCTGCTGAACTGAATCTACTGAAATTTTCAGTT TTGTTAGAAGAAACAGTCAGCCCAGCAGAAGCTGCTTTCTCAAAGGTTCTCTCAATGGGAGAAGCCTGGAGAGGGC CTGCCTTGCAAGTATTTCAATCCCTTGGTTTTCAAAGGGAAGTTCTCTCAATGGGAAAGGGTACTTCTGGCTTT CAAAGCAAAACAAGCAGATTGTTTGCCTTCAGCTAGAGGATGGCAGAAGAGGTTCAATATCGCTGAATAGAACTTG TGTGCTTAAGGCGAGCATGTAGCTTCTTCCCTGCCTGAAAGAGAAAGCAAGAGGTTCAATATCGCTGAATAGAACTTG GGAGGATTCACTTACTGTGTTTGAAAATCTCTCTGAGTTTACATGCCCAAGAAAGTTGATTGTCTACCTCG TTAAAATGTATGACTTTTCTTCCTTTCCAGGTAGAGCGTTCAGGGTTTAGATCATTCGCGGAGGCATTCTTTGC TGCGCGGCTGAGAATACCTCTTTACACACGGTATCCGTAAAAGAGAAGAAGATAACCGCGTTTCTCCCGTTCTGGAAC ATGAATGAGGCAAAACCGTAGTTGTGCTGAGTAAAAACTTTGTTAGTTGCAAGTGAGGTTCCCA >NW00444309.1_9

GTTCACTGCATGAAAGAGAAAGCCAACTACAGTCTTCTCATTCAAGCTGGTTAACTTAATCAACTTAAGTTTTCAAAT TTGGTAGAACAAAGAGTCCTCACACCACCAGCTGCTTTCTCAAACATTCTCTCAATGGGATAAGATTGACGAGTGTC TGCTTTACAAGTACTTGAATCCCCTCTGTTTCAAAGGGAACTTTGCGTTACAGGGAATGAAGGTACATTTTGCATTCT AAGAGAAACAGGCAGATTGTTTGCCTCCAGCTAGATGATGGCAGAAGCAGTTCCGTTTCCCTGAGTAAATCTTGTC TTGTCAGCACGTGAATGTTTCCTATTCACTGCAGAAATGAGAAAACAAAAGATAGTTGCCCTCAGTCCAGTGCATGA ACTGAGAAGGATTCACTTACTGTCTTTGGAAATCCCTCTGTGTTACTTGCCCAGGACTGTTCCGTGTGCACTCTTTA CCTGGGTAAAGACTTCCACTTTCTCCATTTCGACTACATGAAGCTCTGCGGGGTTTCCAGCTCTTGGGAGACACTC TATGCTGTGCAGCCAAGAATGCAGCCTTACAGCCGGTATCGCTGAGGAGAAAACAAAGCCACTCTCTGGGAGAATGCAGCCTTACAGCCGGTATCGCTGAGGAGAAACAAAGCAAAGCCATTTCTCCCATTC GGCATACGACAGCAGTAGAGATGTAGTTTTGTAAAGTAAACACTATGTTAACAGGATGTGCCATGTTCCCA >NW00444309.1_18

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 TGTTAGAAGAAAGTGTCTGCCCAGCACAAGCTGCTTTCTCTAATGTTCCCTGCGTGGGAGAAGGCTGGAGAGGTGCA GGCCTAGCAAGTATTTGAATCCCCTGCTTTTGAAAGGGAATTTTTCCTACCAGGAATGAAGGTACTAATGGCTTTCC AAGCCAAACAGGTAGATTGTCTGCCTTCAGCTAGAGGACAGCAGAAGAGGTTCAGTTTTGCTGACTAAAGCTTGTT TGCTTAACGGGAGCATGTAGCTTCTTCCCTGCCTGAAAGAGAAAGCAAGAGGTTCAGTTTGCTGCCTCCAGTCTGGGGATGA ACGGAGCACGATTCACTTTCTGTTTTTCCAAGTCCCTCTGTGTTACATGCCTAAGAAAGTTGCCTGTGCATTGTCTA CCTGGATAAAGCGTATGACTTTTCTCCCTTCCAGTACTACTTGAAGCCGTTCAGGCTTTCCAGCATTGCCTG CTGTGCTGCGTAGCTGAGAATAGGGCTTTACAGACGCTATCGCTAAAGAGAAAGCTGCCTCCCGGGAGGCATT CTGTGCTGCGTAGCTGAGAATAGGGCTTTACAGACGCTATCGCTAAAGGAGAAACCGCATCCTCCCATTC TGGAACGCGAAGGAGGCAAAACCGGAGTTCTCCTGAGTAAAAACCTTGTTAGCAGTAAGTTTAAGTTCCCT >BMBL01009774.1 2

