UNIVERSIDADE FEDERAL DE MINAS GERAIS
INSTITUTO DE CIÊNCIAS BIOLÓGICAS
DEPARTAMENTO DE BIOLOGIA GERAL
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA

DISSERTAÇÃO DE MESTRADO

Citogenômica de DNAs satélites em macacos-de-cheiro (*Saimiri,* Cebidae, Platyrrhini)

ORIENTADA: Mirela Pelizaro Valeri

ORIENTADORA: Marta Svartman

COORIENTADOR: Gustavo Campos e Silva Kuhn

BELO HORIZONTE

Agosto – 2017

Mirela Pelizaro Valeri

Citogenômica de DNAs satélites em macacos-de-cheiro (*Saimiri,* Cebidae, Platyrrhini)

Dissertação apresentada ao programa de Pós Graduação em Genética da Universidade Federal de Minas Gerais como pré-requisito obrigatório para obtenção do título de Mestre em Genética, área de concentração "Genética Evolutiva e de Populações".

Orientadora: Dra. Marta Svartman

Coorientador: Dr. Gustavo Campos e Silva Kuhn

BELO HORIZONTE

Agosto – 2017

043 Valeri, Mirela Pelizaro.
Citogenômica de DNAs satélites em macacos-de-cheiro (Saimiri, Cebidae, Platyrrhini) [manuscrito] / Mirela Pelizaro Valeri. – 2017.

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

Orientadora: Dra. Marta Svartman. Coorientador: Dr. Gustavo Campos e Silva Kuhn.
Dissertação (mestrado) – Universidade Federal de Minas Gerais, Instituto de Ciências Biológicas.

1. Genética evolutiva. 2. Citogenética. 3. DNA satélite. 4. Sequências de Repetição em Tandem. 5. Bandeamento Cromossômico. I. Svartman, Marta. II. Kuhn, Gustavo Campos e Silva. III. Universidade Federal de Minas Gerais. Instituto de Ciências Biológicas. IV. Título.

CDU: 575

Ficha elaborada pela Biblioteca do Instituto de Ciências Biológicas da UFMG

Pós-Graduação em Genética
Departamento de Biologia Geral, ICB
Universidade Federal de Minas Gerais
Av. Antônio Carlos, 6627 - C.P. 486 - Pampulha - 31270-901 - Belo Horizonte - MG
e-mail: pg-gen@icb.ufmg.br FAX: (+31) - 3409-2570

"Citogenômica de DNAs satélites em macacos-de-cheiro (Saimiri, Cebidae, Platyrrhini) "

Mirela Pelizaro Valeri

Dissertação aprovada pela banca examinadora constituída pelos Professores:

Marta Svartman
UFMG

Francisco Pereira Lobo
UFMG

Fernando Araujo Perini
UFMG

Gustavo Campos e Silva Kuhn
UFMG

Belo Horizonte, 22 de agosto de 2017.

Dedico

Aos meus pais **Sérgio** e **Marli**, razão da minha existência.
Os maiores incentivadores, oferecendo apoio incondicional a cada nova etapa.

AGRADECIMENTOS

Faltam-me palavras para agradecer à todos que (direta ou indiretamente) contribuíram para o desenvolvimento deste trabalho:

À minha orientadora, Dra. Marta Svartman, pela oportunidade, orientação, ensinamentos científicos, paciência, convivência e amizade. Aprendi muito nessa pequena caminhada!

Ao meu coorientador, Dr. Gustavo campos e Silva Kuhn, por toda ajuda, atenção, ensinamentos, paciência e convivência.

Ao Guilherme Borges Dias, por pacientemente ter me ensinado tantas coisas, por ter feito muitos desafios parecerem mais fáceis e principalmente pela amizade! Foram tantos momentos preciosos, você é uma pessoa maravilhosa.

Ao Dr. Roscoe Stanyon, pelos ensinamentos, por gentilmente ter cedido amostras, me ajudado a realizar experimentos e analisar os dados, fora a ótima convivência.

Àos queridos amigos Alice, Radarane e Matías. Além de toda ajuda e convivência no laboratório, não sei o que teria sido de mim sem vocês.

Aos atuais e ex- integrantes do Laboratório de Citogenômica Evolutiva, por todo apoio, amizade, boa convivência e pelo acolhimento de braços abertos quando entrei no grupo. Em especial à Naiara, José Ricardo, Bráulio, Pedro, Carla e Luísa.

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

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 e Camila Nascimento Moreira por terem sido tão solícitas e responsáveis por ceder parte das amostras que usei para desenvolver essa dissertação.

À Fundação Zoo-Botânica de Belo Horizonte, em especial à Valéria do Socorro Pereira, pela ótima colaboração e por nos enviar amostras que foram indispensáveis para o desenvolvimento desse trabalho.

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

Aos amigos que fiz na UFMG: Verônica, André, Fábio, Marliete, Heliana e tantos outros... Vocês deixam a vida mais leve!

Aos meus maravilhosos pais Sérgio Valiengo Valeri e Marli da Guia Pelizaro Valeri e toda a minha família, pelo amor sempre presente e pelo suporte emocional.

Ao Artur, por todo o incentivo e por fazer parte da minha vida.

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

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

SUMÁRIO

LISTA DE FIGURAS

Capítulo 1

Capítulo 2

LISTA DE TABELAS

Capítulo 1

Capítulo 2

LISTA DE ABREVIATURAS

2n - número diplóide, *diploid number*
A - acrocêntrico
AGU - *Alouatta guariba*
AIN - *Aotus infulatus*
BHY - *Brachyteles hypoxanthus*
BLAST - Basic Local Alignment Search Tool
bp - B*ase pairs*
CSA - *Chiropotes satanas*
cyt-b - Citocromo b
DAPI - 4′,6-Diamidine-2′-phenylindole dihydrochloride
ENC - Evolutionary new centromeres
FISH - Hibridação *in situ* Fluorescente, *Fluorescent in situ Hibridization*
FN - Fundamental number
IUCN - International Union for Conservation of Nature
L1 - long interspersed nucleotide element–1
LLA - *Lagothrix lagotricha*
Ma - Milhões de anos
Mya - Million years ago
NA - Número de acrocêntricos
NCBI - National Center for Biotechnology Information
NF - Número fundamental
NOS1AP - *nitric oxid synthase 1 adaptor protein*
NWM - New World monkeys
pb - Pares de base
PIR - *Pithecia irrorata*
satDNAs - DNAs satélite, *satellite DNAs*
SBO - *Saimiri boliviensis*
SINEs - Short interspersed nuclear elements
SM - submetacêntrico
SSC - *Saimiri sciureus*
SUS - *Saimiri ustus*
SVA - *Saimiri vanzolinii*
SXA - *Sapajus xanthosternos*
TEs - Transposable element
α - Alfa

RESUMO

Os macacos-de-cheiro pertencem ao gênero *Saimiri* (Cebidae; Platyrrhini) e são distribuidos ao longo da Bacia Amazônica e em uma pequena parte da América Central. O número de espécies e as relações filogenéticas do gênero *Saimiri* são controversos. A citogenética foi fundamental para auxiliar na identificação das espécies de *Saimiri*: todos os indivíduos do gênero analisados até o momento apresentaram número diploide 2n = 44 com variação de 74 a 78 no número fundamental, de acordo com sua distribuição geográfica. Neste trabalho, investigamos os DNAs satélites de *Saimiri boliviensis* usando o software RepeatExplorer e caracterizamos os dois mais abundantes, o alpha e o CapA. Mapeamos esses DNAs satélites nos cromossomos de *S. boliviensis*, *S. vanzolinii*, *S. sciureus* e *S. ustus*. O DNA satélite alfa possui cerca de 340 pb, corresponde a aproximadamente 1% do genoma de *S. boliviensis* e tem localização centromérica nas quatro espécies. O CapA compreende cerca de 2,2% do genoma de *S. boliviensis* e tem aproximadamente 1.500 pb. Sua distribuição nos cromossomos está relacionada com a heterocromatina constitutiva localizada principalmente nas regiões subteloméricas dos cromossomos submetacêntricos, mas o CapA também está presente em algumas regiões intersticiais. Houve diferença na localização do CapA entre as espécies de *Saimiri*, sugerindo que este DNA satélite possa servir como marcador taxonômico. Também investigamos a presença de CapA em outros primatas, já que foi primeiramente descrito em *Sapajus apella,* e verificamos que está presente em alguns gêneros das três famílias de Platyrrhini. Uma busca mais detalhada revelou que a sequência do CapA está presente em cópia única num trecho do intron do gene *nitric oxid synthase 1 adaptor protein* (NOS1AP) em humanos, bem como na maioria dos eutérios. Provavelmente o DNA satélite CapA observado em Cebidae se originou a partir desta sequência intrônica e os possiveis mecanismos responsáveis por essa amplificação são discutidos. Até onde sabemos, esse é o primeiro relato de uma sequência intrônica de cópia única como origem de um DNA satélite. Informações sobre os DNAs satélites alpha e CapA podem contribuir para o conhecimento sobre a evolução cromossômica de *Saimiri* e de Platyrrhini.

Palavras chave: DNA repetitivo, FISH, bandeamento GTG, repetições em tandem, origem de DNA satélite

ABSTRACT

Squirrel monkeys of the genus *Saimiri* (Cebidae; Platyrrhini) are distributed in the Amazon basin and part of Central America. The number of species and their phylogenetic relationships are controversial. Cytogenetics was fundamental in the identification of *Saimiri* species: all analyzed individuals until now presented a diploid number 2n = 44 with fundamental numbers ranging from 74 to 78, according to their geographical distribution. In this work, we characterized the two most abundant satellite DNAs of *Saimiri boliviensis*, Alpha and CapA, using RepeatExplorer and mapped them in the chromosomes of *S. boliviensis*, *S. vanzolinii*, *S. sciureus* and *S. ustus*. The alpha satellite has ~340 bp, comprises ~1% of the *S. boliviensis* genome and had a centromeric location in the four species. CapA comprises about 2.2% of the *S. boliviensis* genome and has ~1,500 bp. Its distribution on the chromosomes is related to the constitutive heterochromatin, which is mainly located in the subtelomeric regions of submetacentric chromosomes, as well as in some interstitial regions. CapA has different locations among *Saimiri* species, suggesting that it could be used as a taxonomic marker. We investigated the CapA presence in other primates, since it was previously described in *Sapajus apella*, and confirmed its presence in some genera of the three Platyrrhini families. A detailed search revealed the CapA sequence is present as a single copy located within an intron of the *nitric oxid synthase 1 adaptor protein* (NOS1AP) gene in humans and in the genomes of most eutherian mammals. The CapA satellite DNA of Platyrrhini has probably originated from this intronic sequence and the possible mechanisms responsible for its amplification are discussed. To our knowlodge, this is the fisrt report of a single copy intronic sequence giving origin to a satDNA. Information about the alpha and CapA satellite DNAs may contribute to the understanding of chromosome evolution in *Saimiri* and Platyrrhini.

Key-words: repetitive DNA, FISH, G-banding, tandem repeats, satellite DNA origin

1. INTRODUÇÃO

1.1. Primatas Neotropicais

Os primatas da infraordem Simiiformes são divididos nas parvordens Catarrhini e Platyrrhini. A parvordem Catarrhini é composta pelos primatas do Velho Mundo, grandes macacos e humanos, enquanto Platyrrhini agrupa os primatas do Novo Mundo. Estudos moleculares indicam que o início da divergência entre Platyrrhini e Catarrhini aconteceu há aproximadamente 43,5 milhões de anos (Ma) durante o Eoceno. Os Platyrrhini formam um clado monofilético, são endêmicos do continente americano, encontrados no sul do México, passando pela América Central até o norte da Argentina na América do Sul (Perelman *et al.*, 2011; Martin, 2012), e a divergência entre suas famílias ocorreu há apenas cerca de 24,8 Ma (Perelman *et al.*, 2011; Schneider e Sampaio, 2015).

Platyrrhini compreende aproximadamente 140 espécies descritas, pertencentes a 18 gêneros, agrupados em três famílias: Cebidae, Pitheciidae e Atelidae (Figura 1). A família Cebidae inclui os gêneros *Saimiri, Cebus, Sapajus, Aotus, Saguinus, Callithrix, Cebuella, Mico, Callimico* e *Leontopithecus*. Pitheciidae é composta pelos gêneros *Pithecia, Callicebus, Chiropotes* e *Cacajao*, enquanto Atelidae compreende os gêneros *Alouatta, Ateles, Lagothrix* e *Brachyteles* (Perelman *et al.*, 2011; Rylands *et al.*, 2013; Schneider e Sampaio, 2015).

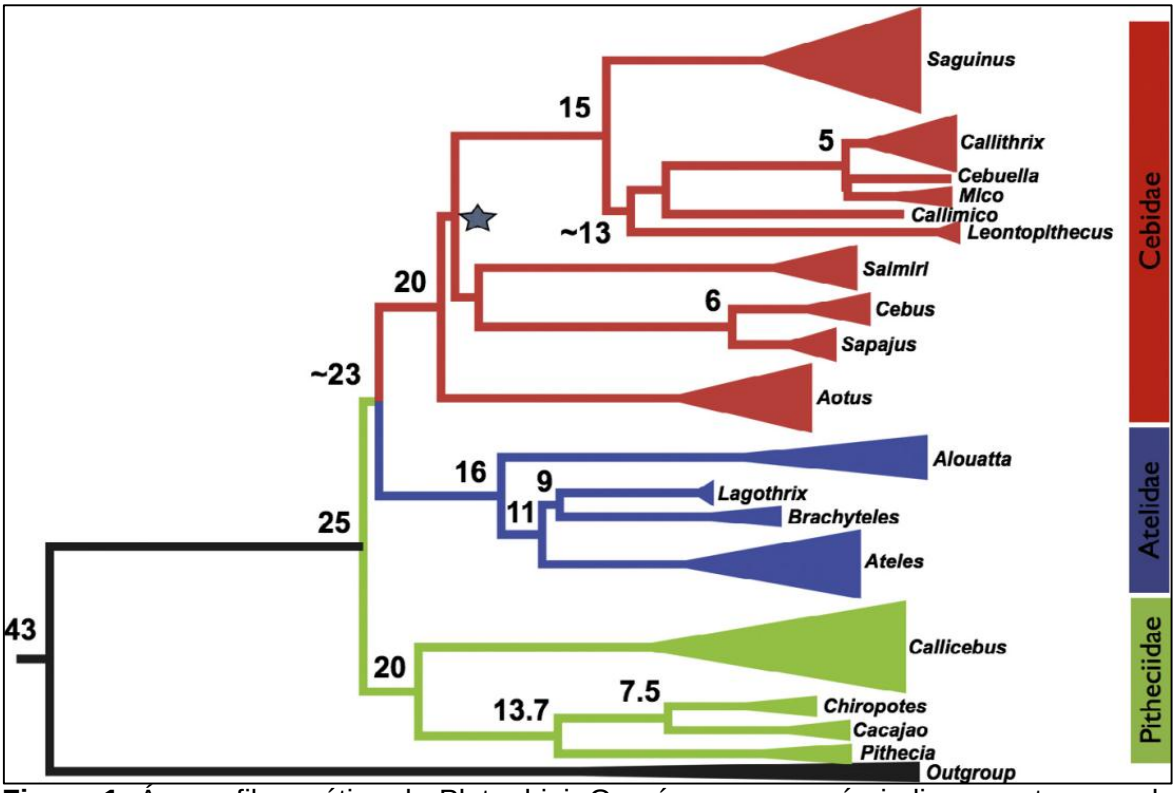


Figura 1. Árvore filogenética de Platyrrhini. Os números nos nós indicam os tempos de divergência em Ma. Fonte: Schneider e Sampaio (2015).

O processo de intensa radiação adaptativa sofrido pelos Platyrrhini resultou num grupo altamente diversificado em morfologia, comportamento e cariótipos, que representa um desafio ao estabelecimento de relações filogenéticas e à identificação precisa de espécies, o que se traduz em mudanças constantes na taxonomia do grupo (Seuánez *et al.*, 2005). A quantidade de espécies e gêneros de Platyrrhini reconhecidos ainda não é exata, com espécies novas ainda sendo descritas e alguns grupos passando por constantes revisões taxonômicas.

1.2. Gênero *Saimiri*

Dentro da família Cebidae, o gênero *Saimiri* é um bom exemplo de um grupo com taxonomia e filogenia confusas. As espécies do gênero são conhecidas como macacos-de-cheiro e são encontradas ao longo da Bacia Amazônica no Brasil, Colômbia, Venezuela, Guiana, Suriname e Guiana Francesa, incluindo algumas áreas de transição para Cerrado, e em uma pequena parte da América Central, no Panamá e Costa Rica (Figura 2) (Hershkovitz, 1984; Lavergne *et al.*, 2010).

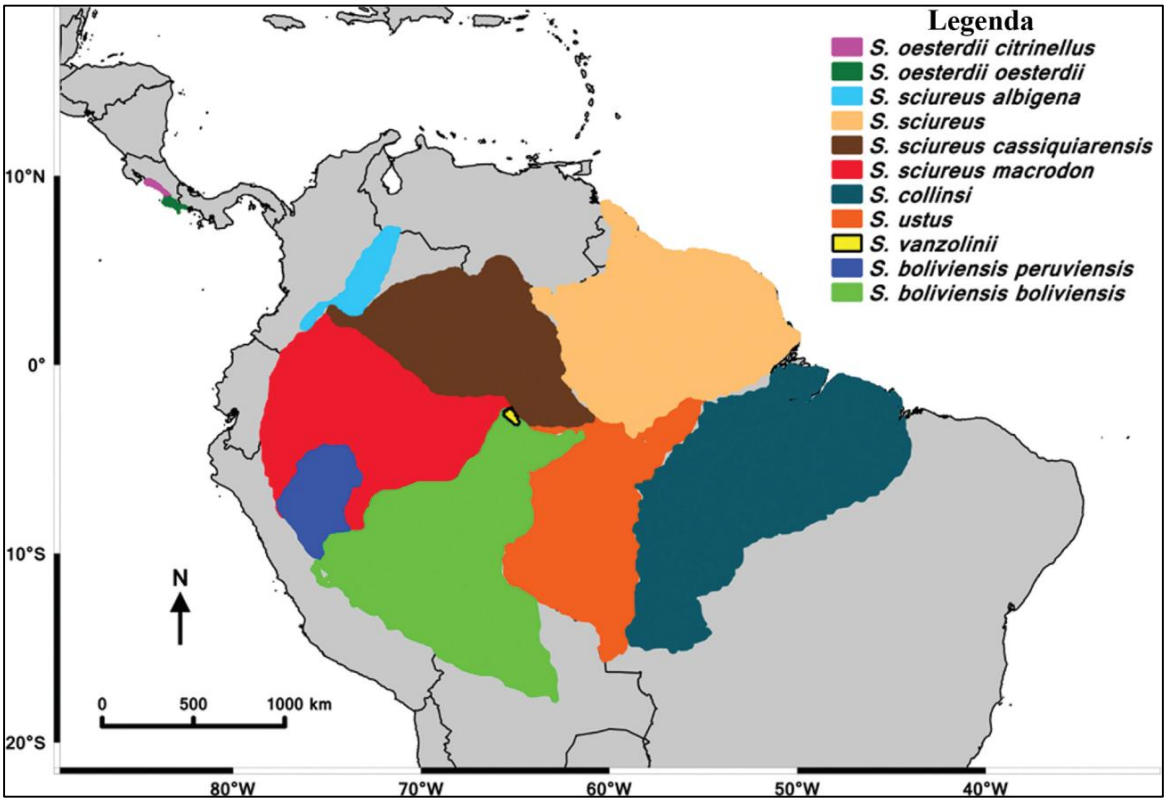


Figura 2. Mapa de distribuição das espécies do gênero *Saimiri*. Imagem adaptada de Carneiro *et al*. (2016).

Estudos realizados desde 1984 e baseados principalmente em análises morfológicas e moleculares ainda não elucidaram a controvérsia existente sobre o

número de taxa (variando de um a 16) e as relações filogenéticas entre as espécies do gênero (Hershkovitz, 1984; Thorington Jr., 1985; Costello *et al.*, 1993; Boinski e Cropp, 1999; Cropp e Boinski, 2000; Groves *et al.*, 2005; Lavergne *et al.*, 2010; Rylands *et al.*, 2013; Alfaro *et al.*, 2015; Merces *et al.*, 2015). A expansão e diversificação das espécies do gênero foram muito rápidas, com os todos eventos de especiação estimados aproximadamente entre 1,5 e 0,5 milhões de anos atrás (Alfaro *et al.*, 2015). Os principais fatores responsáveis pelas dificuldades nos estudos taxonômicos de *Saimiri* são as diferenças morfológicas sutis entre os diferentes taxa (Figura 3), o fato de amostras de localidades diferentes terem sido submetidas a análises diferentes (Hershkovitz, 1984; Alfaro *et al.*, 2015) e ocorrência natural de híbridos (Thorington Jr., 1985; Silva *et al.*, 1992; Costello *et al.*, 1993; Carneiro *et al.*, 2016).

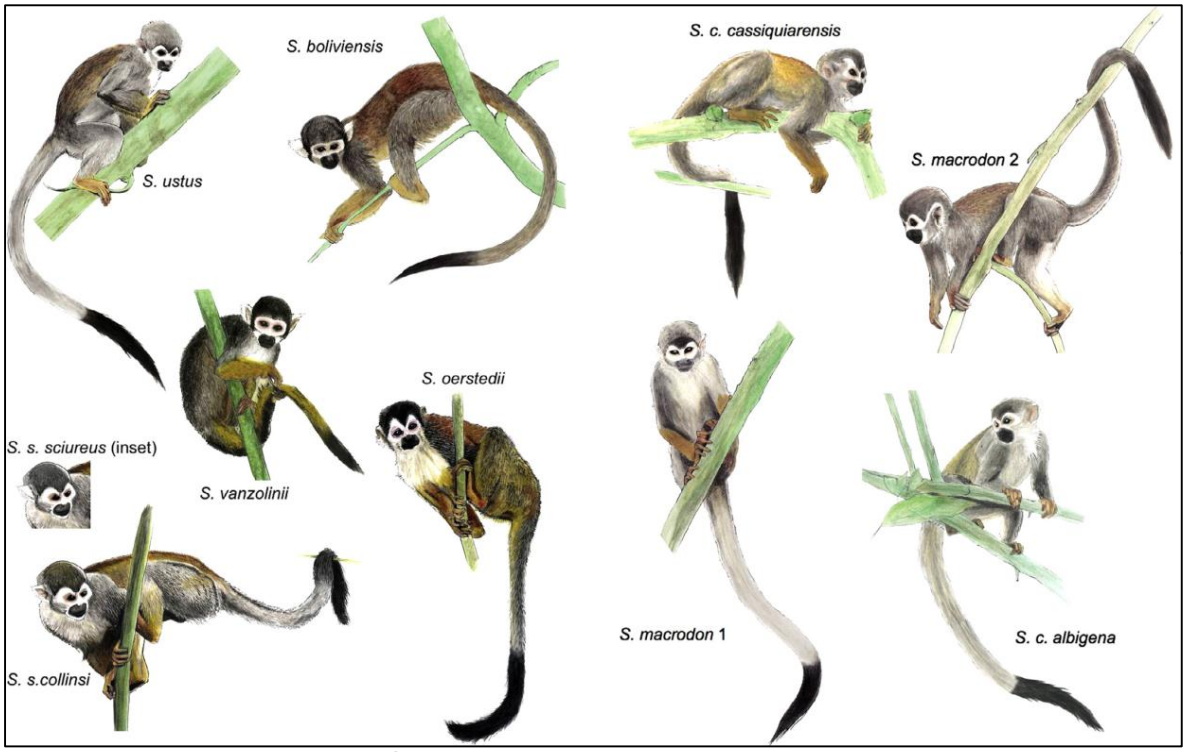


Figura 3. Diversidade morfológica encontrada em *Saimiri*. Fonte: Alfaro *et al.* (2015).