ATTCACCATATGAAAGAGAAAGCAAAAGACAGGCTTCCCTTCAAGCTGCTTAACTGAATCTACTTAAGTTTTCAGTTT TATTAGAAGAAAGAGTCCGCCCAGCACAAGCTGCTTTCTCCAATGTTCTCTGAGTGGGAGAAGGCTGGAGAGGTGCA GGCCTAGCAAGTATTTGAATCCCCTGGTTTTCAAAGGGAATTTTGCATACCAGGGCATGAAGGTACTATTGGCTTTC AAAGCCAAACAGGTAGATTGTCTGCCTTCAGCTAGAGGGAAGGCAGTAGAGGTTCAGGTTTGCTGACTAAAGCTTGT GTGCTTAACAGGAGCATGTAGCTTCTTCCCTGCCTGAAAGAGAAAGCAAGAGCTTGATGTTCTCAGTCTCGGGACG AAGGGAGAATTATTCACTTTCTGTGTTTGAAAATCCCTGTGTGTTACATGACTAAGAAGCTGCCTGGGCATTTTCTA CCTGGATAAAGCGTATGACTTTTCTCAATTTCAAGCAGTTGAAGCCGTTCAGGTTTCCCAGCATTCCGGGAGTCATT CTGGTCTGCGCGGTTGAGAATACCGCTTCACAGGCGGTATCCGTAAAGGAGAAGAGAAACAGCGTTTCTCCCATTC TGGAACGCGAGGGAGTCAAAACTGTAGTTGTGCTGAGTAAAACCTTTGTTAGGAGCAAGTGGAAGTTCCCC >BMBL01009774.1 3

TCAGAGCCAAACAGGCAGATTGCCTGCCTTCAGCTAGAGGAAGGCAGAAGCGGTTCAGTTTTGCTGACTAAACCGT GTGTGCTTAACACGAGCATGTAGCTTCTTCCCTGCCTGAAAGAGGAAGCAAGAGCTTTTTGCTCTCAGTCTGGGGA GGAAGGGAGAACAATTCCCTTTCTGTGTTTGAAAATCCCTCTGTGTTACATGCCTAAGAAAGTTGCCTGTGCATTGT CTACCTGGATGAAGTGTATGACTTTTCTCCCTTTCAAGTAGTTGAAGCTGTTCAGGGTTTCCAGCATTCCGGGAGGC ATTCTGTGCTGCGCGGCTGAGAATACCGCTTTACAGACGGTATCTCTAAAGGAGAAGAGAAACCGCGTTTCTCCCA TTCTGGAACCCGAGGGAGGCAAAACCATAGCTGTGCTGAGTAAAAGCTTTGTTAGGAGCAAGTGGAAGTGCACC >BMBL01009774.1 4

ATTCACCACATGAAAGAGAAAGCAGAAGACAGCACTGTCCTTCAACCTGGTTAACTAAATCTACTGAAGTTTTCAGG TTTGTTAGAAGGAAGTGTCCGCCCAGAACAAGCTGCATTCTTCAAATGTTCTCTCAGTGGGAGAAGACTGGAGAGT GCGGGGCTAGCAATTATTTGAATCCCCTTTTTTTCCAAGGGAACTTGCATACCAGGGAATGAAGGTACTACTGGCT TTCAGATCCAAACAGCCAGATTGTTGCCTTCTGCTAGAGGGACAGCAGAAGAGGGTTCAGTTTTGCTGACTAAACCTTG TGTGGTTAATAGGATCCTGTATCGTCTTCCCTGCATGAAAGTGAAAGCAAGATCTTGTTGCTCTCAGTCTGGGTAGG GAGAACAGTTCACTTTCTGGGTTTGAAAAAACCCTCTGTGTTACATGCCTAAGAAACTTGCCTGTGCATTGTCTACCT GGATAAAGTGTATGACTTTTCTCCCTTTCAAGAGGTTGAAGCCATTCAGGGTTTCCAGCATTGCCTGGGAGGAGATCTT TGCTGCGCGGCTGAAAATACCGCCATACAGGCAGTATCGCTAAAGGAGAAGAGACCCCGTTTCTCCCATTCTGG AACGCGAGGGAGTCAAAACCGTTGTTGTGCTGAGTAATAACTTTGTTAGCAGGAAGTGGAAGTTCCCC >BMBL01009774.1 5