Há registros de híbridos nas bordas de contato entre as áreas de ocorrência de algumas espécies de macaco-de-cheiro. Híbridos entre *S. ustus* e *S. sciureus* foram relatados no banco leste do Rio Tapajós (Thorington Jr., 1985) e na região entre os rios Madeira e Tapajós (Costello *et al.*, 1993). Também foram encontrados híbridos entre *S. boliviensis peruviensis* e *S. sciureus macrodon* em uma região do Rio Ucayali, na Amazônia peruana (Silva *et al.*, 1992). Carneiro *et al*. (2016) descreveu a ocorrência de híbridos entre *S. boliviensis peruviensis* e *S. sciureus macrodon*, entre *S. b. boliviensis* e *S. ustus* e alguns híbridos entre *S. b. boliviensis* e *S. b. peruviensis*.

Com base principalmente em análises morfológicas e algumas informações de citogenética básica, Hershkovitz (1984; 1987) reconheceu quatro espécies de *Saimiri* divididas em dois grupos (Figura 4a). O primeiro, tipo romano, reúne subespécies de *S. boliviensis* (*S. b. boliviensis*, *S. b. peruviensis* e *S. b. vanzolinii*). O segundo, tipo gótico, compreende *S. sciureus* (*S. s. sciureus*, *S. s. macrodon*, *S. s cassiquiarensis* e *S. s albigena*), *S. oerstedi* (*S. o. citrinellus e S. oerstedi oerstedi*) e *S. ustus*. *S. boliviensis* foi considerado grupo-irmão de todas as outras espécies (Hershkovitz, 1984; 1987). Thorington Jr. (1985) propôs a divisão do gênero em apenas duas espécies: *S. ustus* e *S. sciureus*, que incluiria as outras espécies reconhecidas anteriormente (*S. s. sciureus*, *S. s. boliviensis*, *S. s. cassiquiarensis*, *S. s. oerstedi*) (Figura 4b). Costello *et al*. (1993) sugeriram que os macacos-de-cheiro com distribuição geográfica na Bacia Amazônica fossem considerados apenas uma espécie, *S. sciureus*, diferenciada de *S. oerstedi* da América Central (Figura 4c).

Groves (2005) e Rylands e Mittermeier (2009) reconheceram cinco espécies de *Saimiri* com várias subespécies: *S. boliviensis boliviensis*, *S. boliviensis peruviensis*, *S. sciureus sciureus*, *S. sciureus macrodon*, *S. sciureus cassiquiarensis*, *S. sciureus albigena*, *S. oerstedi citrinellus*, *S. oerstedi oerstedi*, *S. ustus* e *S. vanzolinii*. Essa é a classificação que consta na lista vermelha da International Union for Conservation of Nature (IUCN, 2017), com informações de que as populações estão diminuindo, exceto as de *S. vanzolinii*, para a qual não há dados suficientes. As categorias de risco para as espécies são: de menor preocupação para *S. boliviensis* e *S. sciureus,* vulnerável B1 para *S. oerstedi*, quase ameaçada para *S. ustus*, e vulnerável D2 para *S. vanzolinii*.

Baseados em análises de sequências do citocromo *b* (cyt-b), Lavergne *et al*. (2010) identificaram nove taxa terminais: *S. sciureus (S. s. sciureus, S. s. collinsi, S. s. albigena, S. s. macrodon* e *S. s. cassiquiarensis*), *S. ustus, S. oerstedii*, e *S. boliviensis (S. b. boliviensis* e *S. b. peruviensis*). Os autores também sugeriram que o complexo *S. sciureus* seria parafilético. A classificação mais usada atualmente segue Rylands et al. (2013), que reconheceram sete espécies e um total de 11 taxa: *S. oerstedii* (*S. o. oerstedii* e *S. o. citrinellus*), *S. cassiquiarensis* (*S. c. cassiquiarensis* e *S. c. albigena*), *S. macrodon*, *S. ustus*, *S. sciureus* (*S. s. sciureus* e *S. s. collinsi*), *S. boliviensis* (*S. b. boliviensis* e *S. b. peruviensis*) e *S. vanzolinii*.

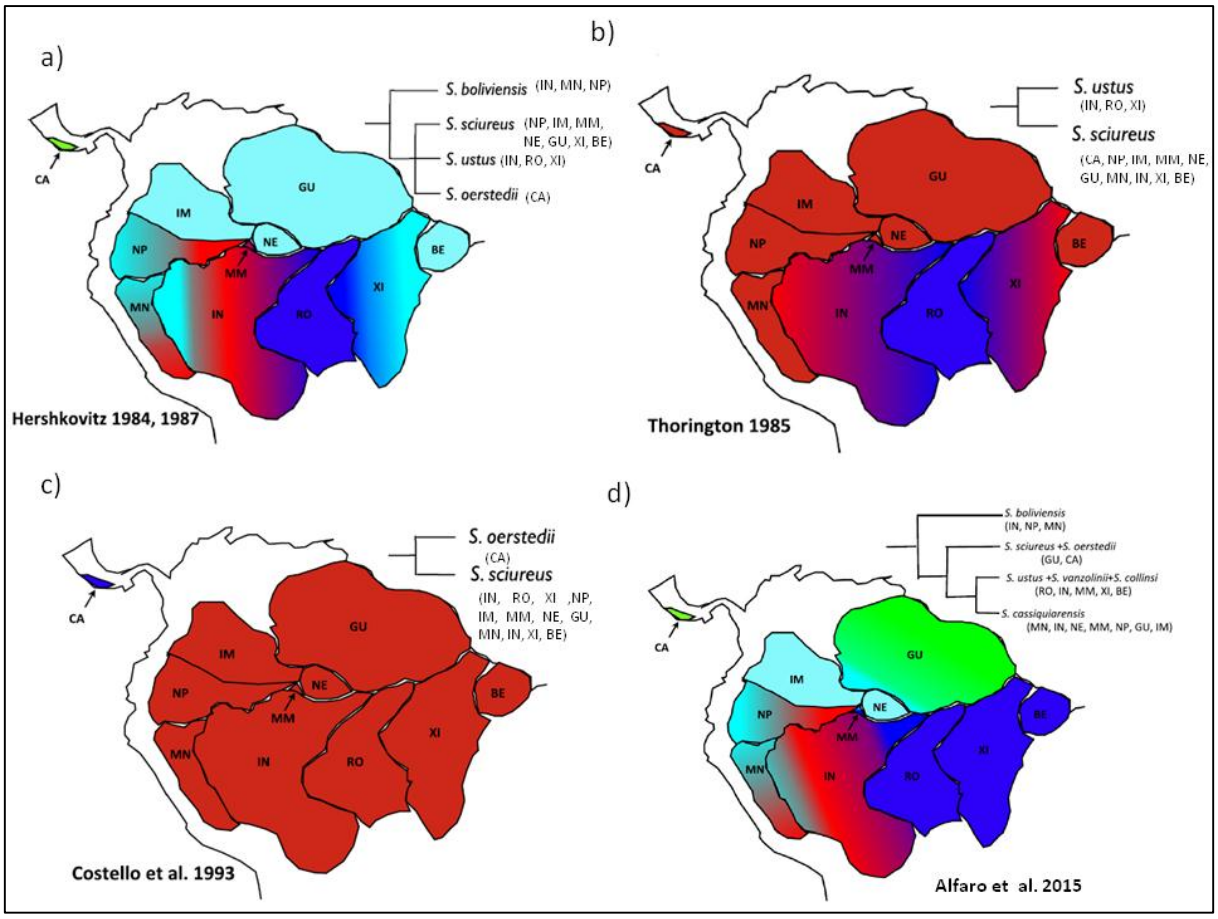


Figura 4. Classificações alternativas de *Saimiri* segundo: a) Hershkovitz (1984; 1987), que dividiu as espécies nos grupos romano e gótico; b) Thorington Jr (1985), que reconheceu *S. ustus* e *S. sciureus*; c) Costello *et al*. (1993), que sugeriram a diferenciação entre indivíduos da Bacia Amazônica e da América Central; d) Alfaro *et al*. (2015), que identificaram 14 taxa agrupados em quatro clados. As cores indicam graus de relações entre as espécies. Imagem adaptada de Alfaro *et al*. (2015).

Mercês *et al*. (2015) realizaram um estudo que incluiu análises morfológicas, morfométricas e de genética molecular e os levou a discordarem da conclusão de Lavergne *et al*. (2010), sugerindo a divisão de *S. sciureus* em duas espécies: *S. sciureus* ao norte do rio Amazonas e *S. collinsi* ao sul do rio Amazonas. Alfaro *et al*. (2015) realizaram análises utilizando as regiões mitocondriais D-Loop e cyt *b* e não encontraram suporte para a divisão de *S. boliviensis* em duas subespécies. Também consideraram *S. vanzolinii* como grupo-irmão de *S. ustus*, discordando da tradicional divisão morfológica gótico vs. romano (Hershkovitz, 1984, 1987). Estes autores também encontraram parafiletismo em três espécies, *S. sciureus*, *S. ustus* e *S. macrodon*, e propuseram uma classificação provisória com base em 14 clados: *S. boliviensis*, *S. cassiquiarensis* (*S. c. cassiquiarensis*, *S. c. albigena*, *S. c. macrodon* A, *S. c. macrodon* B, e *S. c. macrodon* C), *S. collinsi*, *S. oerstedii* (*S. o. oerstedii* e *S. o. citronellus*), *S. sciureus*, *S. ustus* (linhagens

A, B, e C), e *S. vanzolinii*. Esses 14 clados foram agrupados em quatro grupos: 1) *S. boliviensis*; 2) *S. sciureus* + *S. oestedii*; 3) *S. ustus* + *S. vanzolini* + *S. collinsi*; e 4) *S. cassiquiarensis* (Figura 4d). Alfaro *et al*. (2015) indicaram que o gênero precisa passar por uma nova revisão taxonômica e frisaram a necessidade de análises com marcadores nucleares.

1.2.1. Citogenética de *Saimiri*

A citogenética teve um papel fundamental na taxonomia de *Saimiri* por auxiliar na identificação de espécies. Todos os indivíduos do gênero analisados até o momento apresentaram número diploide 2n = 44 (Figura 5a) e a variabilidade cariotípica encontrada relaciona-se com o número de braços cromossômicos ou número fundamental (NF) (Bender e Mettler, 1958; Jones *et al.*, 1973; Jones e Ma, 1975; Ma e Jones, 1975; Hershkovitz, 1984; Yonenaga-Yassuda e Chu, 1985; Assis e Barros, 1987; Moore *et al.*, 1990; Stanyon *et al.*, 2000; Chiatante *et al.*, 2017). O NF varia de 74 a 78 e é descrito de acordo com o número de pares acrocêntricos: cariótipos com sete pares acrocêntricos possuem NF = 74, os com seis pares acrocêntricos têm NF = 76, e complementos com cinco pares acrocêntricos apresentam NF = 78. Os cromossomos envolvidos na variação do NF são os pares 5 e 15, que podem ser submetacêntricos ou acrocêntricos (Figura 5b) (Ma e Jones, 1975; Hershkovitz, 1984; Moore *et al.*, 1990; Chiatante *et al.*, 2016). Os cariótipos registrados até o momento são de três tipos: pares 5 e 15 acrocêntricos (FN = 74), par 5 submetacêntrico e par 15 acrocêntrico (FN = 76), ou os dois pares submetacêntricos (NF = 78).

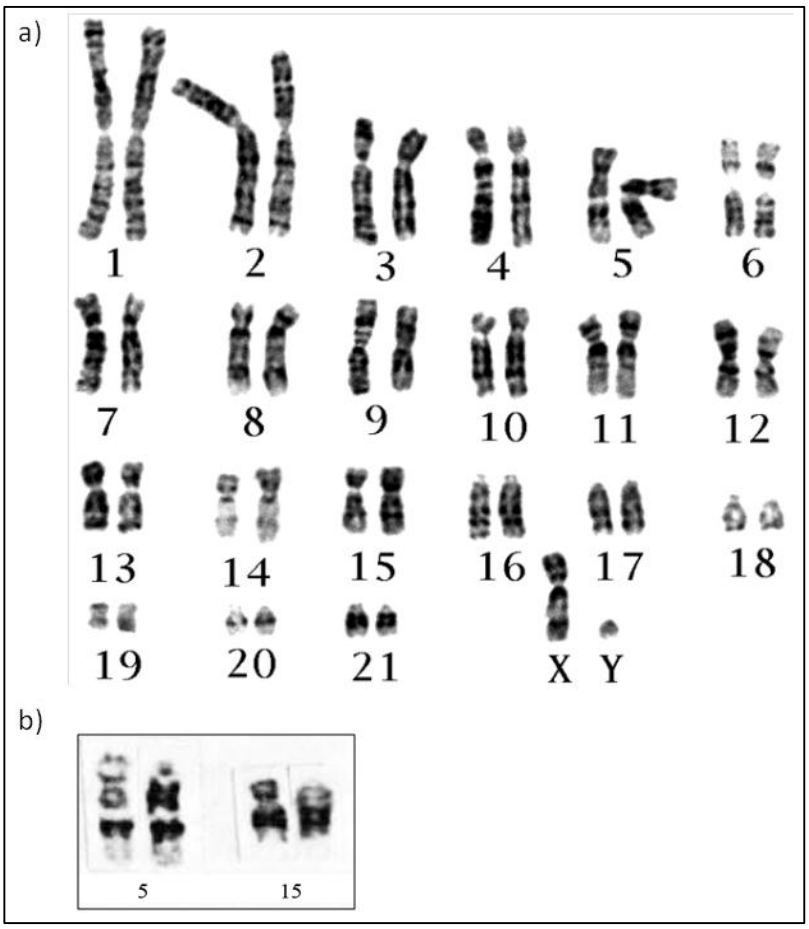


Figura 5. a) Cariótipo de *Saimiri sciureus* (2n = 44, NF = 78) após bandeamento GTG, os pares 5 e 15 desse indivíduo são submetacêntricos; b) pares 5 e 15 heteromórficos de um indivíduo híbrido de *Saimiri*, homólogos acrocêntricos à esquerda e submetacêntricos, à direita. Fontes: a) Stanyon *et al*. 2000; b) Lau e Arrighi, 1976.

Os dados sobre os NFs e as morfologias dos cromossomos 5 e 15 dos indivíduos de *Saimiri* descritos na literatura estão mostrados na Tabela 1. A variação dos NFs parece estar correlacionada com a origem geográfica do indivíduo, mas infelizmente não há dados de citogenética para todas os taxa reconhecidos atualmente e muitos dos indivíduos analisados não possuem origem geográfica conhecida. Segundo Hershkovitz (1984), os indivíduos com sete pares acrocêntricos (NF = 74) originários da Guiana e Nordeste da Bacia Amazônica no Brasil deveriam ser classificados como *S. sciureus sciureus*. Já os indivíduos com seis pares de acrocêntricos com origem no sudeste da Colômbia, Equador, nordeste do Peru e noroeste do Brasil deveria ser classificado com *S. sciureus macrodon*. *S. boliviensis boliviensis* teria seis pares de cromossomos acrocêntricos e é encontrado no sudoeste do Brasil, Bolívia e sudeste do Peru, enquanto *S. b. peruviensis*, com distribuição na região inferior do rio Ucayali-Huallaga no Peru, teria cinco pares de acrocêntricos. Por fim, os animais da América Central, *S. oerstedii*, teriam cinco pares de acrocêntricos. Indivíduos híbridos com os pais de diferentes regiões geográficas e nascidos em cativeiro apresentaram cariótipos com pares heteromórficos.

Tabela 1. Origem geográfica, número de cromossomos acrocêntricos e morfologia dos cromossomos 5 e 15 dos espécimes de *Saimiri* já analisados.

Identificação	Origem	NF	NA	Par 5	Par 15	Referência
S. boliviensis	Desconhecida	76	12	SM	A	Chiatante *et al*. 2017
S. boliviensis	Desconhecida	78	10	SM	SM	Chiatante *et al*. 2017
S. boliviensis boliviensis	Bolívia	76	12	SM	A	Moore *et al*. 1990
S. boliviensis peruviensis	Iquitos, Peru	78	10	SM	SM	Jones *et al*. 1973
S. boliviensis peruviensis	Pucallpa, Peru	78	10	SM	SM	Jones *et al*. 1973
S. boliviensis peruviensis	Peru	78	10	SM	SM	Moore *et al*. 1990
S. oerstedii	Costa Rica e Panamá	78	10	SM	SM	Jones *et al*. 1973
S. sciureus	Desconhecida	76	12	SM	A	Bender e Mettler 1958
S. sciureus	Desconhecida	74	14	A	A	Chiatante *et al*. 2017
S. sciureus	Desconhecida	78	10	SM	SM	Stanyon *et al*. 2000; Chiatante *et al*. 2016
S. sciureus macrodon	Leticia, Colômbia	76	12	SM	A	Jones *et al*. 1973
S. sciureus sciureus	Georgetown, Guiana	74	14	A	A	Jones *et al*. 1973; Ma *et al*. 1974
S. sciureus sciureus	Guiana	76	14	A	A	Moore *et al*. 1990
S. ustus	Rio Jamari -RO	78	10	SM	SM	Assis e Barros 1987
S. vanzolinii	Lago Mamirauá, Tefé - AM	76	12	SM	A	Yonenaga-Yassuda e Chu 1985
S. b. peruviensis x *S. s. macrodon*	Letícia x Iquitos	77	13	SM	SM/A	Jones *et al*. 1973; Ma *et al*. 1974
S. s. sciureus x *S. b. peruviensis* ?	Guiana x Peru?	76	12	SM/A	SM/A	Lau e Arrighi 1976

NF = número fundamental; NA = número de acrocêntricos, SM = submetacêntrico; A = acrocêntrico.

Híbridos nascidos do cruzamento entre pais originários de Letícia, Colômbia e de Iquitos, Peru, tinham o par 15 heteromórfico (Jones e Ma, 1975). Outro híbrido analisado por Lau e Arrighi (1976) apresentou ambos os pares 5 e 15 heteromórficos (Figura 5b). Os autores não sabiam a origem exata dos pais, mas sugeriram que um seria da região de Georgetown, na Guiana, e o outro do nordeste do Peru.

A principal hipótese para as diferenças de NFs nas espécies de *Saimiri* era a ocorrência de inversões pericêntricas (Ma e Jones, 1975; Lau e Arrighi, 1976). Porém, é possível notar que não há alteração na ordem das bandas GTG entre as formas acrocêntrica e submetacêntrica dos pares 5 e 15. Além disto, Chiatante *et al.* (2017) demonstraram recentemente, através da comparação de padrões de bandeamento GTG e de experimentos de hibridação *in situ* fluorescente com marcadores dos cromossomos 5 e 15, que não há alteração do padrão de bandas e na ordem dos marcadores, independentemente da posição do centrômero, excluindo assim a possibilidade de inversão, deleção ou duplicação, e atribuindo a variação ao reposicionamento centromérico.

O reposicionamento centromérico também foi demonstrado no cromossomo X de *Saimiri* que difere em morfologia dos de outros primatas, o que era atribuído a uma inversão pericêntrica (Rocchi *et al.*,2012). Estes novos centrômeros, chamados de neocentrômeros evolutivos, se caracterizam pelo surgimento de um centrômero em uma nova região cromossômica acompanhado da inativação do centrômero antigo.

A heterocromatina constitutiva, evidenciada após bandeamento CBG dos cromossomos de *Saimiri*, foi encontrada nas regiões pericentroméricas e subteloméricas da maioria dos cromossomos submetacêntricos e também em regiões intersticiais de alguns cromossomos (Jones e Ma, 1975; Lau e Arrighi, 1976; Yonenaga-Yassuda e Chu, 1985; Chiatante *et al.*, 2017). O padrão de bandeamento CBG nos cromossomos 5 e 15 varia de acordo com sua morfologia. Quando são acrocêntricos, há heterocromatina apenas na região pericentromérica, mas quando são submetacêntricos, além da banda pericentromérica, também há heterocromatina na região subtelomérica (Lau e Arrighi, 1976; Hershkovitz, 1984).

Também foi descrita variação no conteúdo de heterocromatina do cromossomo 14. Alguns indivíduos possuem um bloco adicional de heterocromatina no braço curto na região acima do centrômero, aumentando consideravelmente o tamanho do cromossomo (Jones e Ma, 1975; Ma e Jones, 1975; Moore *et al.*, 1990; Chiatante *et al.*, 2017). Geralmente este polimorfismo ocorre em heterozigose e foi observado mais frequentemente em indivíduos oriundos da Bolívia (Figura 6) (Moore *et al.*, 1990). Chiatante *et al*. (2017) descreveram o cromossomo 14 mais comum como isoforma A e a variante com mais heterocromatina como isoforma B. Também demonstraram que houve

uma inversão paracêntrica na isoforma B e sugeriram que sequências teloméricas podem ter sido translocadas para a região intersticial do braço curto deste cromossomo.

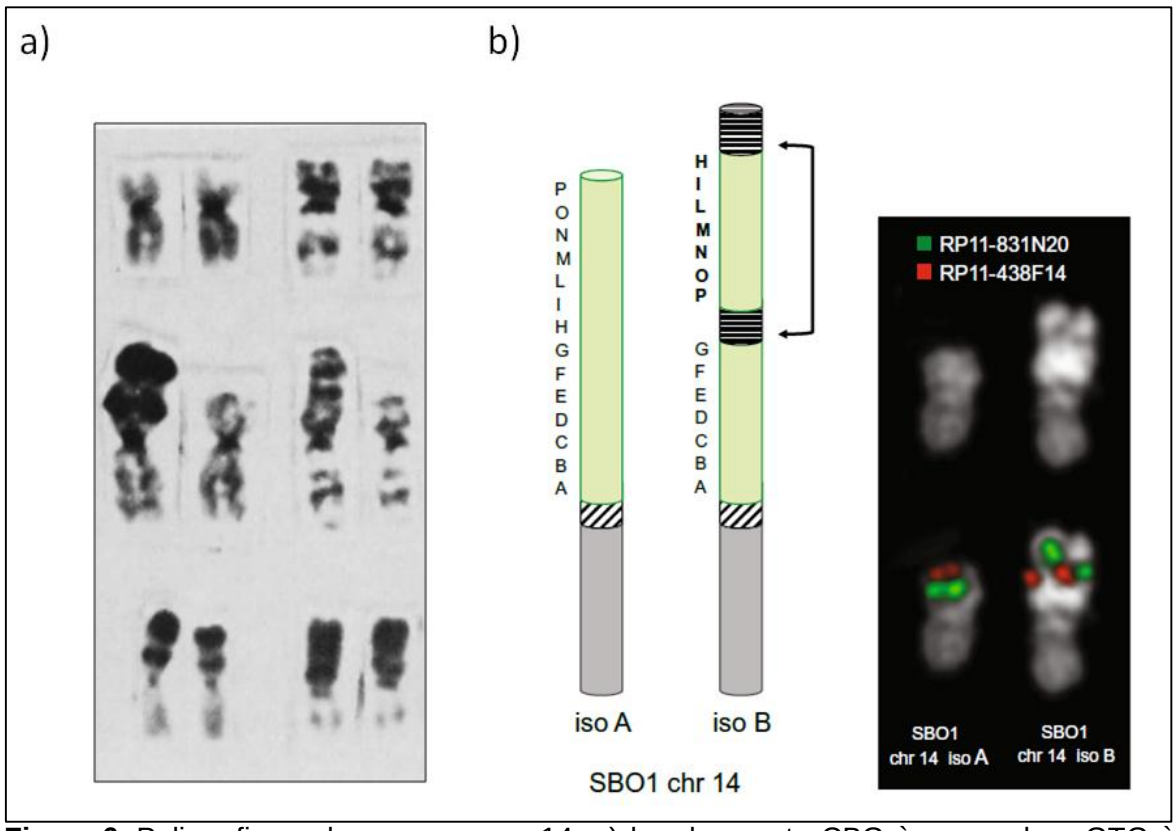


Figura 6. Polimorfismo do cromossomo 14: a) bandeamento CBG à esquerda e GTG à direita, de cima para baixo: o polimorfismo está ausente, em heterozigose e em homozigose; b) diagrama das isoformas A e B (as letras à esquerda indicam os marcadores usados, as linhas horizontais são os blocos de heterocromatina e as linhas diagonais mostram o centrômero) e FISH mostrando os marcadores invertidos no cromossomo 14 (a sonda verde corresponde ao marcador H e a vermelha ao P). Fontes: a) Jones e Ma (1975) e b) Chiatante *et al*. (2017).