ATTCACCACATGGAAGAGAAAGAAAACTACAGCGTTCTCCTTCAAGCTGCTTAACTGAATCTACTGCAGTTTTCAGT TTCTTTGGAAGAAAGTGTCGGCCCAACACAAGCTGCTTTCTCCAATGTTCTCTCAGTGGGATTAGGCTGCAGAGTC CGGGCCTAGCAAGGATTTAAATCCCTTGGTTTTCAAAAGGAACTTTGCATACCAGGGAATGAGGGGGACTACTGGCT TTCAACGCCAAATTGGCAAATTGTCTGCCTTCAGTTAGAGGACGGCAGAAGAGGGTCAGGTTTGCTGACAAAACCG TGTCTGCTTAACAGGAGCATGTAGCTTCTTCCCTGCCTGAAAGAGAAAAGCGAGAGGTTGCTGCTGCTCTCAGTCTGTGG ACAAAGGGAGAATGATTCACTTTCTGTATTTGAAAATCCCTCGGTGTTACATGCCTAGGAAAGTTGCCTGTGCATGG TCTACCTGGATAAAGGGTATGACGTTTCTCCCTTTCAAGTAGTTGAAGCCGTTCAGTGTTCCTACCATTCCTTGAGG CTTTCTGGGCTGCGCGGCTGAGAATACCGCTTTACAGACGGTATCGCTAAAGGAGAAAGCAAGGTTCCCTCCAGTTTCTCCC ACCTGGAACTTGAAGGGGCAAAATCGTAGTTGTGCTGAGTAGAAACTTTGTTAGCAGCAAGAGAAACTCCCC >BMBL01002074.1 1

>BMBL01012666.1 1

ATTCACCGCGTGAAAGAGAAAGAAAAATACAGCATTCTCCTTCAAGCTGCTTAACAGAAGCTACTGAAGTTTTCAGG TATGTTAGAAGAAACAGTCCTCCCACAACCAGCTGCTTTCTCAAACTTTCTCTCAGTGGGCTAAGACTGTAGAGTGC TTGTCTTGCAAGTATTTGAATTCCCTGGTTTTCAAAGGGAAATTTGCATCCCAGGGAATGAAGGTGGTTTTGGCATT CAAAGCAAAGCAGGCAGATTGTTTGCCTTCAGCTAGGGGACGGCCGAATTGTTTCAGTTTGGCGGAATAAACCTTG TGTGCTTAACGCCAGAAAGTAGCTTCTTCCATGCCTGAAGGAGAAAGCAAGAGCAAGTGCCTGTGCAGTCTGGGGAA GAAGTAAGGAGGAGTCACTTACTGTGTTTGAAAATCCCTCCGTGTTACATGCCTAAGAAAGTTGCCTGTGCATTGTC TACCTGGATAAAGCGTATGACTTTTCTCCCTTTCAAGTAGTTGAAGCCGTTCAGTCTTTCCAGCATTCCTGGAGGCA TCCTGTGCTGCGAAGCTGAGAATAGGGCTTTACAGACGGTACCGCTAAAGGAGAAGGAAACCGCGTTTATCCCATT CTGGAACGCGAAGGAGGCAAAACCGTAGTTGTGCTGAGTAGAAATTTTGTTAGCAGCAATTCTAAGTTCCCC >BMBL01012666.1 2

ATTCCCAGTATGAAAGGGAAGGCAAAATACAGTGTTCACATTCAAGCTGGTTAAGTTAAGCAACTTAAGTTTCAGT ATTGGTAGAACAAACGGTCCTCACAGTACCAGCTCATTTCTCAAACATTCTCTCAGTGGAATAAAACTGGCAAGTGT CTGCCTTTCAAGTACTTGAATCCTCTGCTTTTCAAAGGAACTTTTCATTACAGGGACTAAAGGTAGCTTTTGCATTCT AAGCAAAACAGGCAGACTGTTTGCCTTCAGCTAGAGGATGTCAGCAGGGGTTCAGTTTTGCTGACTAATCCTTGTG TGCTTAACACAAGCATTTCTTATTCCTTGCAAGAAAGAGAAAGCAGAAGACAGATGCTCTCAATCTGGGGGATAAAG TGAGAAGAAGTTACTCACTGTGTTTGAAAATCCCTCTCTGTTACTGGCCCAGGACAGCTGCCTGTGCATTGTATACC TGGATAAGAGTTTGAATTTTCTCCCATTGGAGTTCGTGAAGCCCTGCGGGGTTTACAGCTGCTGTGCATTGTATACC TGGCTGCGAGGCCGAGAATTCCGCCTTACAGAGGGGTATCCCTAAAGGAGAAGAAGACGCCTTTTCAGGAGGCACTCT GTGCTGCGAGGCCGAGAATTCCGCCTTACAGAGGGGTATCCCTAAAGGAGAAGAGAGGCGCTTTTCTCCCGTCCTGG TATACGACAGTGGTAGCGCTGTAGTTTTGCTGAGTAAAGATTATGTTAGCAGCAAGTGCAAGTTCCCA >BMBL01099827.1 3

ATTCAATGCACGAAAGAGAAAGCAAACTACAGTGTTCTCCTTCAAGCTGGTTAACTTAACCCACTGAAGCTTTCAGT TTTGGTAGAATACACAGTCATAACACAACCTCCTGCTTTCACAAAATTTCTCTCAGTGGGATAAGACTGTCGAGCGT CTTCCTTGCAAGTATTTGAATCCACTGCGTTTCAAAAGGGAACATGCAGTACAGGGGATGGAGGTACCTTTGGCATT CTAAGCAAAACAGGTGGATTGTTTTCCTTCAGCTAGAAGACAGCAGCAGAAGCAGTTCAGTTTGCTGAGAAAACCTTGT GTGCTTAACACGAGCAGGTATCCTATTCCCTGCAAGATAGAGAAAGCAAAAGACAGTTGGTCTCAGTCTGGGTGAT GAAGTGAGGAGGACTCAATTACTGTGTTTGAATCCCTTGGTGTGAAATGCCTAAGAAAGTTGCCTGAGCATTATCTA CCTGGGTCAAGAGGTTTGACTTTTCCCCGGTTGTGGGGGTTTACAGCTTTCCAGGAGGCCTTTTGTGCTGCGCAGAC GAGAATCTACCCTTACAGACGGTATCACTAAAGGAGAAAGAGAGAAATTGGGTTTCTCCCATTCTGGTATATGAAAGCG GCAGCATTGTAGTTTTGCTGAGTGAAGACTATGTTAGCAGCAAGTGTAAGTTCTCA >BMBL01099827.1 4

ATTCATTGCCTGAAAGAGAAAGCAAAATACAGTGTTCTCATTCAAGCTGGTTAACTTAATCAACTGACATTTTCAGTT TTGGTAGAAGAAACTGTCCTTACAGCACCAGTGTGCTTTCTCAAACATTCTCTCAATGGGATAAGAGTGGCAAGTGT CTGCCTTTCAATTACTTTAATTCCCTGGTTTTCAAAGGGAGCTTTGCATTACAACGAATGAAGGTAACTTTCGCATTC TAGGAGAACTGATTGTTTAACCTTCAGATATTTGACAGCAGAGGGCGGTTCTGTTTTCTGACTAAACCTTGTGTTGTA ACCACATGCATGTTTCCTATTCACTGCAACAAAGAGAAACTAAACAACAGTTGCTGTCAGTCTGGTTGATGAACTGA GAAAGATTCACTTAATATGTTCGATAGTCCCTTTGTCTTACATGTCCAAGAATGTTGCCTGTGAATTCTCTTCCAGGG TAAACAAAGCATTTCACTTTCACTGCCTTAGAGTACACGAAGCCTTACGGAGTTTGCAGCTTTCCCCGGGGAACTCT GTGCTGGGCGCTCAAGAATGTGGGCCTTAGAGACCTTATTCCTTAAGGAGAAAACACAAAGCCCATTTCCCCGTC TGATACATGATCGCAATAGAGACATAGTTTTGCTAAGGAAACACAGTAGCAGAGCCTGCATGTTTCCT SBMBL 01099827 1 8

TCACTGCATAAAAAAGAAAAATCAGAATTCAGTGTGCTCATTCAAGCTGGTTAACTTAATCAACTTATTAGTTTTGAG AAACACAGTTCTCCCACCACCTGCAAGATTTCTCAAACATTCTCTTGATGGGATATGACTGGCGAGTCTCTCGTTTT CCAGTACTTGATTCCCCTGCTTTTCAGAGGGAACTTTGCATTACAGGGAATGAAGTTACCTTTTGCATTCTCAGAAG AACAAGCAGATTGTTTCACCTTCAGGTAAGGGATAGCAGAGGCGGTTCAGTTTTACTGACTAAACCTTGTGTTTTAT CTACATGCCTGTTTCTTATTAACTGCAACAAAGAGAAACCAAACGACAGTTGCTCTCATTCTGGTGGATGAAGTGAG AAGGATTTACTTAATGTGCTTGAAAATCCCTCTGTGTTACATGCCTAAGAATGTCGCCTATGCATTGTCTGCCTGGG TAAAGCATTTCACTTTTCTACCCTTGAAGTACATGAAGCCCTGCGGGGGTTTACAGCTTTCCGGAAGGCACTCTGTTG CGTGGCTGAAAATGTGGTCTTATAGACAGTATGTCTGAATGAGAAAACAAAATGTGTTTCTCCCTGTTCTGGTATACG ACTTCAGCAGAGATGTAGTTTTGCTAAGGAAACACCACGTTAGCAGGACCTGCGTGTTCCAA >BMBL01099827.1 9