1.3. DNAs Satélites

Sequências de DNA repetitivo correspondem a uma proporção grande dos genomas de eucariotos, chegando a compor mais da metade do conteúdo de DNA nuclear (Biscotti *et al.*, 2015). Os principais tipos de DNA repetitivo são as repetições dispersas, grupo dos elementos transponíveis, e as repetições em tandem (Figura 7) (Richard *et al.*, 2008; Biscotti *et al.*, 2015). As repetições em tandem compreendem os microssatélites, minissatélites e DNAs satellites (satDNAs). Os satDNAs são sequências de DNA repetitivo arranjadas em longas cadeias de repetição em tandem e representam uma fração abundante do genoma de quase todos os eucariotos. Essas sequências podem variar em comprimento de pares de bases, número de cópias e organização nos cromossomos (Plohl *et al.*, 2012). O termo DNA satélite se originou a partir de experimentos em que o DNA genômico era submetido a gradientes de densidade, como

os gradientes de cloreto de césio. As bandas compostas por repetições em tandem ficavam separadas do resto do DNA por diferença de densidade (conteúdo AT/GC) e eram chamadas de satélite.

Os satDNAs são o principal componente da heterocromatina constitutiva e geralmente se localizam nas regiões centroméricas, pericentroméricas e também subteloméricas dos cromossomos, normalmente coincidentes com os padrões de bandeamento CBG (Plohl *et al.*, 2012; Biscotti *et al.*, 2015). SatDNAs estão relacionados com a formação e manutenção da heterocromatina, inclusive na região dos centrômeros e telômeros, influenciando a integridade do cromossomo e a estabilidade do genoma (Biscotti *et al.*, 2015).

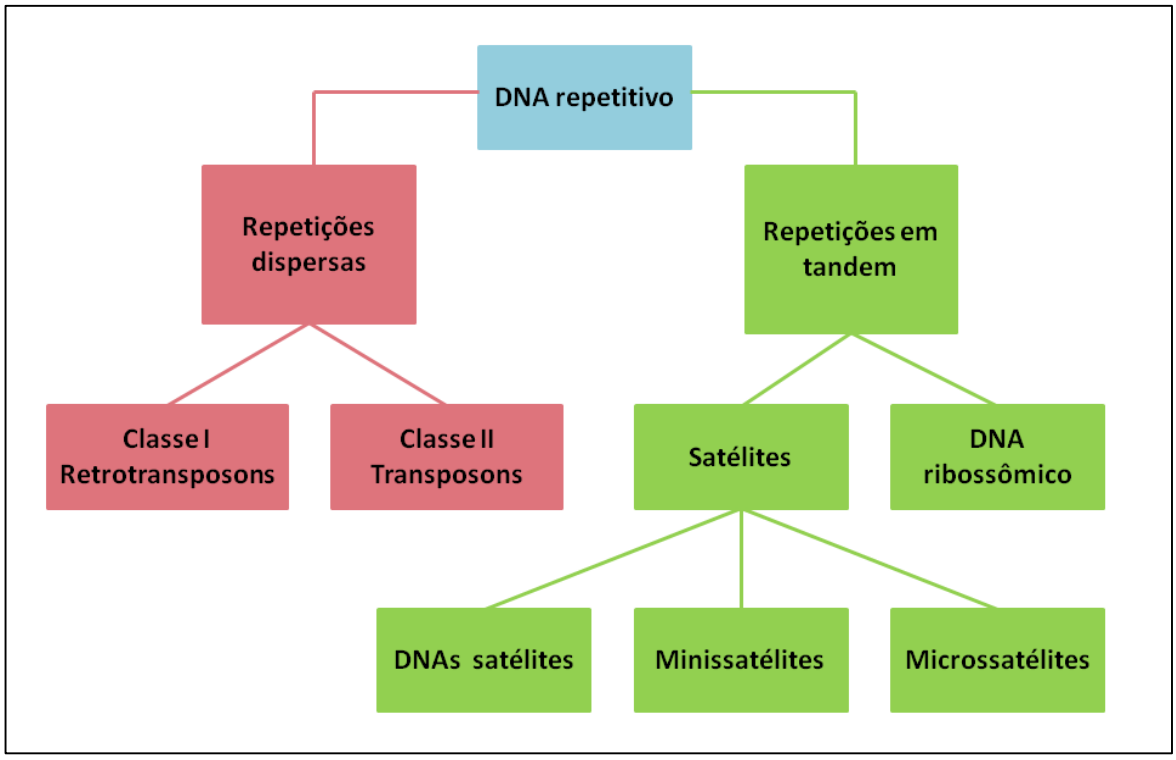


Figura 7. Principais tipos de sequências repetitivas do genoma de eucariotos. Esquema adaptado de Richard *et al.* (2008) e Biscotti *et al.* (2015).

As sequências dos satDNAs estão sujeitas a um mecanismo evolutivo chamado Evolução Combinada, que promove a homogeneização das cadeias de repetições dentro de um genoma e são fixadas numa população, resultando em maior similaridade intra- do que interespecífica (Dover, 1982). Os mecanismos moleculares responsáveis pela homogeneização das sequências dentro do genoma são: recombinação desigual, conversão gênica, transposição e amplificação por círculo rolante (Dover, 1982, 1986; Cohen e Segal, 2009). Os satDNAs evoluem rapidamente e apresentam diferenças significativas mesmo entre espécies filogeneticamente próximas, podendo ser usados como marcadores filogenéticos e até mesmo taxonômicos (Plohl *et al.*, 2012).

Comparações interespecíficas entre as sequências e localização cromossômica de satDNAs podem contribuir para o entendimento da evolução da sequência, bem como para o conhecimento da evolução cromossômica e genômica das espécies analisadas (Kuhn *et al*., 2008; Biscotti *et al.*, 2015).

Com os métodos de sequenciamento de nova geração, a disponibilidade de dados de genomas sequenciados aumentou expressivamente, contudo a quantificação e caracterização das sequências repetitivas ainda é um desafio no processo de montagem dos genomas. Uma ferramenta muito útil para detecção e caracterização de sequências repetitivas é o software RepeatExplorer, acessível por interface web (Novak *et al.*, 2010; Novak *et al.*, 2013). Suas ferramentas foram adaptadas para serem executadas on-line na plataforma Galaxy, que é um gerenciador de workflow científico (Goecks *et al.*, 2010). O software é programado para usar sequências curtas, amostradas aleatoriamente do genoma, ideal para dados gerados por sequenciamento de nova geração. Os principais módulos do RepeatExplorer consistem no *pipeline* de clusterização baseado em grafo, onde as *reads* amostradas são todas comparadas entre si e agrupadas de acordo com sua similaridade. Uma das vantagens é que o RepeatExplorer facilita a identificação *de novo* de sequências repetitivas, sem a necessidade de uma base de dados com sequências conhecidas como referência. Os *clusters* são identificados de acordo com a sua proporção do genoma e de sua similaridade com sequências repetitivas anotadas em bases de dados através de busca pelo RepeatMasker (Smit *et al.*, 2013-2015). O processo de clusterização tende a separar repetições com monômeros grandes, mas há uma ferramenta que permite juntar *clusters* de acordo com a similaridade entre eles. O RepeatExplorer tem sido usado para estudar DNAs repetitivos nos genomas de diversos organismos, como plantas (Kirov *et al.*, 2017; Křivánková *et al.*, 2017), insetos (Shah *et al.*, 2016) e mamíferos (Pagan *et al.*, 2012).

1.3.1. DNAs satélite em Platyrrhini

Os primatas são um grupo muito bem estudado em diversas áreas, já que nossa espécie está incluída neste grupo. Entretanto, assim como para a maioria dos eucariotos, estudos sobre a fração repetitiva do genoma dos Platyrrhini são escassos. Um dos satDNAs mais estudados nos primatas é o DNA satélite α ou alfa, conhecido como principal componente da região centromérica dos primatas Simiiformes (Catarrhini e Platyrrhini). Esse satDNA apresenta divergência entre espécies e até entre cromossomos da mesma espécie, como em humanos (Alves *et al.*, 1998; Alkan *et al.*, 2007).A estrutura básica do satDNA α consiste em repetições de aproximadamente 170 pb de comprimento (Alkan *et al.*, 2007). Na maioria dos Platyrrhini as unidades de repetição do satDNA α têm cerca de 340 pb de comprimento, provavelmente resultado de um evento que aconteceu

no ancestral comum dos Platyrrhini, com duas unidades de repetição diferentes entre si se comportando como uma única unidade de repetição com o dobro do tamanho original (Fanning *et al.*, 1993; Cellamare *et al.*, 2009). Em *Chiropotes satanas* e *Pithecia irrorata,* o monômero do satDNA α tem aproximadamente 550 pb e é constituído por quatro subunidades de 171 pb. A terceira subunidade é incompleta, o monômero tem cerca de 550 pb e provavelmente se originou a partir do monômero de 340 pb no ancestral da família Pitheciidae (Alves *et al.*, 1998).

Fanning *et al*. (1993), usando enzimas de restrição, identificaram um satDNA muito abundante no genoma de *Sapajus apella* (na época identificado como *Cebus apella*). Esse satDNA, denominado CapA, corresponde a 5% do genoma de *S. apella* e tem monômeros com aproximadamente 1.500 pb. O CapA foi mapeado nos blocos de heterocromatina constitutiva dos pares 4, 6, 12, 11, 13, 17 e 21 e na região telomérica de alguns cromossomos de *S. apella* (Figura 8a). Os autores sugeriram, com base em experimentos de Southern blot, que o CapA estaria presente apenas na família Cebidae, excetuando os Callithrichine (Figura 8b).

O uso de técnicas de citogenética associadas com o estudo de sequências repetitivas, especialmente dos DNAs satélites, tem-se mostrado útil por evidenciar variações cromossômicas intra- e interespecíficas, contribuindo significantemente para o entendimento sobre a evolução cromossômica e genômica de diversos grupos, permitindo ainda, em alguns casos, fazer inferências taxonômicas e filogenéticas (Garrido-Ramos *et al*, 1999; Picariello *et al*, 2002; Baicharoen *et al*, 2014; Kirov *et al*, 2017).

Os DNAs satélites do gênero *Saimiri* ainda não foram investigados e, devido aos poucos exemplares analisados, geralmente sem origem geográfica conhecida, informações sobre as sequências repetitivas deste gênero também podem ser úteis na resolução de questões taxonômicas e filogenéticas. Em vista disto, neste trabalho investigamos os dois DNAs satélite mais abundantes no genoma de *Saimiri*, o satDNA alfa e o CapA. Os resultados obtidos estão apresentados na forma de dois artigos.

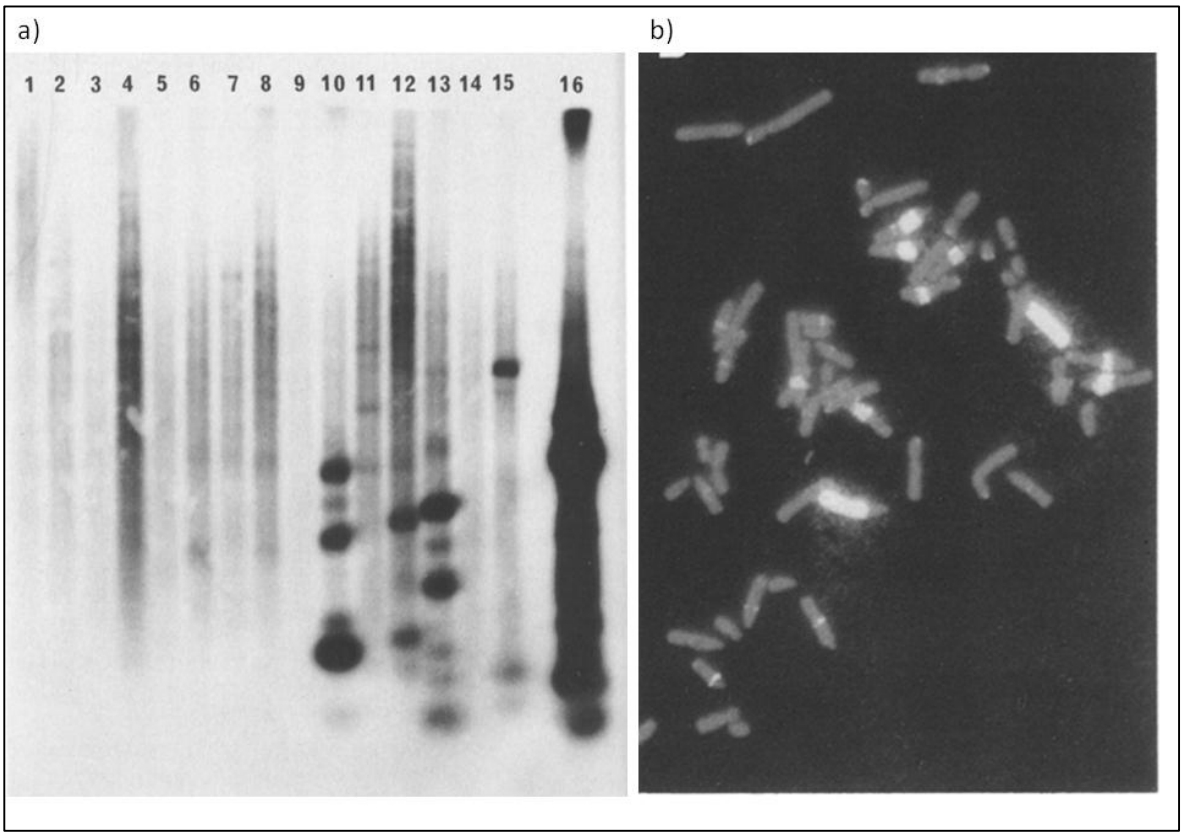


Figura 8. a) Southern Blot com sonda do CapA em primatas do Velho (1-4) e do Novo Mundo (5-16): 1) *Nycticebus coucang*; 2) *Cercopithecus aethiops*; 3) *Coiobus guereza*; 4) *Homo sapiens*; 5) *Callimico goeldii*; 6) *Leontopithecus rosalia*; 7) *Cebuella pygmaea*; 8) *Saguinus labiatis*; 9) *Callithrix jacchus*; 10) *Saimiri boliviensis*; 11) *Ateles fuscieps*; 12) *Aotus trivirgatus*; 13) *Chiropotes satanas*; 14) *Callicebus moloch*; 15) *Lagothrix lagothricha*; 16) *Sapajus apella*. b) FISH com CapA em cromossomos de *S. apella*. Fonte: Fanning *et al*. (1993)

2. OBJETIVOS

2.1. Objetivo Geral

O objetivo geral desta dissertação foi caracterizar sequências de DNA repetitivo e realizar análises citogenéticas em *Saimiri* a fim de contribuir com o conhecimento sobre a evolução cromossômica e genômica do gênero e identificar marcadores citogenômicos que possam ser usados em estudos taxonômicos.

2.2. Objetivos específicos

- Analisar citogeneticamente indivíduos de *S. boliviensis*, *S. vanzolinii*, *S. sciureus* e *S. ustus*, após coloração convencional, padrões de bandeamentos GTG e CBG;

- Isolar e caracterizar sequências de DNA repetitivo do genoma de *S. boliviensis boliviensis*;

- Mapear estas sequências nos cromossomos de *S. sciureus*, *S. boliviensis*, *S. vanzolinii*, e *S. ustus*;

- Comparar as sequências repetitivas encontradas em *Saimiri* com as descritas em outros Platyrrhini.

3. CAPÍTULO 1

O seguinte manuscrito será submetido ao periódico Chromosome Research.

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

Mirela Pelizaro Valeri[1]; Guilherme Borges Dias[1]; Camila Nascimento Moreira[2]; Yatiyo Yonenaga-Yassuda[2]; Roscoe Stanyon[3]; Gustavo Campos Silva Kuhn[1]; Marta Svartman[1*]

[1]Laboratório de Citogenômica Evolutiva, Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais; Belo Horizonte, MG, Brazil.

[2]Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, Brazil.

[3]Department of Biology, University of Florence, Florence, Italy

Running title: Satellite DNAs of *Saimiri*

*Corresponding author
E-mail: svartmanm@ufmg.br
Telephone number: 5531-34092612
Fax number: 5531-34092567

Abstract

The number of *Saimiri* species and their phylogenetic relationships has been under debate for many years and cytogenetics has had an important role in dealing with these questions. All squirrel monkeys analyzed to date presented a diploid number of 2n = 44 with variable fundamental numbers apparently correlated with geographic distribution. We explored the repetitive DNA fraction of the *S. boliviensis* genome and characterized the two most abundant satellite DNAs, alpha and CapA, which were also mapped in *S. sciureus, S. boliviensis*, *S. vanzolinii*, and *S. ustus* chromosomes. Alpha and CapA comprise ~1% and 2.2% of the *S. boliviensis* genome, respectively. Alpha has ~340 bp long monomers and was mapped to the centromeres of all chromosomes of the analyzed species, except in *S. ustus*, where alpha was not found in three chromosomes. CapA, with ~1,500 bp long monomers, is mainly located in subtelomeric regions of the short arms and in the interstitial heterochromatin of some chromosome pairs, and its distribution differs among *Saimiri* species, suggesting a species-specific chromosome marker that could be used with taxonomic purposes.

Key-words: Repetitive DNAs, Tandem Repeats, FISH, GTG-banding, CBG-banding

Introduction

Squirrel monkeys (genus *Saimiri,* family Cebidae) live in large social groups with about 20-50 individuals. They inhabit forest environments of South America in the Amazon basin and the Guianas and part of the coastal area of Central America, extending from about 10ºN to 17ºS (Lavergne et al, 2010; Chiou et al, 2011). As for many New World Monkeys (NWM; Platyrrhini), the number of species and the phylogenetic relationships of *Saimiri* are still under debate (Alfaro et al, 2015).

The uncertain taxonomy is reflected in studies based on morphological and genetic data, that suggest *Saimiri* terminal taxa varying from one to 16 (Hershkovitz, 1984, 1987; Thorington Jr., 1985; Costello et al, 1993; Boinski and Cropp, 1999; Cropp and Boinski, 2000; Groves, 2005; Rylands et al, 2013). In a recent molecular phylogeny, based on mitochondrial D-Loop and cyt *b* analyses, Alfaro et al. (2015) identified 14 clades, stressed the need of a taxonomic revision and suggested a provisional taxonomy, including seven species, some of them with several subspecies: *S. boliviensis*, *S. cassiquiarensis* (*S*. *c*. *cassiquiarensis*, *S*. *c*. *albigena*, *S*. *c*. *macrodon* A, *S*. *c*. *macrodon* B, and *S*. *c*. *macrodon* C), *S*. *collinsi*, *S*. *oerstedii* (*S. o. oerstedii* and *S. o. citronellus*), S*. sciureus*, *S. ustus* (A, B, and C lineages), and *S. vanzolinii*.

Pioneering cytogenetic studies of squirrel monkeys began in 1958 by Bender and Mettler and since then all *Saimiri* analyzed showed a diploid number of 2n = 44. *Saimiri* karyotypes differed in the number of chromosomal arms or fundamental number (FN), which ranged from FN = 74 to 78. Differences in FN in *Saimiri* have traditionally been expressed as the number of acrocentric chromosomes, which ranged from five (FN = 78) to seven pairs (FN = 74). This FN variation is apparently correlated to geographical distribution and taxonomy, making cytogenetic analyses relevant to squirrel monkeys identification. Jones at al. (1973) stated that individuals from Costa Rica, Panama and Iquitos - Peru had five acrocentric pairs (NF = 78), those originating from Georgetown – Guiana had seven (NF = 74) and specimens from Letícia – Colombia presented six pairs of acrocentric chromosomes (NF = 76). Several hypotheses have been raised to explain

this variation, the most frequent concerns pericentric inversions (Ma and Jones, 1975; Lau and Arrighi, 1976), nevertheless reciprocal translocations (Cambefort and Moro, 1978) have also been proposed. Recently, Chiatante et al. (2017) demonstrated that centromere repositioning in *Saimiri* chromosomes 5 and 15 explained the variation in FN. They found that in the karyotypes of *S. sciureus* with FN = 74 both chromosomes 5 and 15 were acrocentric, the karyotypes with FN = 76 of *S. vanzolinii* and *S. boliviensis* had a submetacentric pair 5 and an acrocentric pair 15, whereas in the karyotypes with FN = 78 of *S. sciureus* and *S. boliviensis* both chromosome pairs 5 and 15 were submetacentric.

Satellite DNAs (satDNAs) are a major type of repetitive DNA found in most eukaryotic genomes. They are arranged as tandem repeats of variable unit length, number of copies and chromosome organization (Plohl et al, 2012). SatDNAs are also usually associated with chromosome landmarks such as centromeres, telomeres, and heterochromatic regions (Fanning et al, 1993; Alkan et al, 2007; Plohl et al, 2012; Prakhongcheep et al, 2013).

SatDNA evolution is based on the concerted evolution of the monomers sequences, in which the mutations are homogenized within satDNA arrays in a genome and fixed in a reproductively isolated population (Dover, 1982; Kuhn et al. 2012; Plohl et al, 2012). Mechanisms such as transposition, gene conversion and unequal crossing-over are involved in the evolutionary process known as molecular drive, responsible for sequence homogenization (Dover, 1982). The rapid and concerted evolution of satDNAs results in high intraspecific sequence homogeneity and interspecific differences, making satDNAs good taxonomic markers and in some cases allowing their use for phylogenetic inferences (Bachman et al, 1983; Plohl et al, 2012;). Tandem repeats have already been used as cytogenetic markers in several studies, for example in primates (Baicharoen et al, 2014), frogs (Picariello et al, 2002) fish (Garrido-Ramos et al, 1999) and plants (Kirov et al, 2017).