>BMBL01099827.1 10

>BMBL01099827.1 15

>BMBL01099827.1_16

>BMBL01099827.1 17

ATTCACTGTATGAAAGAGAAACCAAAATACAGTTTCTCTCATCTGCTTGATTAAGTCACATGGTATTATCTGTATTGA TAAAACTCTGTTCCTACATGATGGAGGATCACTCTTATACATTTTCTGAATTTGAAAATAATGTCAAGTTTCTGACTTT CTAGTATGTGAGTGTTTTGGTTTTCAGGAATGAATGTGCCTTTCATAGTCTATGAAGAACAAGGGAGATTGTTTCCC GTTCAACTACAGGACAGCAAAAACTATGTAGTTTTACTAATTAAACCTTGTGTTGTAAGCACGTGCATGTTTCTAATT CATCACATGAAAGATAAACCAAAACACAGTTTCTCTCATTCTGGTTGATTATGTCAGAGCGCTGACTTACTATGTTTG AAAACACTCTGTGTTTCATGCATGAGAATGGTTTGTATGCATTCTCCACCTGGGTAAATAATTTAAATTTGTCCCTT TCAAATAGATGAAACCATGCAGTGTTTAGATGTTTTCAAGAAGTACTCTGTGCTGCCTGACCATGAATGGGCCTTAC AGATAGAATCTCTGAACCAGAAAACAATGCAAGTTTCTCCTGTTCAAGTATAGGACAACAATGTAGTTTTGCTAAGTA AATACTCTGTTACCAGAATGTGCAAGTTCCCA

>BMBL01105564.1 1

>BMBL01043776.1 1

GTGCATGTTCTCA

>BMBL01043776.1_3

ATTCAGTCTATTAĀAAATGAAGGCAAAATGCTGTGTTCTCATTCAAGCCTGTTTCTTTAATGAACTGAAATTTTCACTT TTGTTAGAACAAGCAGTCCTGTCACCATCAGCTTCTTTGTGAAACATTCTCTGAATGGGTGTAACTGGCCTTTCAAG TACTTGGATCCTGTGGTTTTCGAATGGAAAGTTTAATTATAGAGATTGAAGGAACACTTTGCATTGAAAGGAGAACA GTCAGATTGTTTGCCTTCACCGAGAGGACAGCACCAAGCAGCTCAGCTTGTCTGACTAAAGCTTGTGCGGTTAATAT CAGCATGTATCCTATTCTGTTCAAGAAAGAGAAACCAAAAGACAGTTGCTCTCAGGCTGTGGGAATGAAGTGAGAA GGATTCACTTACTGTGTTAGAAATTTCCTCTGTGTTACATGCCCAAGAACTTTCCCTGTGCGGTTATTTCGCCTGGGGATA AGTGTTTGACTTTCTCCCTTTCGAATACGTTTATCCCTGCGGGGGTTTACCGCTTTCCCGTGTGGCTCTCTTTGCCCTGGGTAA AGTGTTTGACTTTTCTCCCTTTCGAATACGTTTATCCCTGCGGGGATTACCGCTTTCCCATGTGGCCCCATTCCCATTCCCATTCCCATTCCCATTCCCATTCCCATATACGG CCTGCGGATAATGCCGCCTTGCATTCCGTATCGCTGAAGGGAAGAAAAGCGGATTTCCCCATTCCCATATACGG CAGCAGCAGTGCTGTACTTTTGCTAACTAAACACTATGTTAGCAACATTTGCATGCTCCCA >BMBL01061436.1 2

TTCTCTGCATGAAGAAGAAGCAAAATACCGTGTTCTCATTCAAGCCTGTTTACTTAATCAACTGAAGATTTTCAGTT TCGGTAGAGCCAACGGTCGTGACATAACCAGACACTTTCTGAAATATTTTCTCAATGGGATAAGACTGGTGATTGTC TGCTTTTCCAGTACTTGAATCCCCTGGTTTTCAAATGGAACTTTGCCTCACAGGCAAAGAAGGTAACTCTCATTCTAA GGAAAATAGGCAGATTGTTTGCCTTCAGCTAGAGGACAGTAGAAGCAATTCAGTTTGCTTACTAAAGCTTGTGTGG TTAACTTGAGCATGTTTCCTATTCCCTGCAAGAAAAAGAAACCAAAAGACAGTTGCTCTCAGTCTGAGCAATGAAGT GACAAGGTTTCACTTACTTTGTTTGAAAAATCCCTCTGTTTTACATGCCCAAGAAAGTTGCCCCTGCGTTGTTTACTG GGGTAAAGAGTTTGACTTTTCTCCCATCCAACCACGTGAAGAACCTGTGGGGGTTTACTGCTTTACAGGAGGCACTCT GTGCTGCTAGGGAGAAAATGCCACTTTACAGACCACGTGAAGACCTGTGGGGAGAAGAGAAAACGCATTTCTCCCTTTCAG GTATATGACAGCAACAGCTGTGTTGTTTGCTAACTAAAGACTATGTTAGTAGCACATGTATCTTCCCA >BMBL01061436.1 4

>BMBL01061436.1 5

ATTCACAACATGAAAGAGAAAGCAAAATGCAGTTTCTCTCATTCAGCTTGAGTATCTCAGAAGGATTAACTCACTGT GTTTGAAAAGCCCTCTGTGTTACATGCATGAGAATGTTGCCTATACATTCTCTACCAGAGAACTTTTCTCCCATTCAA

>BMBL01082959.1 2

>BMBL01041892.1_1

>BMBL01041892.1 3

>BMBL01022471.1_1

ATTCAGTGCATGAAAAAGAAAGCAATATGCAGTGTTCTCTTTCAAGCTGCTTCACTTAAACAACTTAAGTTTTCAGTG TTGGTAGAACAAAGCGTCCTCATACCACCGGCTGCTTTCTCCACCTTTCTTGCAATGGAAGAATACTGGCGAATGTC TGCCTTTCAAGAGATTGCATCCCCTGGTTTTCAAAGGGAACTTTGCAGTACATGGAATGAAGGTAAGATTTGCATTC TGTGAAAAACATGCAGATTTTTTGCCTTCAGCTAGAGAAGGGCGGAAGTGGATCAGTTTTGTTGACTATATCCTGTG TGGTTAACACGAGCAGGTATCTTATTCCTTCCAAGACAGAGAGAAACCAAAAGACAGTTGCTCTCATTCTGGGGGGATG AAGTGAGATGGTTTCACTTACTGTGCTTGAATATCCCTCTATGTTACATGCCCAGGAAATTTTCCTGTCGATTGTCTC CCTGAGTAAAGTGTTTGAATTTTCTCCCCCTTCGAGTACAAGAGGGCCCTGGGGGGTTCACAGCTTTCCCGAACCACTA GTTGCTGCGCGGCCTAGAACACCGCCTTATAGACTGTATGCCTGAAGGAGAAGAGAAAGCGTGTTTTTCCTGTCGTTCT GGTAAGACAGTGGCAATGCTGTAGTTTTGCTAACTAAAGACTATATTAGCAGCACTTGCAAGTTCCTA >BMBL01022471.1 2

Supplementary Sequence File 3. TMAsat monomeric sequences (in fasta) manually isolated from the *Hydrodamalis gigas* assembled reference genome (accession GCA 013391785.1).

>JACANZ010423367.1

CCTTTCAAGTACTTCAATCCCCTGGTTTTCAAAGGAAATTTTGCATTACAGGGAACGAAGGAACTTTCGCATTCTAA GGGAAACAGGCAGATTGCCTTTAGCTAGAGGACGGCAGAAACAGTTCAGTTTTGGATACGAAATCTTGTGTGGTTA ACATGAGTAAGTATCCTATTTCCTGCAACAAAGAGAAACCAAAAAGTTGCTCTCTGTTTCAGGAATGAAGTGAGAAG GATTCATTTACTGTGTTTGAAAATCCCTTTATGTTACTTGCCCAAGAAAGTGGCTGGTGCATTGTCTACCTGGGTAAA AAGTTTGACCTTTCTCCCATTCCGGTACTTCAAGCCCTGCGTGGTTTACAGCTTTCCGGGAGGCATTCTGTGCTGC ACGGCTGAGAATGCGGCCTTACGGACGGTATCCCTGAAGGAGAAGACAAAGCGTGTTTCCCCCATTCTGGTATAT GACAGTGGCAGGGCTGTAGTTTTGCTAAGTAAACACTGTTAGCAGCAAGTACATGTTCCCA >JACANZ010389660 1