The alphoid satDNA family is known to be part of the centromere in simian primates (Catarrhini and Platyrrhini). In most NWM, the alpha satDNA is composed by

units of ~340 bp, probably a duplication-derivative of the ~170 bp sequence found in Old World primates (Alves et al. 1994; Alkan et al. 2007). Small apes of the genus *Nomascus* and *Hylobates* are difficult to distinguish only by pelage characteristics and may produce hybrids. The alpha DNA has a genus-specific chromosome localization in these taxa, which allows the distinction of the parental chromosome sets in hybrids (Baicharoen et al, 2014). Therefore, this satDNA is helpful to precisely identify the specimens, which is important for captive breeding programs and conservation efforts.

Novel markers may help to clarify the taxonomic and phylogenetic relationships among *Saimiri* taxa. As for most eukaryotes, the repetitive DNA fraction of the *Saimiri* genomes are largely unexplored. In this work, we employed bioinformatics and cytogenetics tools to explore the repetitive DNA of the *Saimiri* genomes. 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 and a centromeric location, while CapA has ~1500 bp and is associated with constitutive heterochromatin. CapA mapped differently in the chromosomes of *Saimiri* species, suggesting that it is a potential cytogenetic marker that could be used with taxonomic purposes.

Materials and Methods

Biological samples

Chromosome spreads and genomic DNAs were obtained from fibroblast cultures of 14 squirrel monkeys: four S*. sciureus,* two *S. boliviensis*, two *S. vanzolinii*, and six *S. ustus* (Table 1). We used cell lines established years ago and the precise collection sites of some specimens are unknown. Chromosome preparations from *S. boliviensis* (SBO1 and SBO2) were previously studied by Capozzi et al. (2016) and Chiatante et al. (2017), *S. sciureus* (SSC2 and SSC3) were from the Catoctin Wildlife Preserve and Zoo, Thurmont –Maryland, and previously analyzed by Chiatante et al. (2017). SVA 321 and SVA 322 were previously described by Yonenaga-Yassuda and Chu, 1985.

Satellite DNAs identification

The RepeatExplorer pipeline (Novák et al. 2010, 2013) was used to identify satDNAs based on all to all similarity comparison of the Illumina reads of *S. boliviensis* (NCBI SRA access: SRR317821). A total of 2,230,692 Illumina reads (~100 bp long) of *S. boliviensis* were used in this analysis. The results are represented as graph-based clusters of similar reads and the shape of the clusters is an indicative of the nature of the different repeat families (e.g. globula and ring-like structures suggest tandemly organized repeats). The clustering process tends to split large repeats in several clusters, but the pipeline has a separate re-clustering tool for 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 length used for clustering and assembly was 55 and 40 bp, respectively. Clusters with globula/ring-like structure were analyzed in detail through similarity searches against the *S. boliviensis* reference genome using the BLASTn tool with default parameters (Altschul et al., 1990). Additionally, satDNA clusters were annotated through similarity searches against the whole nucleotide collection (nr/nt).

DNA isolation, PCR amplification, cloning and sequencing of satellite DNAs

Isolation of the two satDNAs identified in the previous step (CapA and alpha) was performed through polymerase chain reaction (PCR) from genomic DNA using the following specific primers sets: CapA-F (ACTTCCTCACTGACCTGTCTT), CapA-R (GGGCTGATGCTTAATGTAGCA); 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 with the pGEM-T Easy vector (Promega) and transformed to *E. coli* strain XL1-BLUE electroporation-competent cells. Recombinant colonies were capillary sequenced with the ABI3130 platform (Applied Biosystems).

Chromosome banding and fluorescent *in situ* hybridization (FISH)

Chromosome preparations were obtained from cultured cells according to Stanyon and Galeni (1991). GTG-banding and CBG-banding were performed according to Seabright (1971) and Sumner (1972), respectively.

Fluorescent *in situ* hybridization (FISH) was performed using alpha and CapA sequences as probes on metaphase spreads of the *Saimiri* species. SatDNAs 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 at 75% for 105 sec. The hybridization mix, composed of 100 ng of labeled probe in 50% formamide/2xSSC, was denatured for 10 min at 98ºC and added to the chromosome spreads. Hybridization followed at 37ºC for 16-20 hours. Post-hybridization washes comprised three baths of 2xSSC at 45ºC for 5 min each. Immunodetection was performed with neutravidin+rhodamine (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).

Results

Chromosome banding

The 14 *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 2). The karyotypes were mounted according to Stanyon et al. (2000). GTG-banding allowed the identification of all chromosomes and suggested that shifts in the centromere position

of pairs 5 and 15 explained the morphological variation of these chromosomes, resulting in different FNs (Figure 1), as already demonstrated by Chiatante et al. (2017).

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* analyzed (SVA321 and SVA 322) had a FN = 76 with a submetacentric pair 5 and an acrocentric pair 15. The second *S. boliviensis* (SBO2), one *S. sciureus* (SSC3) and all the *S. ustus* (SUS739, SUS740, SUS742, SUS746, SUS747 and SUS786) had a FN = 78 and both pairs 5 and 15 were submetacentric.

CBG-banding revealed, in addition to the pericentromeric constitutive heterochromatin, the presence of telomeric and interstitial heterochromatic blocks in many chromosome pairs. The distribution and abundance of telomeric and interstitial heterochromatin varied between species (Table 3). Generally, telomeric CBG bands were located in the short arms of submetacentric chromosomes. Light telomeric and interstitial CBG bands were sometimes difficult to identify with accuracy, so we used the DAPI staining to confirm their distribution. In all specimens the intensity of the telomeric CBG bands in the short arms of pairs 7 and 13 was much weaker in comparison with the telomeric bands of the other chromosomes. Interstitial CBG bands in both arms of chromosome 2 were more evident in *S. sciureus* than in *S. vanzolinii* and were not present in *S. ustus* and *S. boliviensis*. Pairs 5 and 15 had telomeric CBG bands in the short arms only in their submetacentric form. Only the specimens SSC770, SSC782 and SSC2 had the acrocentric form of chromosome 5 and thus did not presented the telomeric CBG bands in this pair. Telomeric CBG bands were detected in pair 15 of SSC3, SBO2 and in all *S. ustus* analyzed. In SSC 782 and SBO1, a polymorphic CBG band was detected only in one homologue of pair 14, in the telomeric and proximal regions of the short arm. The same polymorphism was detected in both homologues of pair 14 in SSC2.

Satellite DNAs identification and chromosome localization

After carefull manual curation of the RepeatExplorer results, we identified clusters that corresponded to tandemly organized repeats. The CL5 cluster comprises 22,193 reads (out of 2,230,692 total) representing ~1% of the *S. boliviensis* genome and is homologous to the alpha satDNA, known to have a centromeric location in simian primates. In *S. sciureus, S. boliviensis*, *S. vanzolinii*, and *S. ustus*, a ~340 bp monomer size was confirmed by PCR with inverse primers and sequence analysis. The alpha satDNA sequence was mapped by FISH in squirrel monkeys chromosomes (Figure 2). In *S. sciureus, S. boliviensis*, and *S. vanzolinii* the centromeric location of the alpha satDNA was in all chromosomes, whereas the repeat appeared to be absent from one homologue of chromosome 7 and from pair 12 of *S. ustus*. Furthermore, the hybridization signal in chromosome 6 of SSC 782 and SSC2 was weaker when compared to the other specimens.

The CL3 cluster was first found splitted into four clusters showing a strong connected component (CL6, CL7, Cl8 and CL12). These connected clusters were merged using the RepeatExplorer cluster merger tool, thus recovering a larger CL3. This cluster comprises ~2.2% of the genome, the second most abundant repeat family in the *S. boliviensis* genome (Supplementary Table), with ~1,500 bp repeat length confirmed by PCR. Similarity searches on the nr/nt database from GenBank using partially assembled contigs from the reads of CL3 as queries revealed that this sequence is related to a satDNA named CapA described in *Sapajus apella* (Fanning et al, 1993).

FISH with the CapA sequence as probe on chromosome spreads of *S. sciureus, S. boliviensis*, *S. vanzolinii*, and *S. ustus* showed that its chromosome location varies among *Saimiri* species and that its distribution is mostly coincident with the heterochromatic regions revealed after CBG-banding (Figures 3 and 4). CapA is mainly located in the subtelomeric regions of the short arms of submetacentric chromosomes and in the interstitial heterochromatin of some chromosome pairs (Figure 5).

In *S. sciureus*, three specimens showed different CapA localization (Figure 3). In SSC 782, CapA was located in the interstitial heterochromatin of both arms of chromosome 2 and in the subtelomeric regions of the short arms of chromosomes 4, 8, 10 and 14. In addition to the subtelomeric location on pair 14, CapA was present in the interstitial heterochromatin of the proximal region of the short arm in one homologue. CapA distribution in SSC2 was similar to SSC 782, but CapA was additionally present in subtelomeric region of pair 7 and in the interstitial heterochromatin and subtelomeric region of the short arms of both homologues of pair 14. SSC3 presented CapA in the subtelomeric regions of chromosomes 4, 5, 7, 8, 10, 13 and 15. CapA was mapped to the subtelomeric regions of the short arms of pairs 4, 5, 7, 8, 10, 11 and 13 in *S. boliviensis*, in which CapA was present in the short arm in only one homologue of pair 14, in the subtelomeric region and in the interstitial heterochromatin located in the proximal area of the short arm. In *S. vanzolinii,* CapA was mapped to the subtelomeric regions of the short arms of pairs 4, 5, 8, 10 and 11, and in *S. ustus*, CapA was evidenced in the subtelomeric regions of the short arms of chromosomes 4, 5, 7, 8, 10, 11 and 15 (Figure 4). The CapA signal on pair 2 of SSC3, *S. boliviensis*, *S. vanzolinii* and *S. ustus* displayed a much weaker intensity than that found in SSC 782 and SSC2.

Discussion

We identified two abundant satDNAs, alpha and CapA, in the genome of *Saimiri boliviensis*, and characterized them regarding their repeat size proportion in the genome and chromosome localization in *S. boliviensis*, *S. vanzolinii*, *S. sciureus* 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. The alpha repeat structure with ~340 bp found in most NWMs is probably the ancestral form in the group, and evolved from the ~170 bp monomer found in Catarrhini, and behaving as a dimeric unit (Cellamare et al, 2009).

The alpha DNA centromeric localization was observed in all chromosomes of *S. sciureus, S. boliviensis*, *S. vanzolinii* and only in *S. ustus* it appears to be absent from three chromosomes (or undetectable by FISH due to the low number of repeats in these chromosomes). Variable FNs among squirrel monkey species is due to centromere repositioning in pairs 5 and 15, which regardless of their acrocentric or submetacentric form, are apparently mature centromeres (Chiatante et al, 2017). Our sample includes specimens with both morphologies of each chromosome 5 and 15 and the alpha satDNA was always detected in the centromeres of these chromosomes, supporting the idea that these evolutionary new centromeres (ENCs) are mature centromeres.

CapA has ~1,500 bp monomers, comprises ~2.2% of the *S. boliviensis* genome and is associated with the constitutive heterochromatin revealed after CBG-banding (Figure 3). It is the second most abundant repetition in *S. boliviensis*, with only Alu elements (Short Interspersed Nuclear Elements - SINEs) representing a larger proportion of the squirrel monkey genome. CapA repeats on *Saimiri* chromosomes are mostly enriched in the subtelomeric heterochromatin of submetacentric chromosomes, but the overall distribution and abundance differed among species. The CapA distribution in chromosomes 5 and 15 was coincident with the CBG bands: they were absent in their acrocentric forms and present in the subtelomeric 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 (Stanyon et al, 2000; Stanyon et al, 2008). During the process of ENCs formation in *Saimiri,* CapA may have been lost from pair 5 when it became acrocentric, whereas it may have colonized the centromeric heterochromatic region of chromosome 15 when it became submetacentric.

SBO1 has been previously analyzed by Chiatante et al. (2017), who reported pair 14 as heteromorphic, formed by isoform A and its homologue, isoform B, that differed by a

paracentric inversion and two heterochromatic blocks, one in the telomeric and the other in the proximal region of the short arm. The same polymorphism was described in several specimens of squirrel monkeys, more frequently in those originated from Bolivia, in which heterozygotes for pair 14 were more common than homozygotes (Jones and Ma, 1975; Moore et al, 1990). Chiatante et al. (2017) suggested that the inversion in the isoform B may have carried some telomeric DNA sequences to the interstitial region of the short arm, what could explain the heterochromatic blocks. Only three specimens in our sample (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 subtelomeric region of isoform B. This may be an indication that CapA is present in the isoform B of all specimens with this variant 14 chromosome and that it may be related with the paracentric inversion, instead of the previously suggested telomeric sequences (Chiatante et al, 2017). In SSC 782, CapA was also detected in the isoform A in the subtelomeric region of the short arm. One hypothesis is that CapA was located in the subtelomeric region and part of it was relocated by the paracentric inversion to the interstitial region of the short arm. SSC2 had both homologues of pair 14 as isoform B, whereas SSC 782 presented a heteromorphic pair 14, but in both specimens the isoform B had a smaller heterochromatic block when compared with *S. boliviensis* (SBO1) or with all the specimens described by Moore et al. (1990).

The differences in CapA distribution among *Saimiri* species are not restricted to the heterochromatin, whose distribution is much less variable than that of CapA (Figures 3 and 4). Even individuals of different species with the same karyotype and FN differ in their CapA localization, as for instance SVA 322 and SBO1, both with FN = 76 (Figure 5). The specimens identified as *S. sciureus* had different CapA distributions among them and when compared with the other species (Figure 3). Without information about the collection sites of SSC2 and SSC3, it is difficult to make any inference about their relationship and they could actually be different taxa. The variable chromosomal localization of CapA among the *Saimiri* species analyzed herein suggests that it may be a species-specific

cytogenetic marker. This specificity of CapA distribution in squirrel monkeys needs to be confirmed with further studies, but it may be a valuable tool in taxonomic identification and also in the characterization of hybrids. Hybridization in squirrel monkeys has been reported in captivity and in nature (Hershkovitz, 1984; Thorington Jr, 1985; Carneiro et al, 2016) and CapA mapping may reveal the origin of chromosome sets more precisely than chromosome morphology or banding patterns.

CapA was first identified using restriction enzymes in *Sapajus apella* (previously classified as *Cebus apella*), in which it was shown to be ~1,500 bp long and to comprise about 5% of the genome (Fanning et al, 1993; Malfoy et al, 1986). This satDNA was suggested to be common to all Cebidae, except Callitrichines. The localization of CapA in *S. apella* was associated with the heterochromatin, comprising most of the long arm of chromosome 11 and some telomeres (Fanning et al. 1993).

Squirrel monkeys and the *Cebus*/*Sapajus* group are sister clades, which diverged ~15.5 million years ago (Mya; Schneider and Sampaio, 2015) and have rearranged karyotypes. The karyotype of *S. apella* (2n = 54) is considered similar to the putative ancestral NWM complement (de Oliveira et al, 2012). There are around 2.5 times more copies of CapA in *S. apella* (120,620 copies) than in *Saimiri* (47,335 copies) and its chromosome distribution also differs between *Saimiri* and *Sapajus*. These differences illustrate how dynamic the evolution of satDNAs can be.

The rapid change of satDNAs is based in its concerted evolution: the sequences are homogenized in a genome by molecular drive, a process that allows mutations to accumulate through transposition, gene conversion and unequal crossing-over and then be fixed in a population (Dover, 1982), favoring species-specific satDNA organization even in phylogenetically closely related species (Bachman et al, 1983; Plohl et al, 2012; Kuhn et al. 2008; Baicharoen et al, 2014; Kirov et al, 2017).

The rapid expansion and diversification of squirrel monkeys occurred in the Amazon basin with all the speciation events estimated in the range between 1.4 - 0.8 Mya (Alfaro et al, 2015) starting with the divergence between *S. boliviensis* and the ancestor to

all squirrel monkeys at 1.4 - 1.6 Mya (Chiou et al, 2011). Even with this recent diversification, we detected different CapA chromosomal distributions among the few *Saimiri* specimens analyzed, which suggests that this satDNA may be a potential species-specific marker for the genus. 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 do not have their precise geographical origin known. Until now, the karyotypes of only five out of the seven recently recognized species (Alfaro et al, 2015) have been reported and they are very similar, including their banding patterns. In order to validate CapA as a marker more georeferenced specimens need to be analyzed. The inclusion of species still not karyotyped would also help to clarify the current *Saimiri* taxonomic and phylogenetic questions. The availability of a cytogenomic marker would be very useful for species identification, which has important implications for taxonomic, phylogenetic and conservation studies.

In conclusion, in this work we characterized the two most abundant satellite DNAs in *Saimiri,* the alpha and CapA satDNAs, and showed variation in CapA chromosomal distribution, suggesting that it may be used as a taxonomic marker for squirrel monkeys.

Acknowledgments

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 received a Masters fellowship and GBD received a doctoral fellowship, both from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

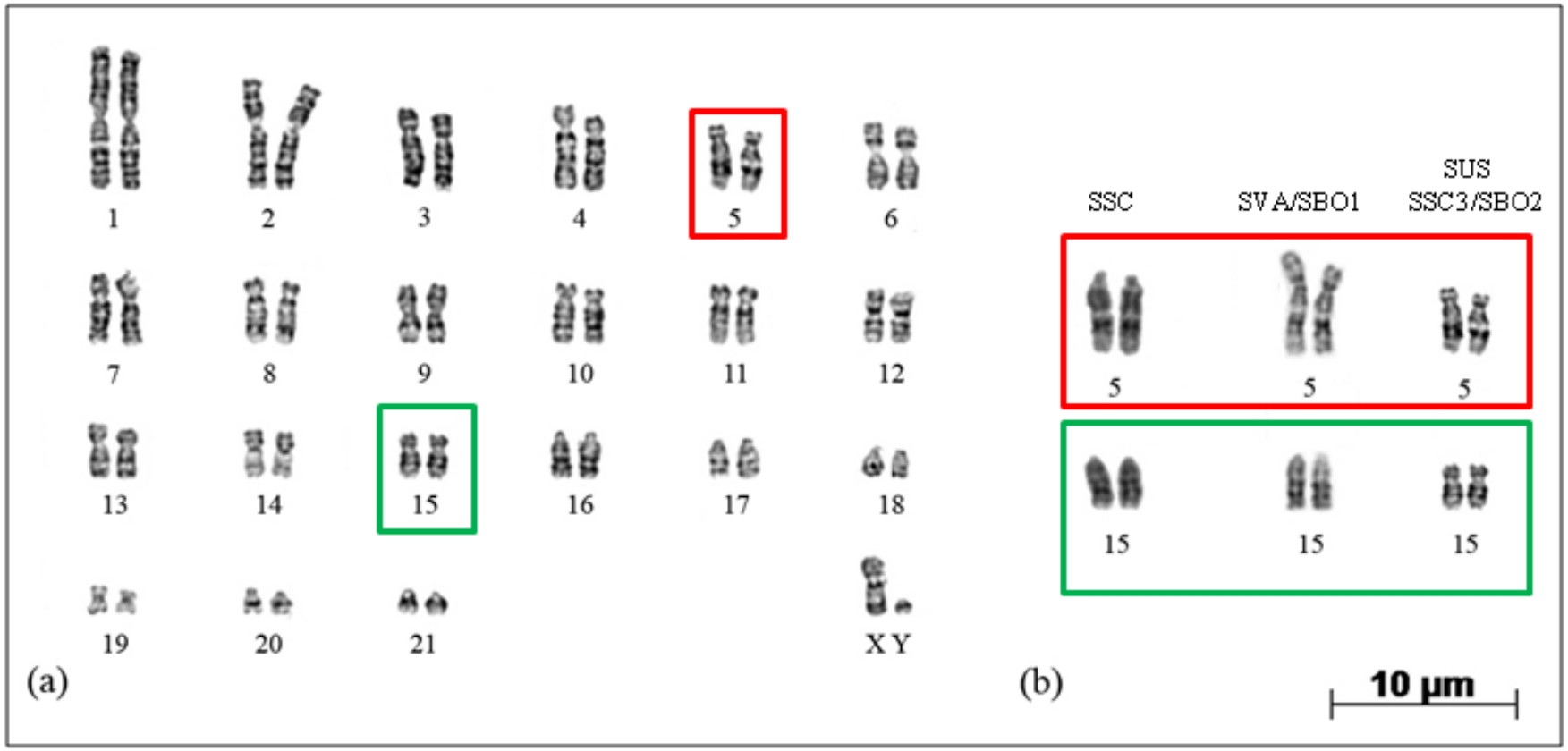


Figure 1. (a) Representative *Saimiri* karyotype with FN = 78 and five acrocentric pairs, pairs 5 and 15 are submetacentric (SUS739). (b) Acrocentric pairs 5 and 15 (SSC 770, SSC 782, SSC2); submetacentric pair 5 and acrocentric 15 (SVA 321, SVA 322 and SBO1), submetacentric pairs 5 and 15 (SUS, SSC3 and SBO2). Bar = 10 µm.

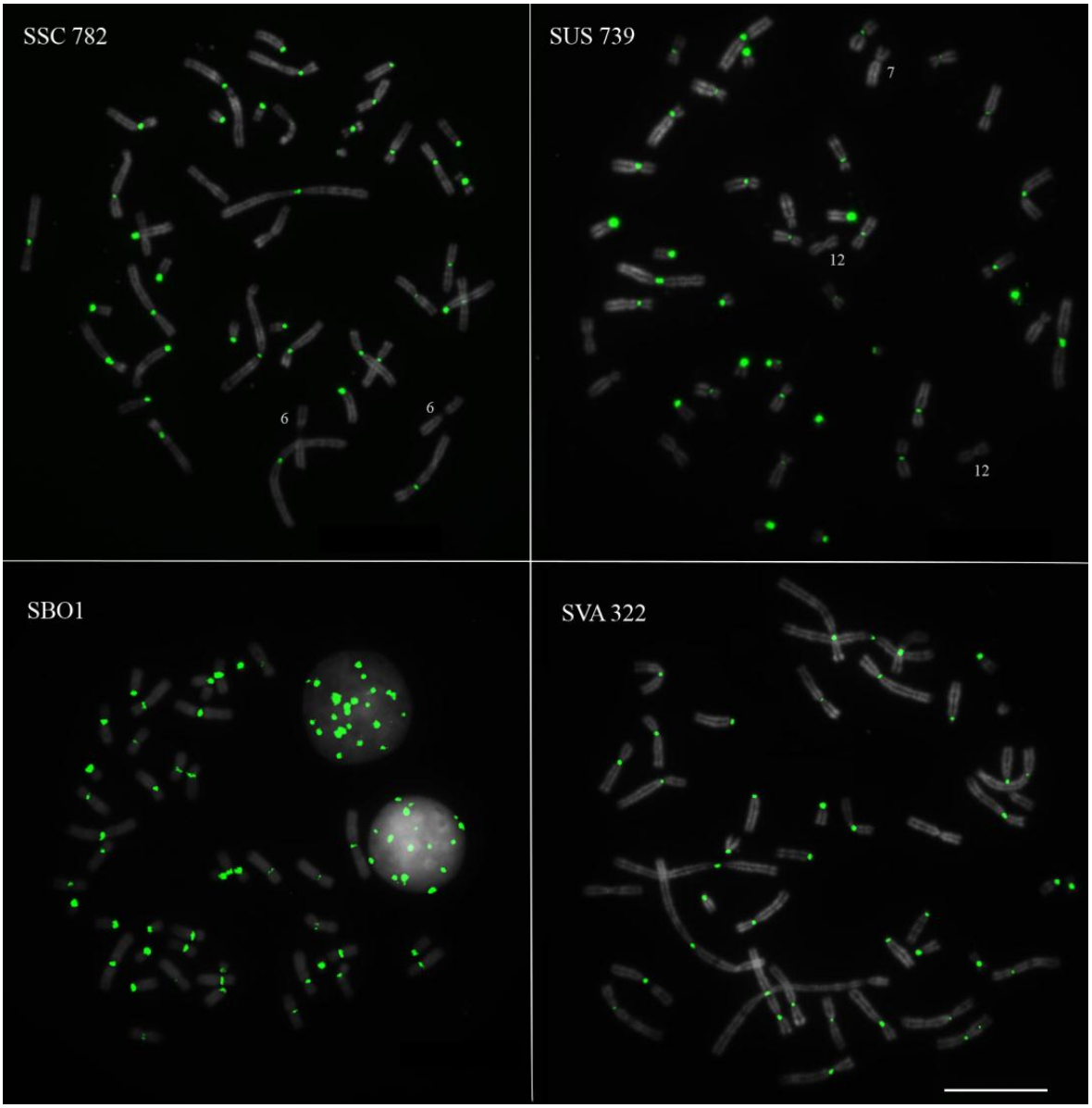


Figure 2. Metaphases of *S. sciureus* (SSC 782), *S. boliviensis* (SBO1), *S. vanzolinii* (SVA 322) and *S. ustus* (SUS 739) after FISH using the alpha satDNA as probe. The signal in pair 6 of SSC 782 was weaker compared to the other species; pair 12 and one homologue of pair 7 of SUS 739 did not show any signal. Bar = 10 µm.

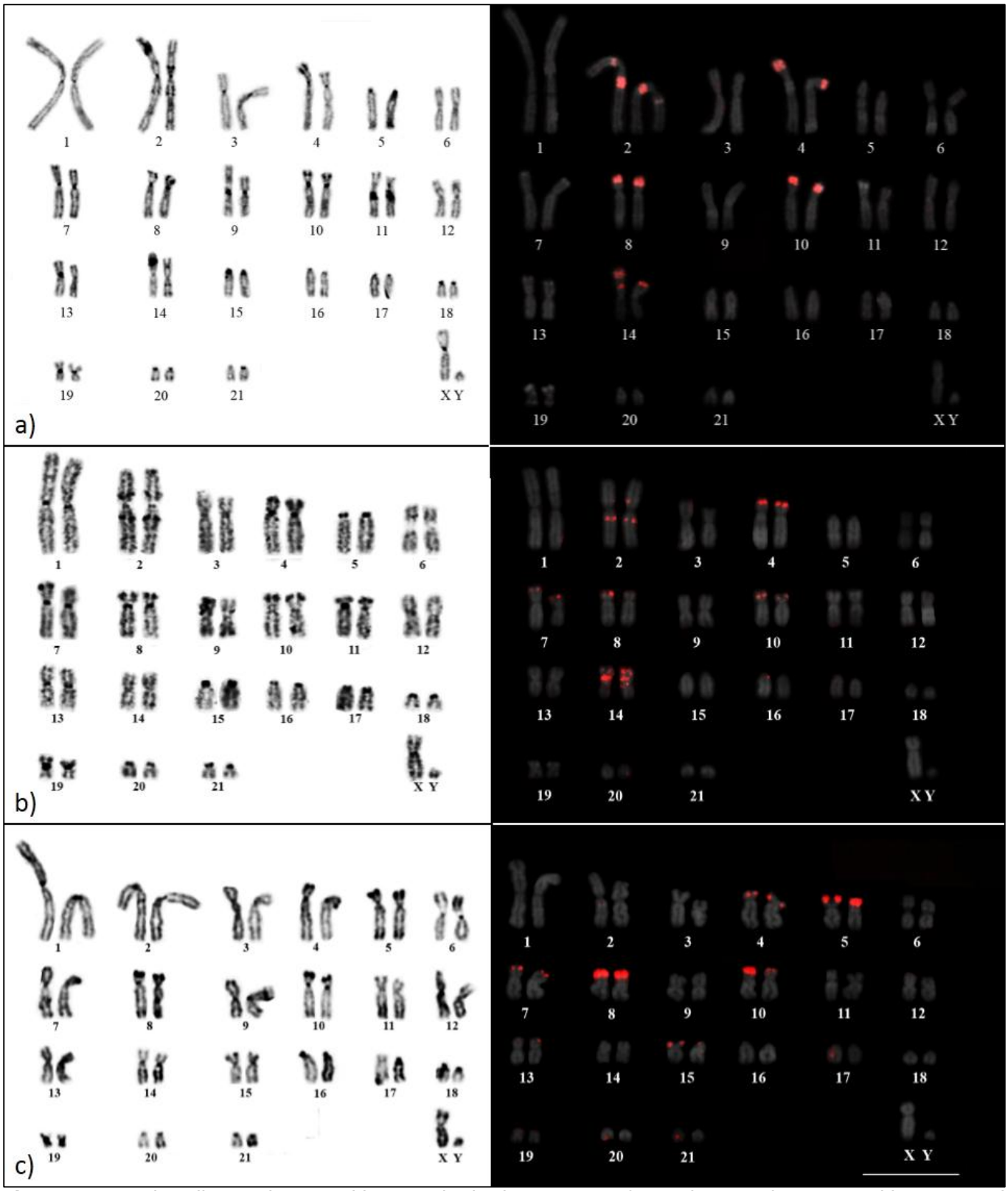


Figure 3. CBG-banding and FISH with CapA in the karyotypes of *S. sciureus*: a) SSC 782 b) SSC2 and c) SSC3. Bar = 10 μm.

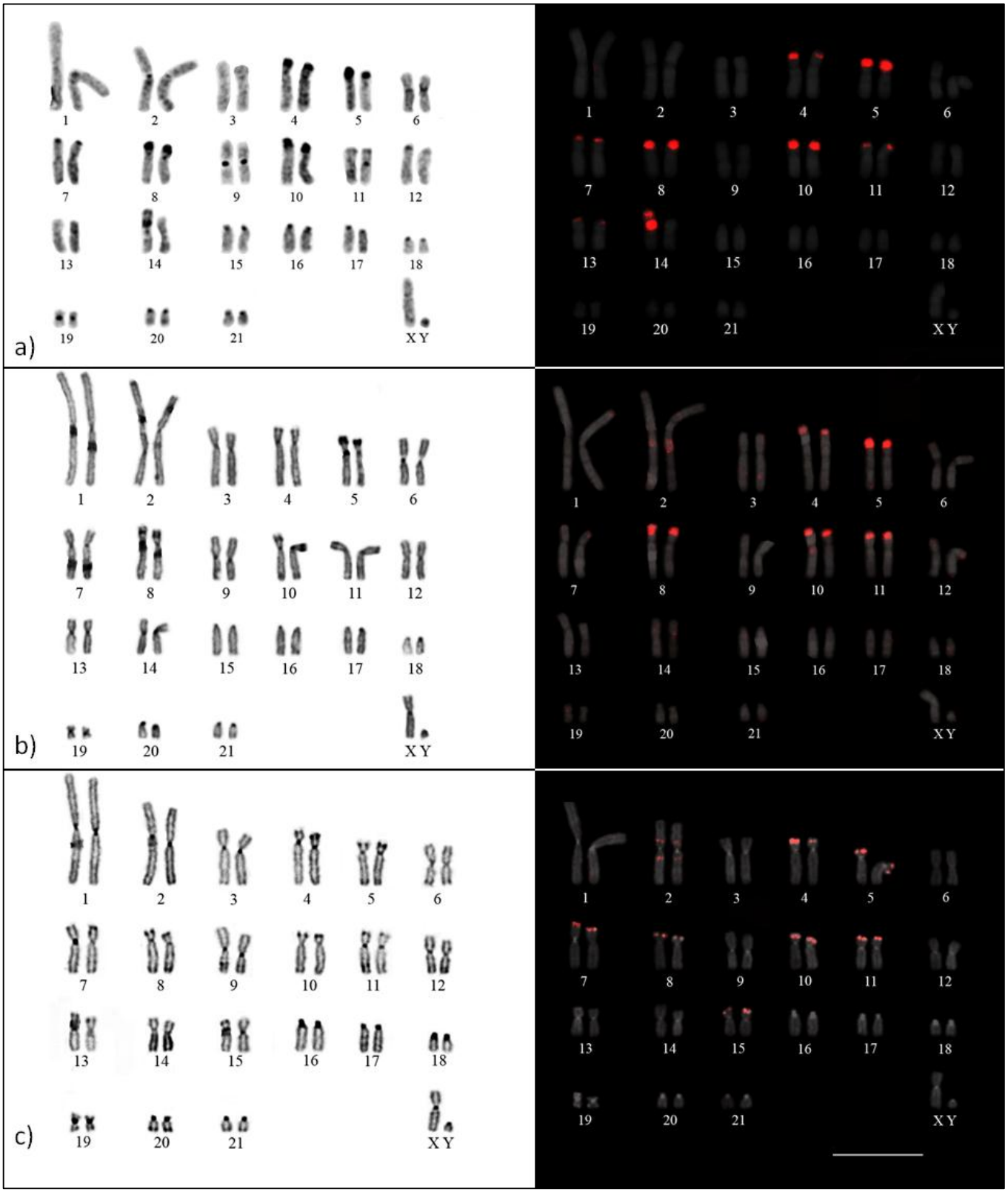


Figure 4. CBG-banding and FISH with CapA in the karyotypes of: a) *S. boliviensis* b) *S. vanzolinii* and c) *S. ustus*. Bar = 10 µm.

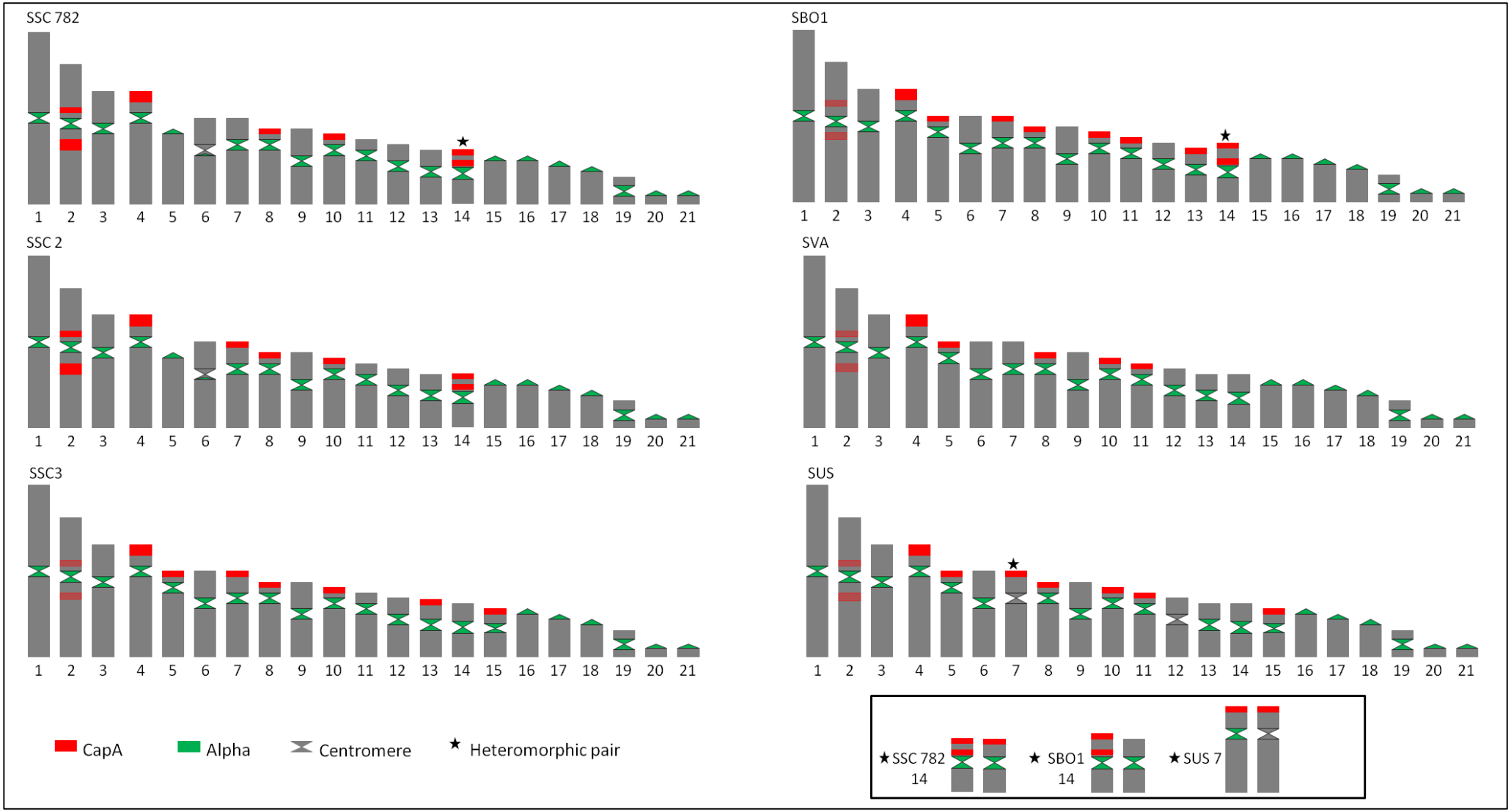


Figure 5. Ideograms of *Saimiri* chromosomes showing the localization of the alpha and CapA satDNAs in *S. sciureus* (SSC 782, SSC2 and SSC3), *S. boliviensis* (SBO), *S. vanzolinii* (SVA) and *S. ustus* (SUS). The heteromorphic pairs for satDNAs distribution are in the insets.

Table 1. Identification, sex and collection site of the analyzed specimens.

Specimen	Species	Sex	Locality
782 SSC	*S. sciureus*	Male	Presidente Figueiredo – Amazonas
770 SSC	*S. sciureus*	Female	Santarém – Pará
SSC2	*S. sciureus*	Male	Unknown
SSC3	*S. sciureus*	Male	Unknown
SBO1	*S. boliviensis*	Male	Unknown
SBO2	*S. boliviensis*	Male	Unknown
321 SVA	*S. vanzolinii*	Female	Lake Mamirauá, Tefé - Amazonas
322 SVA	*S. vanzolinii*	Male	Lake Mamirauá, Tefé - Amazonas
739 SUS	*S. ustus*	Male	Hydroelectric plant, Samuel - Rondônia
740 SUS	*S. ustus*	Female	Hydroelectric plant, Samuel - Rondônia
742 SUS	*S. ustus*	Female	Hydroelectric plant, Samuel - Rondônia
746 SUS	*S. ustus*	Female	Hydroelectric plant, Samuel - Rondônia
747 SUS	*S. ustus*	Female	Hydroelectric plant, Samuel - Rondônia
786 SUS	*S. ustus*	Female	Hydroelectric plant, Samuel - Rondônia

Table 2. Diploid and fundamental numbers, and chromosome morphology of the *Saimiri* karyotypes analyzed.

Specimen	2n	FN	Metacentrics	Submetacentrics	Acrocentrics
782 SSC 770 SSC SSC2	44	74	1, 2, 6, 9, 19	3, 4, 7, 8, 10, 11, 12, 13, 14, X	5, 15, 16, 17, 18, 20, 21
SBO1 321 SVA 322 SVA	44	76	1, 2, 6, 9, 19	3, 4, 5, 7, 8, 10, 11, 12, 13, 14, X	15, 16, 17, 18, 20, 21
SSC3 SBO2 739 SUS 740 SUS 742 SUS 746 SUS 747 SUS 786 SUS	44	78	1, 2, 6, 9, 19	3, 4, 5, 7, 8, 10, 11, 12, 13, 14, 15, X	16, 17, 18, 20, 21

2n = diploid number; FN = fundamental number

Table 3. CBG bands location in *Saimiri* chromosomes.

Species	Telomeric		Interstitial	
	Short arm	Long arm	Short arm	Long arm
S. sciureus (SSC 782)	4, 7, 8, 10, 11, 13, 14,	20, 21	2,14	2, 5, 7, 8, 9, 10, 11,14
S. sciureus (SSC2)	4, 7, 8, 10, 11, 13, 14	20, 21	2, 14	2, 5, 7, 8, 9, 10, 11,14
S. sciureus (SSC3)	4, 5, 7, 8,10, 11, 13, 15	20, 21	2	2, 5, 7, 8, 9, 10, 11, 14
S. boliviensis (SBO1)	4, 5, 7, 8,10, 11, 12, 13, 14	20	14	5, 7, 8, 9, 10, 11, 14
S. boliviensis (SBO2)	4, 5, 7, 8,10, 11, 12, 13, 14, 15	20	-	5, 7, 8, 9, 10, 11, 14
S. vanzolinii	4, 5, 7, 8,10, 11, 13	20, 21	2	2, 5, 7, 8, 9, 10, 11, 14
S. ustus	4, 5, 7, 8,10, 11, 13, 15	20, 21	-	5, 7, 8, 9, 10, 11, 13, 14

Supplementary Table. Description of the clusters retrieved by RepeatExplorer from a sample of the sequencing reads of *Saimiri boliviensis* (NCBI/SRA accession nº: SRR317821).

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

[a] Only clusters with an estimated coverage 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 outside clusters. It is a measure of uniqueness 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.

References

Alfaro JWL, Boubli JP, Paim FP, Ribas CC, da Silva MNF, 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.

Alkan C, Ventura M, Archidiacono N, Rocchi M, Sahinalp SC, et al (2007) Organization and evolution of primate centromeric DNA from whole-genome shotgun sequence data. PLoS Comput Biol, 3: 1807-1818.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol, 215: 403-410.

Alves G, Seuánez HN, Fanning T (1994) Alpha-satellite DNA in neotropical primates (Platyrrhini). Chromosoma, 103: 262-267.

Bachmann L, Schibel JM, Raab M, Sperlich D (1993) Satellite DNA as a taxonomic marker. Biochem Syst Ecol, 21: 3-11.

Baicharoen S, Miyabe-Nishiwaki T, Arsaithamkul V, Hirai Y, Duangsa-ard K, et al (2014) Locational diversity of alpha satellite DNA and intergeneric hybridization aspects in the *Nomascus* and *Hylobates* genera of small apes. PLoS One, 9: e109151.

Bender MA, Mettler LE (1958) Chromosome Studies Of Primates. Science, 128: 186-190.

Boinski S, Cropp SJ (1999) Disparate data sets resolve squirrel monkey (*Saimiri*) taxonomy: Implications for behavioral ecology and biomedical usage. Int J Primatol, 20: 237-256.

Cambefort Y, Moro F (1978) Cytogenetics and taxonomy of some south bolivian monkeys. Folia Primatol, 29: 307-314.

Capozzi O, Archidiacono N, Lorusso N, Stanyon R, Rocchi M (2016) The 14/15 association as a paradigmatic example of tracing karyotype evolution in New World monkeys. Chromosoma, 125: 747-756.

Carneiro J, Rodrigues LFD, Schneider H, Sampaio I (2016) Molecular data highlight hybridization in squirrel monkeys (*Saimiri*, Cebidae). Genet Mol Biol, 39: 539-546.

Cellamare A, Catacchio CR, Alkan C, Giannuzzi G, Antonacci F, et al (2009) New insights into centromere organization and evolution from the white-cheeked gibbon and marmoset. Mol Biol Evol, 26: 1889-1900.

Chiatante G, Capozzi O, Svartman M, Perelman P, Centrone L, et al (2017) Centromere repositioning explains fundamental number variability in the New World monkey genus *Saimiri*. Chromosoma, 126: 519-529.

Chiou KL, Pozzi L, Alfaro JWL, Di Fiore A (2011) Pleistocene diversification of living squirrel monkeys (*Saimiri* spp.) inferred from complete mitochondrial genome sequences. Mol Phylogenet Evol, 59: 736-745.

Costello RK, Dickinson C, Rosenberger AL, Boinski S, Szalay FS (1993) Squirrel monkey (Genus *Saimiri*) taxonomy. In: Martin WHKB, ed. Species, species concepts and primate evolution. Springer US, pp 177-210.

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

de Oliveira EHC, Neusser M, Muller S (2012) Chromosome evolution in New World monkeys (Platyrrhini). Cytogenet Genome Res, 137: 259-272.

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

Fanning TG, Seuánez HN, Forman L (1993) Satellite DNA-sequences in the New-World primate *Cebus apella* (Platyrrhini, Primates). Chromosoma, 102: 306-311.

Garrido-Ramos MA, de la Herran R, Jamilena M, Lozano R, Rejon CR, et al (1999) Evolution of centromeric satellite DNA and its use in phylogenetic studies of the Sparidae family (Pisces, Perciformes). Mol Phylogenet Evol, 12: 200-204.

Groves C, Wilson D, Reeder D (2005) Order Primates. In: Wilson DE, Reeder DM, eds. Mammal species of the world, 3th ed. Baltimore: The Johns Hopkins University Press, pp 111-184.

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.

Hershkovitz P (1987) 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.

Jones T, Thorington R, Hu M, Adams E, Cooper R (1973) Karyotypes of squirrel monkeys (*Saimiri sciureus*) from different geographic regions. Am J Phys Anthropol, 38: 269-277.

Jones TC, Ma, NSF (1975) Cytogenetics of the squirrel monkey (*Saimiri sciureus*), In: Primate Research. Springer US, pp 13-21.

Kirov IV, Kiseleva AV, Van Laere K, Van Roy N, Khrustaleva LI (2017) Tandem repeats of *Allium fistulosum* associated with major chromosomal landmarks. Mol Genet Genomics, 292: 453-464.

Kuhn GCS, Sene FM, Moreira-Filho O, Schwarzacher T, Heslop-Harrison JS (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*. Chromosome Res. 16: 307-324.

Kuhn GCS, Kuettler H, Moreira-Filho O, Heslop-Harrison JS (2012). The 1.688 tandem repeat DNA family of *Drosophila melanogaster*: Concerted evolution at different genomic scales. Mol Biol Evol. 29: 7-11.

Lau YF, Arrighi FE (1976) Studies of the squirrel monkey, *Saimiri sciureus*, genome. Cytogenet Genome Res, 17: 51-60.

Lavergne A, Ruiz-Garcia M, Catzeflis F, Lacote S, Contamin H, et al (2010) Phylogeny and phylogeography of squirrel monkeys (genus *Saimiri*) based on cytochrome b genetic analysis. Am J Primatol, 72: 242-253.

Ma NSF, Jones TC (1975) Added Heterochromatin Segments In Chromosomes Of Squirrel-Monkeys (*Saimiri sciureus*). Folia Primatol, 24: 282-292.

Malfoy B, Rousseau N, Vogt N, Viegas-Pequignot E, Dutrillaux B, et al (1986) Nucleotide sequence of an heterochromatic segment recognized by the antibodies to Z-DNA in fixed metaphase chromosomes. Nucleic Acids Res, 14: 3197-3214.

Moore C, Harris C, Abee C (1990) Distribution of chromosomal polymorphisms in three subspecies of squirrel monkeys (genus *Saimiri*). Cytogenet Genome Res, 53:118-122.

Novak P, Neumann P, Macas J (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. Bmc Bioinformatics, 11: 378.

Novak 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.

Picariello O, Feliciello I, Bellinello R, Chinali G (2002) 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.

Plohl M, Meštrović N, Mravinac B (2012) Satellite DNA evolution. Repetitive DNA, 7: 126-152.

Prakhongcheep O, Hirai Y, Hara T, Srikulnath K, Hirai H, et al (2013) Two types of alpha satellite dna in distinct chromosomal locations in azara's owl monkey. DNA Res, 20: 235-240.