ATTCCCAGTATGAAAGGGAAGGCAAAATACAGTGTTCACATTCAAGCTGGTTAAGTTAAGCAACTTAAGTTTCAGN NTGGTAGAACAAAAGGTCCTCACAGTACCAGCTCATTTCTCAAACATTCTCTAAGTGGAATAAAACTGGCAAGTGTC TGCCTTTCAAGTACTTGAATCCTCTGCTTTTCAAAGGAACTTTGCATTACAGGGACTAAAGGTAGCTTTTGCATTCTA AGCAAAACAGGCAGACTGTTTGCCTTCAGCTGGAGGATGTCAGCAGGGGGTTCAGTTCTGCTGGCGAATACCTTGTG GTTTAACACAAGCATTTCTTATTCCTTGCAAGAAAGAGAAAGCAGAAGCAGAAGACAGATGCTCTCAATCTGGGGGGATAAAGT GAGAAGAAGTTACTCACTGTGTTTGAAAATCCCTCTCTGTTACTGGCCCAGGACAGCTGCCTGTGCATTGTATACCT GGATAAGAGTTTGAATTTTCTCCCCATTGGAGGTTCCGTGAAGCCCTGCGGGGTTTACAGCTTTCAGGAGGCCACTCTG TGCTGTGAGGCCAAGAATTCTGCCTTACAGAGGGGTATCCCTAAAGGAGAAGAAGACGAGAGCACATTTCTCCCCATCCTGGTA TACAACAGTGGTAGCGCTGTAGTTTTGCTGAGTAAAGATTATGTTAGCAGCAAATGCAAGTTCCCAC >JACANZ010229584.1

>JACANZ010159291.1_1

>JACANZ010159291.1 2

>JACANZ010122094.1_1

>JACANZ010122094.1_2

>JACANZ010120854.1 1

ATTCACCACATGAAAGAGAAAACACATTACAACGTTCGCCTTCCAGCTGCTTAACTGAATCTACTGAAGTTTCCAGT TTTGTTAGAAGAAAGTGTCCGCCCAGCACAAGCTACTTTCTCCAATGTTATCAGAGTGGGAGAAGGCTGGAGAGGTG CGGGCCAAGCACGTATTTGAATCCCCTGATTTTCAAAGGGATCTTTGCGTACCAGGGAATGAAGGGATTACTGGCT TTCAAAGCCACACCTGCCCATTGTCTGCCTTCAGCTAGAGGATGGTAGAAGAGGGTCAGGTTTCCTAAGTAAACCT TGTGTGCTTAACACGATCCTGTAGCTTCTTCCCTGCCTGAAAGACAAAGCAAAGCAAGAGCCTGTTGCTCTCAGTCCAGGG TTGAAGGGTGAACGACTCACTTTCTGTGTTTCAAAATCCCTCTGTGTTACATGCCTAAGGAAGTTGCCTGTGCATTG TCTACCCGGATACAGTGTATGACGTTTCTCCCTTTCCAGCAGTTGAAGCCATTCAGAGGTTTAGAGCATTCTGGAAGG GATTCTGTGCTGCGTGGCTGAGAATACCGCTTTACAGACTATTCTGAAAGGAGAAAACCGCGTACCTCCGT TCTGGAACACGAACGAGGCAAAACTGTAGTGGTGGAGAGACAACACTTTGTTAGCAGCAAGTGTAAGTTCCCA >JACANZ010120854.1 2

ATTCCCCACGTGAAAGAGAAAGCAAAATGCAGCGTTCTCCTTCAAGTTGCTCAACTGAAGCTACTGAAATTTCCCAT TTGGTTAGAAGAAAATGTCCGCCCACCAGAAGCTGGTTTCTCAAACGTTCTCTCAGTGGGCAAAGGCTGGAGAGTG CCTGCCTTGGAAGTATTTGAATCCTCTGGTTTTCAACGGAAATTTGGAATAAGAGGGGAATAGAGGGTACTACTGGCTT TCAAAGCAAAACGCAGAGATTGTTTGACTTCACCTAGAAGATGAAAGAAGAAGAGGGTTCAGTTTTACTGACTAAAACTTG AGTGCTTACGCGAGTGTGTAGACTCTTCCCTGCCTGAAAGAAGAAGAAGAAGAGGCTTGTTGCTGTCAGTCTGGGGATG AAGGGAGAACGATTCACTTTCTGTGTTTGAAAATCCCTCTGTGTTCCATGCCTAAGAAAGTTGCCTGTGCAGTCTGGGGATG ACCTGGAAAAAGTGTATGACTTTTCTCCCTTTCAAGTAGTTGAAGCCGCTCAGGGTTTCCAACATTTGGTAGGCATT GTTTGCTGCACGGCTGAGAATACAGCTTTACAGACGGTAGCGCTAAAGGAGAAGAGAAACTGCCTTTCTCCCATTC TGGAACACGAATGATACAAAACCGTAGTGGTGCTGATTAAAAACTTTGTTAGCAGCAATTCTTAGTTCCCA >JACANZ010078570.1 1

ATTCACCACATGAAAGAGAAAGCAAAATACACCGTTCTCCTTCAAGCTGCTTAACAGAATCTACGGAAGTTTTCAGT TTTGTTAGAGGAAAGAGTCCGCCCAGGACAAGCTGCTTTCTCAAACGTTCCTTGAGCGGGAGAAGGCTGGAGTGT GGAGGCCTAGCAAAGATTTGAATGCCCTGGTTTTCAAAGGGAACTTTGCATTCCAGGGAATGAAGGTACTACTGGA TTTCAAAGCAAAACAGTCAGATTGCGTTCAGCAAGAGGGACGGCAGACGAAGTTCAGTTTTGCTGACTAAACCTTGT GTGCTTAACACACACACATGTAGCTTTTTCCCTGCCTGAAAGAGGAAAGCAAGAGTTTGTTGCTCTCAGATTGGGGATGA AGGGAGAACGATTCCCGTTCTGTGTTTGAAAATCCCTCTGTGTTACATGCCTAGGAAAGTTGCCTGTGCATTGTCTC CCTGGATAAAGTGTATGACGTTTCTCCCTTTCAAGTAATTGAAGCCTTTTAGGGGTTTCCAGCATTCCGGGAGGCATT CTGTGCTGCTCGGCTGAGCATACCGCTTTACACACGGTATCGCTAAAGGAGAAAAGAAACCGCGTTTCTCCCATTC TGGAACGTGAAGTAGGCAAAACTGTTGTTCTGCTTAGTAAACGTTTGTTAGCAGCAAGGGCAAGATCCCC >JACANZ010078570.1 2

>JACANZ010255388.1

>JACANZ010129270.1 2

>JACANZ010175637.1

>JACANZ010035624.1 1

ATTCAGTGCAGAAAAGAGAAAACAAAATACGGTGTTCCTTTGAAGAAGGTTAACTTAAGTTGATTAGGTCATTAGTAT TGATAAAACACAGTAGTCACACCATCCACAAGCTTTGTCAAACATTCTCTCAGTGGGATAAGAATGGAAAGTCTCTC CCTTTCAAGTACTTGAATTCCCCGGATGTTCAGAGGTTTTCTAGGAGAACTTTGCATTACCCGGTATGAAAGTACCT TTTTCTCTCCAGGAAGAACAAGGCAGATTGCTTAGCATTCATGTGAAGGATGGAAGAAGTGGTGCAGTTTTGCTGA CTAAGCCTTGTGTTGTAAGCACGTGCATGTTTCCAGTTCACTGCAAGAAACAGAAACCAAAAAGGCAGTTGCTCTCA GTCTGGTGGATGAAATCAGAAGGATTCACTACTGTTTGGAAATCCCTGTGTTACATGCCCAAGAAAGTAGCCATG CTTTCTCCCATGTGGGTAAAGAGTTTGACTTTCCCCTTTGAGAAATCCCTGTGTTACATGCCCAAGAAAGTAGCCATG CTTTCTCCCATGTGGGTAAAGAGTTTGACTTTCCCCCTTTGAGCAAGTGAAGCCCTGCGAGGTTTAGAGGTTTCCG AGAGCCACTGTGTGCTGCCTGGCTGAGAATGTGGCCTTCTGGAAGGTATCTCTGAAGGAGAAACAATGCACGTTTC TCCCATTCAGGTATACGACAGCAGTAGAGATGTGCTTCTGCTAAGTAAACACTGTATTAGCAGAACGTGCATGTTCC CAG

>JACANZ010035624.1 2

>JACANZ010035624.1 3

>JACANZ010035624.1 4

>JACANZ010107986.1 1

>JACANZ010107986.1_2

AATGAGTCACTTAACCGGTTTGAAAATCTGTCTGTGTTACATGCCCAAGAATGTTGCCTATGAATTGTCCATCTGGA TAAAAGCTTGGATATTTCTCCCTTTTAAGTACTTGATTTAAGTACTGCGGGGATTTACAGCTTTCCAGGAGGGCATTCTA TGCTGCCTGGCGAGAATGCAGCCTTACAGACAGTATATCTGAGGGGGAAAAAGAAACCACGTTTTCTTCCATTCTGG TATACCACAGCGGTACAGATGGAGTGTTGCTAAGTAAACCCTATGTTAGGCAGACGTGCATGTTCCCT >JACANZ010061387.1 1

>JACANZ010190329.1 2

>JACANZ010402190.1

>JACANZ010202422.1

AATTAGTGCCTTAAAGAGAAACCAGAATACAGTTNNNNNNNNNTTGGGTAACTTAATCTACTTGAGTAAGTTATCA GTTTGCATAAAACACACTGCCCACAGGATCCACAACGTTCCTCCTACATTCTCTGTAAGGAAGAAGAATGTCAAGTG TCTGCCTTTAAAGTACTTGAATTCCCCGGGCTTTCAGAGGTGTTCTGGGGAAAATGTGTGCTACCAGCAGTGAATG

>JACANZ010316682.1

>JACANZ010334892.1

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 wholegenome 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., Noväk, 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: Specieslevel 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.