Rylands AB, Mittermeier RA, Bezerra BM, Paim FP, Queiroz HL (2013) Family Cebidae (Squirrel Monkeys and Capuchins). In: Mittermeier RA, Rylands AB, Wilson DE, eds, Handbook of the Mammals of the World. 3. Primates. Lynx Edicions, pp 348-413.

Schneider H, Sampaio I (2015) The systematics and evolution of New World primates - A review. Mol Phylogenet Evol, 82: 348-357.

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

Stanyon R, Consigliere S, Muller S, Morescalchi A, Neusser M, et al (2000) Fluorescence in situ hybridization (FISH) maps chromosomal homologies between the dusky titi and squirrel monkey. Am J Primatol, 50: 95-107.

Stanyon R, Galleni L (1991) A rapid fibroblast-culture technique for high-resolution karyotypes. Boll Zool, 58: 81-83.

Stanyon R, Rocchi M, Capozzi O, Roberto R, Misceo D, et al (2008). Primate chromosome evolution: ancestral karyotypes, marker order and neocentromeres. Chromosome Res, 16(1), 17-39.

Sumner AT (1972) A simple technique for demonstrating centromeric heterochromatin. Exp Cell Res, 75: 304-306.

Thorington Jr RW (1985) The Taxonomy and distribution of squirrel monkeys (*Saimiri*). In: Rosenblum LA, Coe CL, eds. Handbook of squirrel monkey research. Springer US, pp 1-33.

Yonenaga-Yassuda Y, Chu T (1985) Chromosome banding patterns of *Saimiri vanzolinii* Ayres, 1985 (Primater, Cebidae). Pap Avulsos Zool, 36: 165-168.

3. CAPÍTULO 2

O seguinte manuscrito será submetido ao periódico Biology Letters.

An eutherian intronic sequence gave rise to a major satellite DNA in Platyrrhini

Mirela Pelizaro Valeri[1¶]; Guilherme Borges Dias[1¶]; Valéria do Socorro Pereira[2]; Gustavo Campos Silva Kuhn[1]; Marta Svartman[1*]

[1]Laboratório de Citogenômica Evolutiva, Departamento de Biologia Geral, Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais; Belo Horizonte, MG, Brazil

[2]Fundação Zoo-Botânica de Belo Horizonte,Avenida Otacílio Negrão de Lima, 8000 – Pampulha, Belo Horizonte, MG 31365-450, Brazil

¶ These authors contributed equally to this work

*Corresponding author
E-mail: svartmanm@ufmg.br
Telephone number: 5531-34092612
Fax number: 5531-34092567

Abstract

Satellite DNAs (satDNAs) are major components of eukaryote genomes. However, because of their quick divergence, the evolutionary origin of a given satDNA family can rarely be determined. Here we took advantage of a large collection of available primate sequenced genomes to pinpoint the origin of the CapA satDNA (~1,500 bp long repeats), first described in *Sapajus apella* (previously classified as *Cebus apella,* Platyrrhini). We show that CapA is an abundant satDNA in members of Platyrrhini, while only a single copy located within a large intron from the gene NOS1AP (*nitric oxid synthase 1 adaptor protein*) is present in the genomes of most eutherian mammals, including humans. We suggest that this intronic CapA sequence gave rise to the CapA satDNA and discuss possible mechanisms implicated in this event. To the best of our knowledge, this is the fisrt report of a single copy intronic sequence giving origin to a satDNA, with as much as 100,000 copies in some genomes.

Keywords: Repetitive DNA, Satellite DNA origin, Platyrrhini, Mammals

Eukaryote genomes are replete with repetitive DNA sequences amongst which satellite DNAs (satDNAs) are usually prominent components [1]. These tandem repeats (TRs) are commonly found as very long arrays located in heterochromatic regions of chromosomes, such as pericentric and centric heterochromatin, and subtelomeric regions [2]. Because satDNAs are fast-evolving, their exact origin is often hard to determine [1, 2].

Primate genomes are very important for medical genetics and comparative evolutionary studies. Nevertheless, as for most eukaryotes, their satDNA fraction is often unexplored. Herein, we took advantage of the available set of primate sequenced genomes to study the evolution of a satDNA called CapA. This satDNA was first described in the New World monkey (NWM) *Sapajus apella* (previously classified as *Cebus apella,* Cebidae, Platyrrhini), has large monomers of ~1,500 bp, and DNA-DNA hybridization experiments revealed its presence in other NWMs [3, 4]. However, because of the paucity of DNA sequence data available at the time of its description, CapA origin and evolution remained elusive.

Blast searches using the CapA sequence as a query described in Malfoy et al. [3] returned similar sequences from several mammals (Supplementary Table 1). Interestingly, one of these sequences (77% coverage with 79% identity) was annotated as the *Homo sapiens nitric oxid synthase 1 adaptor protein* (NOS1AP), located at the proximal region of the long arm on chromosome 1 (1q23.3). A closer inspection revealed that the region with similarity with CapA is ~1,500 bp long and is situated in the second intron of this gene (Supplementary Figure 1). Subsequent Blast searches using this intronic CapA sequence against the whole human genome retrieved no additional matches, indicating that this CapA-like sequence is not repetitive in humans.

We then used this CapA-like intronic sequence as a query against all vertebrate genomes available in NCBI. This search retrieved matches with a single CapA-like sequence present in eutherians, which is also apparently located in the orthologous region corresponding to the human NOS1AP gene. Data from species with partial chromosome assembly and chromosome painting data with human probes confirm this probable

correspondence (Supplementary Table 2). However, some eutherian clades appears do not have this CapA-like sequence, most notably the Chiroptera, some Euliplotypla and some Rodentia (Supplementary Table 2). These findings suggest that a single CapA-like sequence was present in the ancestor of eutherians and that it was likely the precursor sequence that gave rise to the CapA satellite DNA (Figure 1).

The only genomes in which we found multiple copies of CapA were those of NWMs from the family Cebidae, in which some contigs revealed the presence of a few CapA TRs (Supplementary Table 3). However, because the assembly of TRs is usually incomplete, especially for TRs with monomers larger than read size (as in the case of CapA), we also used raw Illumina data to estimate CapA abundance in available NWMs genomes (Supplementary Table 2). Strikingly, this approach revealed that CapA is an abundant tandem repeat in *Cebus capucinus* (genome proportion of 4.21%), *S. boliviensis* (1.48%) and *A. nancymaae* (0.27%) (Supplementary Table 4). In fact, CapA divergence landscapes are very similar across the three Cebidae genera, despite the significant differences in its abundance (Figure 4). A slightly higher overall divergence was detected in *S*. *boliviensis* (19.67%), compared to 17.8 and 18.46% in *A*. *nancymaae* and *C*. *capucinus*, respectively (Supplementary Figure 2). In *Callithrix jacchus,* we found CapA in low copy numbers but also tandemly arranged. Interestingly, even in the genomes where CapA achieved high abundance, the intronic ancestral locus persisted with a single CapA sequence, with the exception of *Callithrix jacchus*. In this species the intronic CapA suffered a rearrangement involving a small duplication and the insertion of an unrelated sequence of ~320 bp (Supplementary Figure 3).

All the currently available genomes of NWMs belong to species of Cebidae, preventing us from assessing the amplification status of CapA in the other two NWMs families, Atelidae and Pitheciidae. To expand our analysis to species without sequence data, we assessed CapA abundance by performing fluorescent in situ hybridization (FISH) with the human intronic CapA-like sequence as a probe onto cells of several NWMs (Table 1). We detected signs of CapA expansion in members of the three Platyrrhini families (Figure 2).

Within Cebidae, CapA was present in *Sapajus xanthosternos*, *Saimiri boliviensis* and *Aotus infulatus* and was not detected in representatives of Callithrichinae (a Cebidae subfamily). CapA also occurred as a high copy sequence in *Alouatta guariba*, *Lagothrix lagotricha* and *Brachyteles hypoxanthus*, of the family Atelidae, in which CapA was less abundant in *A. guariba* than in *L. lagotricha* and *B. hypoxanthus*. In Pitheciidae, *Chiropotes satanas* and *Pithecia irrorata*, also presented signs of CapA expansion, while *Callicebus nigrifrons* did not. CapA was found in only one small acrocentric chromosome pair of *P. irrorata* and was very abundant in *C. satanas*. In all the species in which CapA was abundant, it was associated with heterochromatin. In the species that did not present CapA, it is possible that it occurs in low repeat number or that the sequence has quite diverged, preventing its detection by FISH.

These data suggest that CapA suffered an expansion within Platyrrhini, less than ~25 million years ago (Mya), when the NWMs families first diverged [5]. CapA expansion could have predated the NWMs divergence and the satDNA could have been lost in some taxa. A second less parsimonious hypothesis is that CapA would have become a satDNA independently in multiple taxa, as only some species within each family showed CapA expansion (Figure 3). To deeply investigate the alternative possibilities, representatives of all NWMs species need to be analyzed using cytogenetic and bioinformatic tools.

We found that the large CapA satDNA present in Platyrrhini has probably originated from an intronic sequence that is still present in most eutherians. Nevertheless, it is difficult to reconstruct the steps of CapA amplification. One possibility is that CapA arose through segmental duplications. In fact, segmental duplications have already been evoked to explain hyper expansion of sequences in primate genomes [6]. A single intronic precursor sequence is still present in the ancestral locus of CapA. Therefore, prior to amplification, this CapA sequence would have had to undergo transposition to another genome region. Transposable elements (TEs) may also have participated in the process, such as L1 retrotransposons (long interspersed nucleotide element–1), through transduction events. L1 is associated with the indirect spread of other retrotranscripts, but they can also carry non-L1 DNA sequences, that are flanking L1 3’ ends, to new genomic locations [7, 8].

Duplicative transposition followed by expansion of particular euchromatic segments have been described in the pericentromeric regions of human and in the subterminal ends (generally heterochromatic) of great apes chromosomes [6,9]. CapA duplication and expansion in NWMs may be explained by a similar mechanism, in which transposition of the CapA intronic segment to heterochromatic regions in the ancestral Platyrrhini genome followed by hyper expansion through unequal crossing over would have given rise to the CapA satDNA in some species. Although we favor this segmental duplication hypothesis, the incompleteness of current NWMs genome assemblies prevents the scrutiny of this possibility to exhaustion.

In conclusion, we characterized CapA, a satDNA formed less than 25 Mya, with ~1,500bp monomers present in species of the three Platyrrhini families. In Cebidae, with the exception of Callitrichines, CapA abundance ranges from 0.27 to 5% of the genome. The CapA-like ancestral sequence is present in most eutherians, most likely embedded in the second intron of the *nitric oxide synthase 1 adaptor protein* (NOS1AP) gene, such as in *Homo sapiens*. One hypothesis for the CapA expansion in NWMs is through duplicative transposition followed by expansion through unequal crossing over.

The origin of satDNAs remains elusive. There are several reports of satDNAs that originated from TE sequences. These reports include cases involving different TE classes/superfamilies and different organisms representing both animal and plant species [10]. To the best of our knowledge, this is the fisrt report of a single copy intronic sequence giving origin to a satDNA, with as much as 100,000 copies in some genomes.

Material and Methods

We searched for sequences similar to CapA in the non-redundant nucleotide collection of GenBank using the CapA monomer described in Fanning et al. (1993) as query and the BLASTn tool [11]. This search returned hits from several mammals, particularly NWMs, and included a hit in the intron of the *nitric oxide synthase 1 adaptor protein* (NOS1AP) gene from *Homo sapiens* (Supplementary Table 1). After checking the human

reference genome (hg38) at UCSC using BLAT, we confirmed that this CapA-like sequence exists at a single locus, inside the NOS1AP gene on chromosome 1. We then used this human intronic sequence as query in BLASTn searches against all mammalian assembled genomes available at NCBI. A hit was included when query cover was ≥30% and e-value ≤ e^{-5}. Number of hits with the CapA-like sequence, query cover, identity and E-value of each search are available in Supplementary table 2.

In the genomes where multiple hits of CapA were found, we also ran RepeatMasker [12] to estimate CapA abundance. We included data from an Old World monkey (*Chlorocebus aethiops*) and a great ape (*Homo sapiens*) as negative controls. All sequence reads used in this step were obtained from the Short Read Archive at NCBI (available at http://www.ncbi.nlm.nih.gov/sra/), with accession numbers as follows: *Aotus nancymaae* SRR1692991, *Callithrix jacchus* SRR1746970, *Cebus capucinus imitator* SRR3144006, *Saimiri boliviensis* SRR317821, *Chlorocebus aethiops sabaeus* SRR5251202, and *Homo sapiens* ERR016352. Sequence reads were downloaded in fastq format using the software SRAToolkit (available at https://github.com/ncbi/sra-tools) and random samples of two million reads ranging from 101 to 125 bp were produced using the software seqtk (available at https://github.com/lh3/seqtk/). To determinate the CapA fraction of the reads, RepeatMasker was used in the following setup: sensitive mode, without searching for low complexity or bacterial insertion sequences, using wublast as search engine and a custom library containing the NOS1AP intronic CapA-like sequence. We used the alignment files generated by RepeatMasker to calculated Kimura distances of CapA fragments against the ancestral sequence (NOS1AP intronic CapA-like) using the utility script calcDivergenceFromAlign.pl from the RepeatMasker package. Results were then imported to RStudio and plotted.

To investigate CapA abundance in species for which no genome data was available, we performed fluorescent in situ hybridization (FISH) with the intronic CapA probe. This probe was obtained after PCR amplification of human DNA with primers flanking the CapA-like intronic sequence from NOS1AP (CapA-F: ACTTCCTCACTGACCTGTCTT; CapA-R: GGGCTGATGCTTAATGTAGCA). The PCR products were purified, cloned and sequenced

to ensure specificity. Chromosome spreads of several NWMs were obtained from fibroblast or lymphocyte cultures (Table 1) and their geographic origin is mostly unknown. FISH was performed with 200 ng of biotin labeled probes. Chromosomes were denatured in 70% formamide/2xSSC at 75ºC for 105-120 sec. The hybridization mix (200 ng of labeled probe in 50% formamide/2xSSC) was denatured for 10 min at 98ºC and added to the chromosome spreads. Hybridization followed at 37ºC for 16-20 hours. Post-hybridization washes comprised two 2xSSC baths at 37ºC for 5 min each. Immunodetection was performed with neutravidin+rhodamine (Roche Applied Science) and the slides were mounted with DAPI in Slowfade 1:500 (Life Technologies).

ACKNOWLEDGEMENTS

The authors thank the following people for their assistance in obtaining samples for this study: Dr. Yatiyo Yonenaga-Yassuda and MSc Camila do Nascimento Moreira, Universidade de São Paulo; Dr. Roscoe Stanyon, University of Florence; Prof. Alan Lane de Melo, Universidade Federal de Minas Gerais. 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 received a Masters fellowship and GBD received a doctoral fellowship, both from Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES).

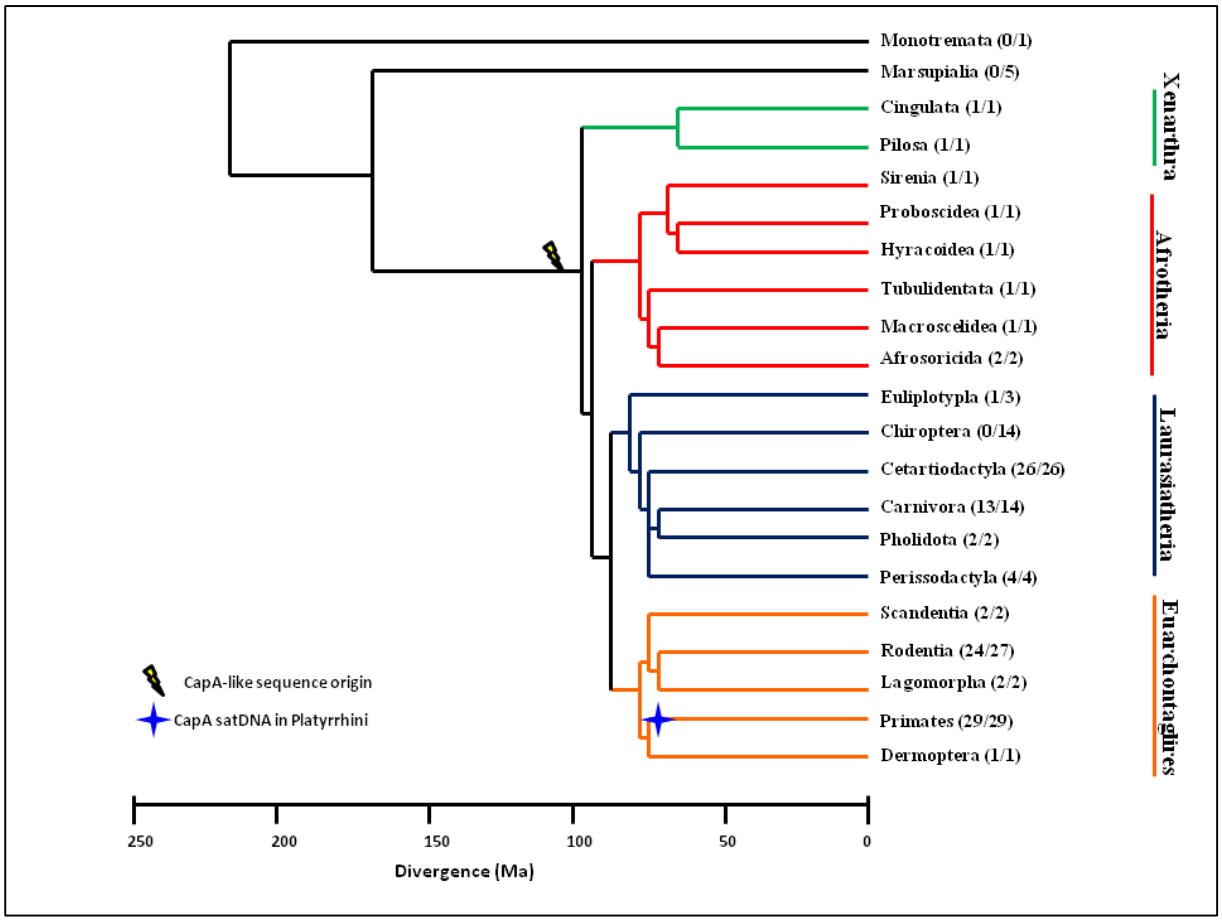


Figure 1. Diagramatic representation of the mammalian phylogeny adapted from Foley et al. [13]. The CapA precursor probably predates the eutherian radiation and was likely lost in a few clades. CapA was found as a satDNA in several Platyrrhini. In parentheses is the number of species in wich CapA was found in the total analyzed.

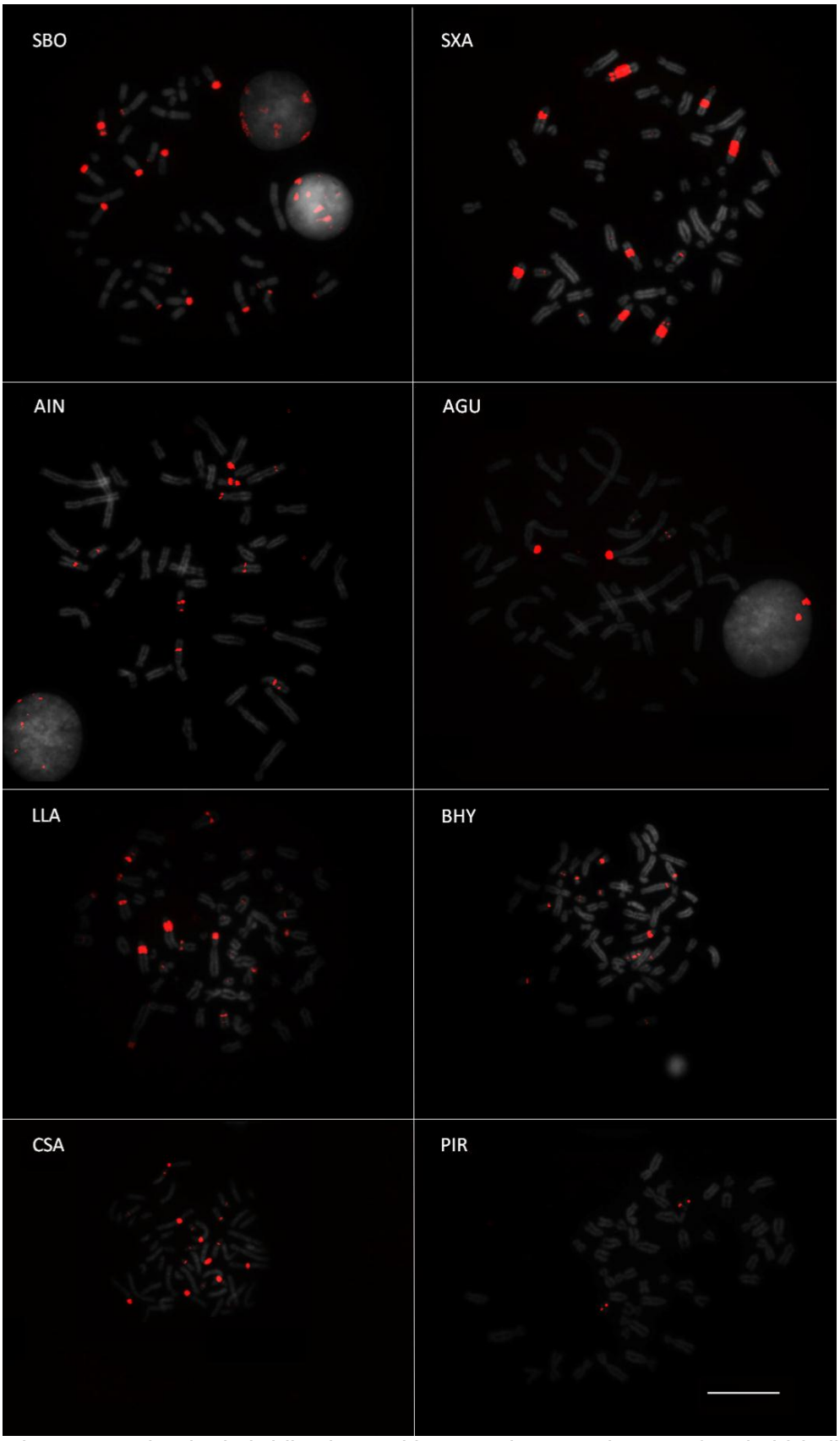


Figure 2. Fluorescent *in situ* hybridizations with CapA in metaphases of *Saimiri boliviensis* (SBO), *Sapajus xanthosternos* (SXA), *Aotus infulatus* (AIN), *Alouatta guariba* (AGU), *Lagothrix lagotricha* (LLA), *Brachyteles hypoxanthus* (BHY), *Chiropotes satanas* (CSA) and *Pithecia irrorata* (PIR). Bar = 10μm

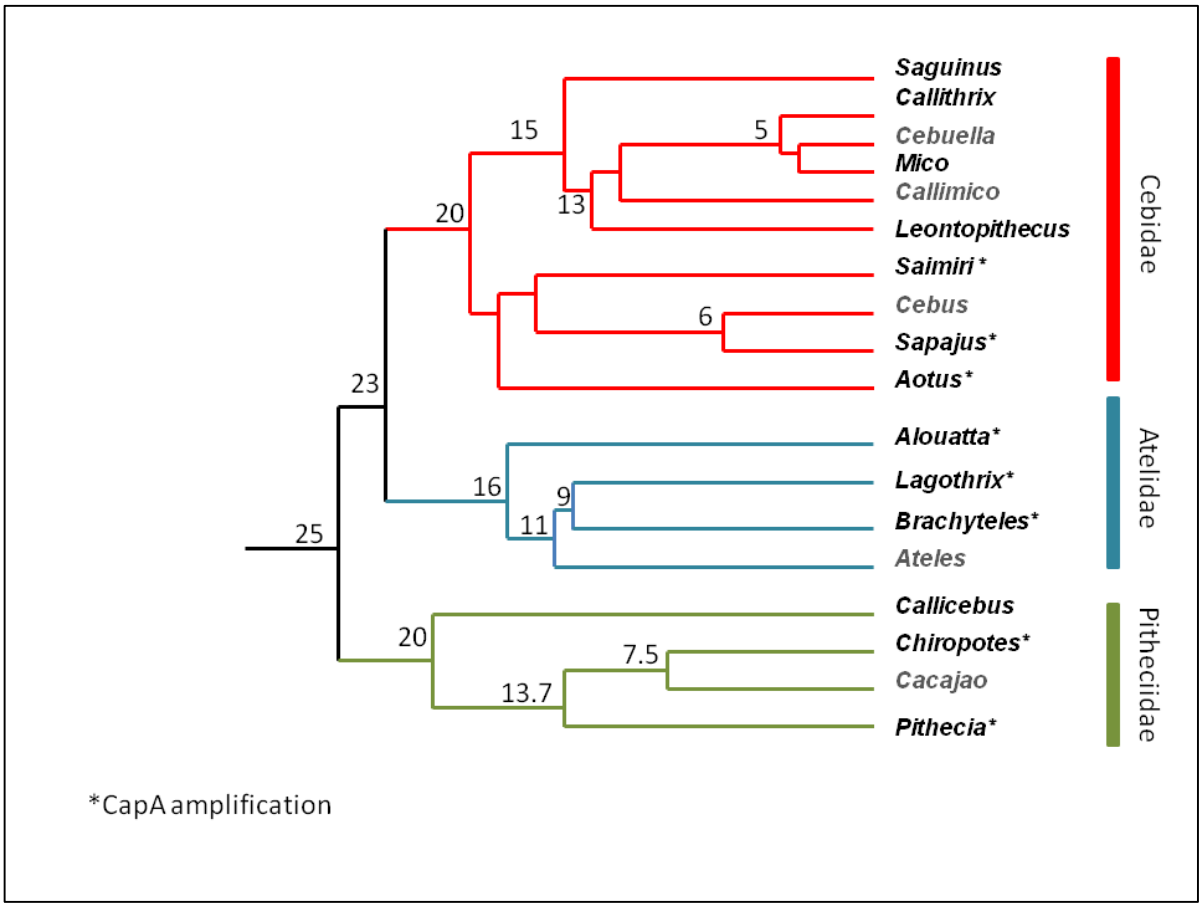
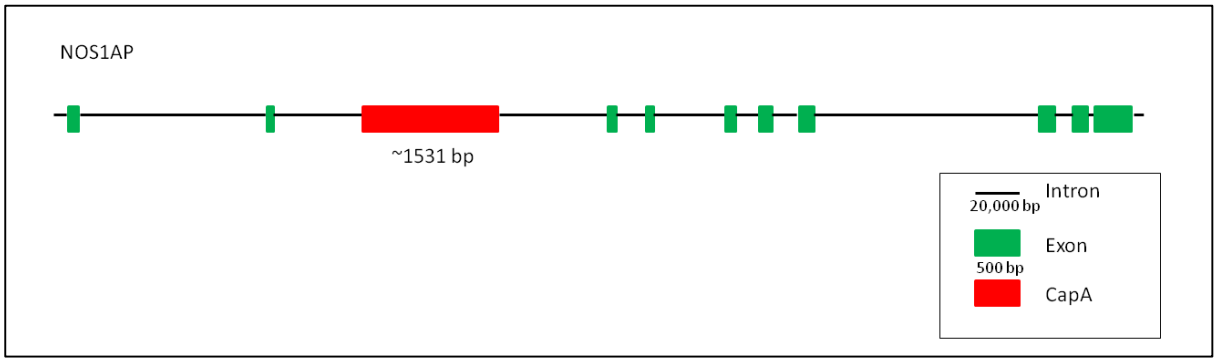
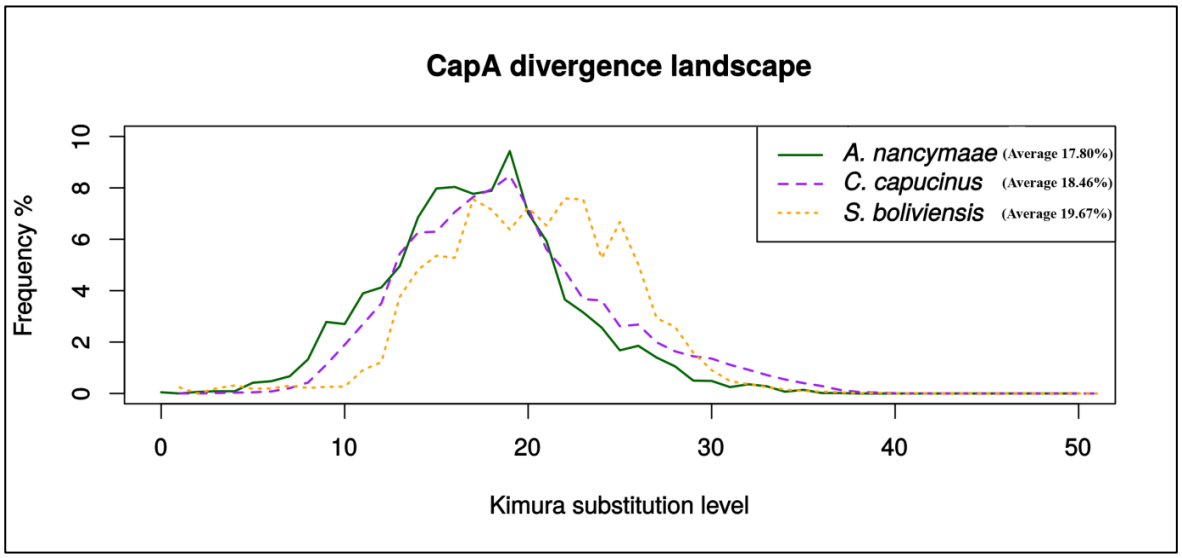


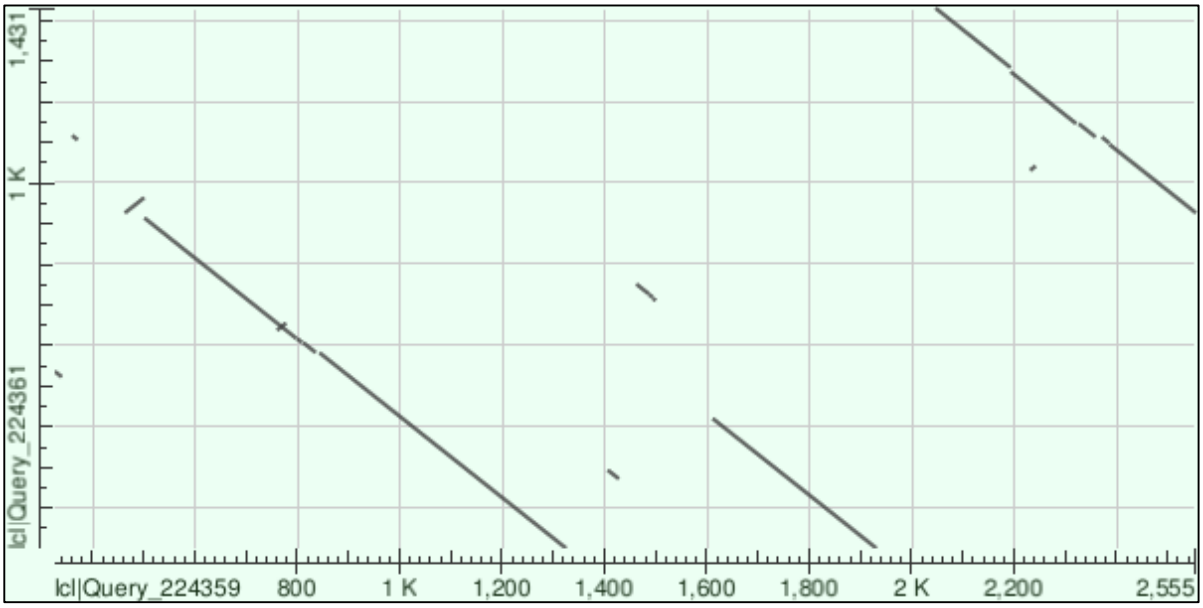
Figure 3. Distribution of CapA among Platyrrhini families. Diagramatic representation of the Platyrrhini phylogeny adapted from Schneider & Sampaio [5]. The number at nodes indicates Mya. The genera with representatives analyzed are shown in bold and the asterisk indicates CapA amplification.



Supplementary Figure 1. Schematic representation of the *Homo sapiens nitric oxid synthase 1 adaptor protein* (NOS1AP) gene, with the CapA-like sequence located at intron 2.



Supplementary Figure 2. CapA divergence landscape within Cebidae: *Aotus nancymaae*, *Cebus capucinus* and *Saimiri boliviensis*.



Supplementary Figure 3. Dot blot comparison of the CapA-like sequence from NOS1AP and contig with CapA hit in *C. jacchus* chr1c06988 (accession number BBXK01006988.1).

Table 1. Species in which FISH with the human intronic CapA-like were performed.

Species	Family	Procedence
Saguinus imperator (SIM)	Cebidae	Fundação Zoo-Botânica de Belo Horizonte
Callithrix penicillata (CPE)	Cebidae	Universidade Federal de Minas Gerais
Mico argentata (MAR)	Cebidae	Universidade de São Paulo
Leontopithecus rosalia (LRO)	Cebidae	Fundação Zoo-Botânica de Belo Horizonte
Sapajus xanthosternos (SXA)	Cebidae	Fundação Zoo-Botânica de Belo Horizonte
Saimiri boliviensis (SBO)	Cebidae	University of Florence
Aotus infulatus (AIN)	Cebidae	Fundação Zoo-Botânica de Belo Horizonte
Alouatta guariba (AGU)	Atelidae	Fundação Zoo-Botânica de Belo Horizonte
Lagothrix lagotricha (LLA)	Atelidae	Fundação Zoo-Botânica de Belo Horizonte
Brachyteles hypoxanthus (BHY)	Atelidae	Fundação Zoo-Botânica de Belo Horizonte
Callicebus nigrifrons (CNI)	Pitheciidae	Fundação Zoo-Botânica de Belo Horizonte
Chiropotes satanas (CSA)	Pitheciidae	Fundação Zoo-Botânica de Belo Horizonte
Pithecia irrorata (PIR)	Pitheciidae	Fundação Zoo-Botânica de Belo Horizonte

Supplementary Table 1. Blast searches using the CapA sequence as query against the non-redundant nucleotide collection of GenBank.

Order	Species	Description	Coverage [%]	Identity [%]	E-value	Accession number
Primates	*Aotus lemurinus*	Aotus lemurinus marker 289 genomic sequence	33	75	2E-44	JQ933059.1
Primates	*Aotus lemurinus*	Aotus lemurinus marker 667 genomic sequence	35	79	4E-34	JQ932269.1
Primates	*Aotus lemurinus*	Aotus lemurinus marker 757 genomic sequence	27	81	2E-50	JQ932337.1
Primates	*Aotus trivirgatus*	Aotus trivirgatus (clone A-15) satellite DNA	89	70	3E-79	L78708.1
Primates	*Cacajao calvus*	Cacajao calvus marker 011 genomic sequence	47	71	6E-69	JQ932173.1
Primates	*Cacajao calvus*	Cacajao calvus marker 092 genomic sequence	59	71	3E-85	JQ932467.1
Primates	*Cacajao calvus*	Cacajao calvus marker 248 genomic sequence	34	79	3E-60	JQ932469.1
Primates	*Cacajao calvus*	Cacajao calvus marker 337 genomic sequence	32	79	2E-62	JQ932156.1
Primates	*Cacajao calvus*	Cacajao calvus marker 546 genomic sequence	26	84	5E-51	JQ931610.1
Primates	*Cacajao calvus*	Cacajao calvus marker 569 genomic sequence	59	70	8E-74	JQ931662.1
Primates	*Cacajao calvus*	Cacajao calvus marker 639 genomic sequence	40	79	1E-66	JQ931882.1
Primates	*Cacajao calvus*	Cacajao calvus marker 662 genomic sequence	34	75	2E-43	JQ931861.1
Primates	*Cacajao calvus*	Cacajao calvus marker 910 genomic sequence	33	76	5E-45	JQ931773.1
Primates	*Cacajao calvus*	Cacajao calvus marker 966 genomic sequence	21	84	7E-18	JQ931299.1
Primates	*Cacajao calvus*	Cacajao calvus marker 997 genomic sequence	21	81	3E-23	JQ931749.1
Primates	*Cebus apella*	Cebus apella DNA interacting with antibodies to Z-DNA (clone pCH11) (satellite DNA 1520 bp repeat HaeIII fragment of R-positive heterochromatin segments of fixed metaphase chromosomes)	100	100	0.0	X03835.1
Primates	*Cebus capucinus imitator*	PREDICTED: Cebus capucinus imitator uncharacterized LOC108293577 (LOC108293577), ncRNA	22	78	2E-32	XR_001820741.1
Primates	*Cebus capucinus imitator*	PREDICTED: Cebus capucinus imitator uncharacterized LOC108298982 (LOC108298982), transcript variant X1, ncRNA	21	88	2E-76	XR_001822774.1
Primates	*Homo sapiens*	Homo sapiens nitric oxide synthase 1 adaptor protein (NOS1AP), RefSeqGene on chromosome 1	77	79	4E-65	NG_015979.1
Primates	*Homo sapiens*	Human DNA sequence from clone RP11-384L19 on chromosome 1, complete sequence	71	79	4E-65	AL450163.12
Primates	*Macaca fascicularis*	Macaca fascicularis complete genome, chromosome chr1	77	82	3E-61	LT160000.1
Carnivora	*Mustela putorius furo*	PREDICTED: Mustela putorius furo uncharacterized LOC101686638 (LOC101686638), ncRNA	20	68	0,000006	XR_001179155.1

Supplementary Table 2. Blast searches using the CapA sequence as query against all mammalian assembled genomes available at NCBI.

Order	Species	Hits	Coverage [%]	Identity [%]	E-value	Accession Number	Chromosomes
Afrosoricida	*Chrysochloris asiatica*	1	92	69	1E-150	AMDV01273478.1	-
Afrosoricida	*Echinops telfairi*	1	61	68	7E-50	AAIY02149348.1	-
Carnivora	*Acinonyx jubatus*	1	84	75	1E-47	LLWD01001073.1	-
Carnivora	*Ailuropoda melanoleuca*	1	80	72	1E-144	ACTA01121436.1	-
Carnivora	*Canis lupus familiaris*	1	71	74	4E-51	AOCS01133076.1	-
Carnivora	*Felis catus*	1	81	74	1E-50	ACBE01568514.1	-
Carnivora	*Leptonychotes weddellii*	1	87	72	3E-143	APMU01086890.1	-
Carnivora	*Mustela putorius furo*	1	61	70	4E-107	AGTQ01057656.1	-
Carnivora	*Odobenus rosmarus divergens*	1	90	72	4E-141	ANOP01021768.1	-
Carnivora	*Panthera tigris*	1	71	73	4E-46	ATCQ01118692.1	-
Carnivora	*Ursus arctos*	0	-	-	-	-	-
Carnivora	*Ursus maritimus*	1	78	73	6E-145	AVOR01015736.1	-
Carnivora	*Ailurus fulgens styani*	1	72	71	9E-118	LNAC01000762.1	-
Carnivora	*Lycaon pictus*	1	69	73	5E-49	LPRB01000038.1	38
Carnivora	*Neomonachus schauinslandi*	1	87	72	1E-141	NINY01007742.1	-
Carnivora	*Panthera pardus*	1	71	73	8E-49	LQGZ01114650.1	-
Cetartiodactyla	*Bison bison*	1	67	77	3E-169	JPYT01044221.1	-
Cetartiodactyla	*Bos indicus*	2	51	77	6E-171	AGFL01160664.1	3; 20
Cetartiodactyla	*Bos mutus*	1	73	77	3E-169	AGSK01136958.1	-
Cetartiodactyla	*Bos taurus*	1	74	77	2E-169	DAAA02006932.1	3
Cetartiodactyla	*Bubalus bubalis*	1	74	72	3E-131	AWWX01468899.1	-
Cetartiodactyla	*Camelus bactrianus*	1	98	71	0	JARL01015881.1	-
Cetartiodactyla	*Camelus dromedarius*	1	98	71	0	JDVD01029756.1	-
Cetartiodactyla	*Camelus ferus*	1	98	71	0	AGVR01001381.1	-
Cetartiodactyla	*Capra aegagrus*	1	72	76	3E-159	JXYW01026932.1	-
Cetartiodactyla	*Capra hircus*	1	72	76	3E-159	LWLT01000004.1	3
Cetartiodactyla	*Capreolus capreolus*	1	73	75	8E-151	CCMK010066276.1	-

Cetartiodactyla	*Odocoileus virginianus*	1	62	77	1E-167	MLBE01010430.1	-
Cetartiodactyla	*Ovis aries*	1	72	73	1E-132	AMGL02020864.1	-
Cetartiodactyla	*Pantholops hodgsonii*	1	72	75	1E-148	AGTT01187235.1	-
Cetartiodactyla	*Sus scrofa*	1	93	69	3E-162	LUXT01055614.1	-
Cetartiodactyla	*Vicugna pacos*	1	98	71	0	JEMW01024141.1	-
Cetartiodactyla	*Ammotragus lervia*	1	71	76	8E-157	NIVO01000753.1	-
Cetartiodactyla	*Giraffa camelopardalis tippelskirchi*	1	72	76	2E-159	LVKQ01026506.1	-
Cetartiodactyla	*Okapia johnstoni*	1	79	76	1E-161	LVCL010210031.1	-
Cetartiodactyla	*Balaenoptera acutorostrata*	1	99	71	0	ATDI01015654.1	-
Cetartiodactyla	*Balaenoptera bonaerensis*	1	85	73	0	BAUQ01712362.1	-
Cetartiodactyla	*Lipotes vexillifer*	1	98	72	0	AUPI01082027.1	-
Cetartiodactyla	*Orcinus orca*	1	99	72	0	ANOL02062978.1	-
Cetartiodactyla	*Physeter catodon*	1	98	72	0	AWZP01028655.1	-
Cetartiodactyla	*Tursiops truncatus*	1	99	72	0	MRVK01001109.1	-
Cetartiodactyla	*Eschrichtius robustus*	1	79	72	0	NIPP01002163.1	-
Chiroptera	*Eidolon helvum*	0	-	-	-	-	-
Chiroptera	*Eptesicus fuscus*	0	-	-	-	-	-
Chiroptera	*Megaderma lyra*	0	-	-	-	-	-
Chiroptera	*Miniopterus natalensis*	0	-	-	-	-	-
Chiroptera	*Myotis brandtii*	0	-	-	-	-	-
Chiroptera	*Myotis davidii*	0	-	-	-	-	-
Chiroptera	*Myotis lucifugus*	0	-	-	-	-	-
Chiroptera	*Pteronotus parnellii*	0	-	-	-	-	-
Chiroptera	*Pteropus alecto*	0	-	-	-	-	-
Chiroptera	*Pteropus vampyrus*	0	-	-	-	-	-
Chiroptera	*Rhinolophus ferrumequinum*	0	-	-	-	-	-
Chiroptera	*Rousettus aegyptiacus*	0	-	-	-	-	-
Chiroptera	*Hipposideros armiger*	0	-	-	-	-	-
Chiroptera	*Rhinolophus sinicus*	0	-	-	-	-	-
Cingulata	*Dasypus novemcinctus*	1	66	74	3E-169	AAGV03262613.1	-

Dasyuromorphia	*Sarcophilus harrisii*	0	-	-	-	-	-
Dermoptera	*Galeopterus variegatus*	1	96	74	0	JMZW01046728.1	-
Didelphimorphia	*Monodelphis domestica*	0	-	-	-	-	-
Diprotodontia	*NotaMacropus eugenii*	0	-	-	-	-	-
Diprotodontia	*Notamacropus eugenii*	0	-	-	-	-	-
Diprotodontia	*Phascolarctos cinereus*	0	-	-	-	-	-
Eulipotyphla	*Condylura cristata*	1	66	67	8E-41	AJFV01056784.1	-
Eulipotyphla	*Erinaceus europaeus*	0	-	-	-	-	-
Eulipotyphla	*Sorex araneus*	0	-	-	-	-	-
Hyracoidea	*Procavia capensis*	1	70	72	6E-58	ABRQ02098865.1	-
Lagomorpha	*Ochotona princeps*	1	72	71	2E-141	ALIT01067400.1	-
Lagomorpha	*Oryctolagus cuniculus*	1	83	76	6E-121	AAGW02000242.1	-
Macroscelidea	*Elephantulus edwardii*	1	71	70	2E-70	AMGZ01101199.1	-
Monotremata	*Ornithorhynchus anatinus*	0	-	-	-	-	-
Perissodactyla	*Ceratotherium simum simum*	1	99	78	0	AKZM01014557.1	-
Perissodactyla	*Equus asinus*	1	96	77	0	JREZ01000739.1	-
Perissodactyla	*Equus caballus*	1	99	77	0	ATDM01010189.1	-
Perissodactyla	*Equus przewalskii*	1	99	77	0	ATBW01050075.1	-
Pholidota	*Manis pentadactyla*	1	88	67	2E-105	JPTV01068612.1	-
Pholidota	*Manis javanica*	1	93	67	5E-115	JSZB01014022.1	-
Pilosa	*Choloepus hoffmanni*	1	90	72	0	ABVD02306311.1	-
Primates	*Aotus nancymaae*	>50	87	86	0	JYKP02110125.1	-
Primates	*Callithrix jacchus*	19	59	89	0	BBXK01041799.1	1, 6, 7, 18
Primates	*Cebus capucinus*	>75	98	91	0	LVWQ01113849.1	-
Primates	*Cercocebus atys*	1	99	94	0	JZLG01012077.1	-
Primates	*Chlorocebus sabaeus*	1	99	94	0	AQIB01138673.1	-
Primates	*Colobus angolensis*	1	99	92	0	JYKR01111310.1	-
Primates	*Daubentonia madagascariensis*	1	99	80	0	AGTM011566514.1	-
Primates	*Eulemur flavifrons*	1	99	78	0	LGHW01000499.1	-
Primates	*Eulemur macaco*	1	99	77	0	LGHX01000497.1	-

Primates	*Gorilla gorilla*	1	100	99	0	CYUI03000164.1	1
Primates	*Homo sapiens*	1	100	100	0	NIOH01000275.1	1
Primates	*Macaca fascicularis*	1	99	94	0	CAEC01309666.1	1
Primates	*Macaca mulatta*	1	99	94	0	AEHK01163883.1	-
Primates	*Macaca nemestrina*	1	99	94	0	JZLF01038469.1	-
Primates	*Mandrillus leucophaeus*	1	99	93	0	JYKQ01166594.1	-
Primates	*Microcebus murinus*	1	99	78	0	ABDC03003258.1	-
Primates	*Nasalis larvatus*	1	99	92	0	JMHX01319529.1	-
Primates	*Nomascus leucogenys*	1	99	96	0	ADFV01184158.1	-
Primates	*Otolemur garnettii*	1	99	78	0	AAQR03039340.1	-
Primates	*Pan paniscus*	1	100	99	0	AJFE02086751.1	-
Primates	*Pan troglodytes*	1	100	99	0	AACZ04056320.1	-
Primates	*Papio anubis*	1	99	94	0	AHZZ02017098.1	-
Primates	*Pongo abelii*	1	100	98	0	ABGA01181501.1	-
Primates	*Propithecus coquereli*	1	99	77	0	JZKE01044125.1	-
Primates	*Rhinopithecus roxellana*	1	99	92	0	JABR01056658.1	-
Primates	*Saimiri boliviensis boliviensis*	>30	99	87	0	AGCE01139118.1	-
Primates	*Tarsius syrichta*	1	92	83	0	ABRT02453745.1	-
Primates	*Carlito syrichta*	1	92	83	0	ABRT02453745.1	-
Primates	*Rhinopithecus bieti*	1	99	91	0	MCGX01000550.1	-
Proboscidea	*Loxodonta africana*	1	98	70	5E-160	AAGU03056741.1	-
Rodentia	*Apodemus sylvaticus*	1	58	69	2E-28	LIPJ01005224.1	-
Rodentia	*Cavia aperea*	0	-	-	-	-	-
Rodentia	*Cavia porcellus*	1	59	72	5E-77	AAKN02043146.1	-
Rodentia	*Chinchilla lanigera*	1	68	72	3E-80	AGCD01049440.1	-
Rodentia	*Cricetulus griseus*	1	43	72	7E-55	APMK01073751.1	5
Rodentia	*Dipodomys ordii*	1	78	71	8E-80	ABRO02060052.1	-
Rodentia	*Fukomys damarensis*	1	52	76	2E-63	AYUG01070188.1	-
Rodentia	*Heterocephalus glaber*	1	64	73	6E-73	AHKG01082377.1	-
Rodentia	*Ictidomys tridecemlineatus*	1	79	76	6E-129	AGTP01111227.1	-

Rodentia	*Jaculus jaculus*	1	48	74	3E-54	AKZC01198456.1	-
Rodentia	*Marmota marmota*	1	82	76	9E-131	CZRN01000030.1	1
Rodentia	*Mesocricetus auratus*	1	54	70	1E-39	APMT01062006.1	-
Rodentia	*Microtus agrestis*	1	33	68	2E-19	LIQJ01002821.1	-
Rodentia	*Microtus ochrogaster*	1	34	69	5E-44	AHZW01166309.1	-
Rodentia	*Mus musculus*	1	53	72	5E-42	LVXP01002560.1	-
Rodentia	*Mus spretus*	1	54	71	3E-36	LVXV01014015.1	-
Rodentia	*Myodes glareolus*	1	49	71	4E-36	LIPI01003193.1	-
Rodentia	*Nannospalax galili*	1	82	68	3E-49	AXCS01163983.1	-
Rodentia	*Octodon degus*	1	34	72	3E-74	AJSA01153615.1	-
Rodentia	*Peromyscus maniculatus*	2	62	70	5E-52	AYHN01125060.1	-
Rodentia	*Rattus norvegicus*	0	-	-	-	-	-
Rodentia	*Castor canadensis*	1	89	76	1E-128	MTKA01001748.1	-
Rodentia	*Meriones unguiculatus*	1	55	68	1E-34	NHTI01000294.1	-
Rodentia	*Mus caroli*	1	57	72	1E-46	FMAL02019406.1	-
Rodentia	*Mus pahari*	1	39	71	1E-45	FMBV02006109.1	-
Rodentia	*Neotoma lepida*	1	56	73	1E-59	LZPO01007985.1	-
Rodentia	*Phodopus sungorus*	0	-	-	-	-	-
Scandentia	*Tupaia belangeri*	1	60	78	2E-124	AAPY01291272.1	-
Scandentia	*Tupaia chinensis*	1	60	78	3E-124	ALAR01016877.1	-
Sirenia	*Trichechus manatus latirostris*	1	87	75	0	AHIN01112077.1	-
Tubulidentata	*Orycteropus afer*	1	75	74	9E-139	ALYB01072602.1	-

Supplementary Table 3. Examples of contigs from Cebidae species that have the CapA sequence in tandem.

Species	Contig	Accession
Saimiri boliviensis boliviensis	Saimiri boliviensis boliviensis isolate 3227 contig148367, whole genome shotgun sequence	AGCE01148367.1
Cebus capucinus imitator	Cebus capucinus imitator isolate Cc_AM_T3 Cebus_imitator-1.0-38.466, whole genome shotgun sequence	LVWQ01021601.1
Aotus nancymaae	Aotus nancymaae isolate 86115 Contig82194, whole genome shotgun sequence	JYKP02082194
Callithrix jacchus	Callithrix jacchus Contig433.58, whole genome shotgun sequence	ACFV01140206.1

Supplementary Table 4. CapA abundance and number of copies in some Primates.

Species	Genome size (Mb)	Sample size (bp)	CapA abundance %	Copy Number
Aotus nancymaae	3300	202000000	0.27	6016
Chlorocebus aethiops sabaeus	3931.56	200000000	0.00	-
Callithrix jacchus	3354.54	183706306	0.00	-
Cebus capucinus imitator	3716.4	250000000	4.21	104254
Homo sapiens	3423	215442168	0.00	-
Saimiri boliviensis	3227.4	201136039	1.48	31895

References

1. Plohl M, Meštrović N, Mravinac B. (2012) Satellite DNA evolution. *Genome Dyn. Basel*, **7**, 126–152. (doi: 10.1159/000337122)

2. Biscotti MA, Olmo E, Heslop-Harrison JS. (2015) Repetitive DNA in eukaryotic genomes. *Chromosome Res***, 23**, 415-420. (doi: 10.1007/s10577-015-9499-z)

3. Malfoy B, Rousseau N, Vogt N, Viegas-Pequignot E. (1986) Nucleotide sequence of an heterochromatic segment recognized by the antibodies to Z-DNA in fixed metaphase chromosomes. *Nucleic Acids Res*. **14**, 3197-214. (doi: 10.1093/nar/14.8.3197)

4. Fanning TG, Seuánez, HN, Forman, L. (1993) Satellite DNA sequences in the New World primate *Cebus apella* (Platyrrhini, Primates). *Chromosoma*. **102**, 306-311.

5. Schneider H, Sampaio I. (2015) The systematics and evolution of New World primates - A review. *Mol Phylogenet Evol*. **82**, 348-57. (doi: 10.1016/j.ympev.2013.10.017)

6. Cheng Z, Ventura M, She X, Khaitovich P, Graves T, Osoegawa K, Church D, DeJong P, Wilson RK, Paabo S, Rocchi M, Eichler EE. (2005) A genome-wide comparison of recent chimpanzee and human segmental duplications. *Nature*. **437**. 88-93. (doi:10.1038/nature04000)

7. Pickeral OK, Makalowski W, Boguski MS, Boeke JD. (2000) Frequent human genomic DNA transduction driven by LINE-1 retrotransposition. *Genome Res*. **10**,411-415. (doi: 10.1101/gr.10.4.411)

8. Ostertag EM, Kazazian HH. (2001) Biology of mammalian L1 retrotransposons. *Annu Rev Genet*. **35**, 501-38. (doi: 10.1146/annurev.genet.35.102401.091032)

9. Eichler EE, Budarf ML, Rocchi M, Deaven LL, Doggett NA, Baldini A, Nelson DL, Mohrenweiser WH. (1997) Interchromosomal duplications of the adrenoleukodystrophy locus: A phenomenon of pericentromeric plasticity. *Hum Mol Genet*. **6**, 991-1002. (doi: 10.1093/hmg/6.7.991)

10. Meštrović N, Mravinac B, Pavlek M, Vojvoda-Zeljko T, Šatović E, Plohl M. (2015) Structural and functional liaisons between transposable elements and satellite DNAs. *Chromosome Res*. **23**, 583-596. (doi: 10.1159/000080800)

11. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. (1990) Basic Local Alignment Search Tool. *J Mol Biol*. **215**, 403-10. (doi: 10.1016/S0022-2836(05)80360-2)

12. Smit AFA, Hubley R, Green P. (2015) *RepeatMasker Open-4.0*. 2013-2015. Institute for Systems Biology. http://repeatmasker. org.

13. Foley NM, Springer MS, Teeling EC. (2016) Mammal madness: is the mammal tree of life not yet resolved? *Phil Trans R Soc B*. **371**, 11. (doi: 10.1098/rstb.2015.0140)

5. CONCLUSÕES

- Este é o primeiro estudo que caracterizou os DNAs satélites do genoma de *S*. *boliviensis* e os mapeou nos cromossomos de *S. sciureus, S. boliviensis, S. vanzolinii*, e *S. ustus*;
- Os dois principais DNAs satélites encontrados em *Saimiri* foram o DNA satélite alfa e o CapA, que correspondem, respectivamente, a 1% e 2,2% do genoma de *S*. *boliviensis*;
- O tamanho do monômero do DNA satélite alfa em *Saimiri*, com 340 pb, é igual ao descrito em outros Platyrrhini e sua localização centromérica foi confirmada.
- O DNA satélite CapA possui monômeros de 1.500 pb, está associado com a heterocromatina constitutiva e sua distribuição nos cromossomos variou entre as espécies de *Saimiri*, sugerindo que possa ser um possível marcador taxonômico para o grupo. CapA também pode estar relacionado com o polimorfismo do cromossomo 14.
- O CapA está presente em alguns gêneros das três famílias de Platyrrhini.
- A sequência ancestral do DNA satélite CapA está presente na maioria dos eutérios e provavelmente também no íntron do gene NOS1AP, como em humanos. Para o nosso conhecimento, esse é o primeiro relato de uma sequência intrônica de cópia única como origem de um DNA satellite. Uma hipótese para a origem desse DNA satélite seria a duplicação/transposição da sequência ancestral para outra região do genoma e sua amplificação por crossing-over desigual.

6. REFERÊNCIAS

ALFARO, J. W. L.; BOUBLI, J. P.; PAIM, F. P.; RIBAS, C. C.; da SILVA, M. N. F.; MESSIAS, M. R.; ROHE, F.; MERCES, M. P.; SILVA JUNIOR, J. S.; SILVA, C. R.; PINHO, G. M.; KOSHKARIAN, G.; NGUYEN, M. T. T.; HARADA, M. L.; RABELO, R. M.; QUEIROZ, H. L.; ALFARO, M. E.; FARIAS, I. P. Biogeography of squirrel monkeys (genus *Saimiri*): South-central Amazon origin and rapid pan-Amazonian diversification of a lowland primate. **Molecular Phylogenetics and Evolution,** v. 82, p. 436-454, 2015.

ALKAN, C.; VENTURA, M.; ARCHIDIACONO, N.; ROCCHI, M.; SAHINALP, S. C.; EICHLER, E. E. Organization and evolution of primate centromeric DNA from whole-genome shotgun sequence data. **Plos Computational Biology,** v. 3, n. 9, p. 1807-1818, 2007.

ALVES, G.; SEUÁNEZ, H. N.; FANNING, T. A clade of new world primates with distinctive alphoid satellite DNAs. **Molecular Phylogenetics and Evolution,** v. 9, n. 2, p. 220-224, 1998.

ASSIS, M. F. L.; BARROS, R. Karyotype pattern of *Saimiri ustus*. **International Journal of Primatology,** v. 8, n. 5, p. 552, 1987.

BAICHAROEN, S.; MIYABE-NISHIWAKI, T.; ARSAITHAMKUL, V.; HIRAI, Y.; DUANGSA-ARD, K.; SIRIAROONRAT, B.; DOMAE, H.; SRIKULNATH, K.; KOGA, A.; HIRAI, H. Locational diversity of alpha satellite DNA and intergeneric hybridization aspects in the *Nomascus* and *Hylobates* genera of small apes. **PloS One**, v. 9, n. 10, p. e109151, 2014.

BENDER, M. A.; METTLER, L. E. Chromosome studies of primates. **Science,** v. 128, n. 3317, p. 186-190, 1958.

BISCOTTI, M. A.; OLMO, E.; HESLOP-HARRISON, J. S. Repetitive DNA in eukaryotic genomes. **Chromosome Research,** v. 23, n. 3, p. 415-420, 2015.

BOINSKI, S.; CROPP, S. J. Disparate data sets resolve squirrel monkey (*Saimiri*) taxonomy: Implications for behavioral ecology and biomedical usage. **International Journal of Primatology,** v. 20, n. 2, p. 237-256, 1999.

CARNEIRO, J.; RODRIGUES, L. F. D.; SCHNEIDER, H.; SAMPAIO, I. Molecular data highlight hybridization in squirrel monkeys (*Saimiri*, Cebidae). **Genetics and Molecular Biology,** v. 39, n. 4, p. 539-546, 2016.

CELLAMARE, A.; CATACCHIO, C. R.; ALKAN, C.; GIANNUZZI, G.; ANTONACCI, F.; CARDONE, M. F.; DELLA VALLE, G.; MALIG, M.; ROCCHI, M.; EICHLER, E. E.; VENTURA, M. New insights into centromere organization and evolution from the white-cheeked gibbon and marmoset. **Molecular Biology and Evolution,** v. 26, n. 8, p. 1889-1900, 2009.

CHIATANTE, G.; CAPOZZI, O.; SVARTMAN, M.; PERELMAN, P.; CENTRONE, L.; ROMANENKO, S. S.; ISHIDA, T.; VALERI, M.; ROELKE-PARKER, E. M.; STANYON, R. Centromere repositioning explains fundamental number variability in the New World monkey genus *Saimiri*. **Chromosoma**, v. 126, p. 519-529, 2017.

COHEN, S.; SEGAL, D. Extrachromosomal circular DNA in eukaryotes: possible involvement in the plasticity of tandem repeats. **Cytogenetic and Genome Research,** v. 124, n. 3-4, p. 327-338, 2009.

COSTELLO, R. K.; DICKINSON, C.; ROSENBERGER, A. L. B., S.; SZALAY, FREDERICK S. Squirrel monkey (genus *Saimiri*) taxonomy. In: MARTIN, W. H. K. B. **Species, species concepts and primate evolution**. Nova York: Springer US, 1993. p.177-210.

CROPP, S.; BOINSKI, S. The Central American squirrel monkey (*Saimiri oerstedii*): Introduced hybrid or endemic species? **Molecular Phylogenetics and Evolution,** v. 16, n. 3, p. 350-365, 2000.

DOVER, G. Molecular Drive - A cohesive mode of species evolution. **Nature,** v. 299, n. 5879, p. 111-117, 1982.

DOVER, G. A. Molecular drive in multigene families - How biological novelties arise, spread and are assimilated. **Trends in Genetics,** v. 2, n. 6, p. 159-165, 1986.

FANNING, T. G.; SEUÁNEZ, H. N.; FORMAN, L. Satellite DNA-sequences in the New-World primate *Cebus apella* (Platyrrhini, Primates). **Chromosoma,** v. 102, n. 5, p. 306-311, 1993.

GARRIDO-RAMOS, M. A.; DE LA HERRAN, R.; JAMILENA, M.; LOZANO, R.; REJÓN, C. R.; REJÓN, M. R. Evolution of centromeric satellite DNA and its use in phylogenetic studies of the Sparidae family (Pisces, Perciformes). **Molecular Phylogenetics and Evolution**, v. 12, n. 2, p. 200-204, 1999.

GOECKS, J.; NEKRUTENKO, A.; TAYLOR, J.; GALAXY, T. Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. **Genome Biology,** v. 11, n. 8, p. 13, 2010.

GROVES, C. P. Order Primates. In: WILSON, D. E.; REEDER, D. M. **Mammal species of the world**. 3th ed. Baltimore: The Johns Hopkins University Press, 2005. p. 111–184.

HERSHKOVITZ, P. Taxonomy of squirrel monkeys genus *Saimiri* (Cebidae, Platyrrhini): a preliminary report with description of a hitherto unnamed form. **American Journal of Primatology**, v. 7, n. 2, p. 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. **American Journal of Primatology**, v. 12, n. 1, p. 1-53, 1987.

JONES, T.; THORINGTON, R.; HU, M.; ADAMS, E; COOPER, R. Karyotypes of squirrel monkeys (*Saimiri sciureus*) from different geographic regions. **American Journal of Physical Anthropology,** v. 38, n. 2, p. 269-277, 1973.

JONES, T. C.; MA, N. S. F. Cytogenetics of the squirrel monkey (*Saimiri sciureus*). In: GOODWIN W. J. E AUGUSTINE J. **Primate Research,** v.6. Nova York: Springer US, 1975. p.13-21.

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. **Molecular Genetics and Genomics,** v. 292, n. 2, p. 453-464, 2017.

KŘIVÁNKOVÁ, A.; KOPECKÝ, D.; STOČES, Š.; DOLEŽEL, J.; HŘIBOVÁ, E. Repetitive DNA: A versatile tool for karyotyping in *Festuca pratensis* Huds. **Cytogenetic and Genome Research,** v. 151, n. 2, p. 96-105, 2017.

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*. **Chromosome Res**, v. 16, n.2, p. 307-324, 2008.

LAU, Y.-F.; ARRIGHI, F. E. Studies of the squirrel monkey, *Saimiri sciureus*, genome. **Cytogenetic and Genome Research,** v. 17, n. 1, p. 51-60, 1976.

LAVERGNE, A.; RUIZ-GARCÍA, M.; CATZEFLIS, F.; LACOTE, S.; CONTAMIN, H.; MERCEREAU-PUIJALON, O.; LACOSTE, V.; THOISY, B. Phylogeny and phylogeography of squirrel monkeys (genus *Saimiri*) based on cytochrome b genetic analysis. **American Journal of Primatology,** v. 72, n. 3, p. 242-253, 2010.

MA, N. S. F.; JONES, T. C. Added heterochromatin segments in chromosomes of squirrel-monkeys (*Saimiri sciureus*). **Folia Primatologica,** v. 24, n. 4, p. 282-292, 1975.

MARTIN, R. D. Primates. **Current Biology,** v. 22, n. 18, p. 785-790, 2012.

MERCÊS, M. P.; ALFARO, J. W. L.; FERREIRA, W. A.; HARADA, M. L.; JÚNIOR, J. S. S. Morphology and mitochondrial phylogenetics reveal that the Amazon River separates two eastern squirrel monkey species: *Saimiri sciureus* and *S. collinsi*. **Molecular Phylogenetics and Evolution**, v. 82, p. 426-435, 2015.

MOORE, C.; HARRIS, C.; ABEE, C. Distribution of chromosomal polymorphisms in three subspecies of squirrel monkeys (genus *Saimiri*). **Cytogenetic and Genome Research,** v. 53, n. 2-3, p. 118-122, 1990.

NOVAK, P.; NEUMANN, P.; MACAS, J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. **BMC Bioinformatics,** v. 11, p. 12, 2010.

NOVAK, 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,** v. 29, n. 6, p. 792-793, 2013.

PAGAN, H. J. T.; MACAS, J.; NOVAK, P.; MCCULLOCH, E. S.; STEVENS, R. D.; RAY, D. A. Survey sequencing reveals elevated DNA transposon activity, novel elements, and variation in repetitive landscapes among vesper bats. **Genome Biology and Evolution,** v. 4, n. 4, p. 575-585, 2012.

PERELMAN, P.; JOHNSON, W. E.; ROOS, C.; SEUÁNEZ, H. N.; HORVATH, J. E.; MOREIRA, M. A. M.; KESSING, B.; PONTIUS, J.; ROELKE, M.; RUMPLER, Y.; SCHNEIDER, M. P. C.; SILVA, A.; O'BRIEN, S. J.; PECON-SLATTERY, J. A molecular phylogeny of living Pprimates. **PLoS Genetics,** v. 7, n. 3, p. 17, 2011.

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**, v. 45, n. 1, p. 63-70, 2002.

PLOHL, M.; MEŠTROVIĆ, N.; MRAVINAC, B. Satellite DNA evolution. **Repetitive DNA,** v. 7, p. 126-152, 2012.

RICHARD, G. F.; KERREST, A.; DUJON, B. Comparative genomics and molecular dynamics of DNA repeats in eukaryotes. **Microbiology and Molecular Biology Reviews,** v. 72, n. 4, p. 686, 2008.

ROCCHI, M.; ARCHIDIACONO, N.; SCHEMPP, W.; CAPOZZI, O.; STANYON, R. Centromere repositioning in mammals. **Heredity,** v. 108, n. 1, p. 59-67, 2012.

RYLANDS, A. B.; MITTERMEIER, R. A. The diversity of the new world primates (Platyrrhini): an annotated taxonomy. In: GARBER, P.A., ESTRADA, A., BICCA-MARQUES, J.C., HEYMANN, E.W., STRIER, K.B. **South American primates:** Comparative perspectives in the study of behavior, ecology, and conservation. New York: Springer, 2009. p. 23–54.

RYLANDS, A. B.; MITTERMEIER, R. A.; BEZERRA, B. M.; PAIM, F. P.; QUEIROZ, H. L. Species accounts of Cebidae. In: Mittermeier, R.A., Rylands, A.B., Wilson, D.E. **Handbook of the mammals of the world**, vol. 3. Primates. Barcelona: Lynx Edicions, 2013, p. 390–413.

SCHNEIDER, H.; SAMPAIO, I. The systematics and evolution of New World primates - A review. **Molecular Phylogenetics and Evolution,** v. 82, p. 348-357, 2015.

SEUÁNEZ, H. N.; BONVICINO, C. R.; MOREIRA, M. A. M. The primates of the neotropics: genomes and chromosomes. **Cytogenetic and Genome Research,** v. 108, n. 1-3, p. 38-46, 2005.

SHAH, A. B.; SCHIELZETH, H.; ALBERSMEIER, A.; KALINOWSKI, J.; HOFFMAN, J. I. High-throughput sequencing and graph-based cluster analysis facilitate microsatellite development from a highly complex genome. **Ecology and Evolution,** v. 6, n. 16, p. 5718-5727, 2016.

SILVA, B. T. F.; SAMPAIO, M. I. C.; SCHNEIDER, H.; SCHNEIDER, M. P. C.; MONTOYA, E.; ENCARNACION, F.; SALZANO, F. M. Natural hybridization between *Saimiri* taxa in the peruvian amazonia. **Primates,** v. 33, n. 1, p. 107-113, 1992.

SMIT, A.; HUBLEY, R.; GREEN, P. ***RepeatMasker Open-4.0*** 2013-2015. <http://www.repeatmasker.org>

STANYON, R.; CONSIGLIERE, S.; MULLER, S.; MORESCALCHI, A.; NEUSSER, M.; WIENBERG, J. Fluorescence in situ hybridization (FISH) maps chromosomal homologies between the dusky titi and squirrel monkey. **American Journal of Primatology,** v. 50, n. 2, p. 95-107, 2000.

THORINGTON JR., R. W. The Taxonomy and distribution of squirrel monkeys (*Saimiri*). In: Rosenblum, L. A. e Coe C. L. **Handbook of squirrel monkey research**. Nova York: Springer US, 1985. p.1-33.

The IUCN Red List of Threatened Species. Version 2017-1. Disponível em: <www.iucnredlist.org>. Acesso em: 08 de Julho de 2017.

YONENAGA-YASSUDA, Y.; CHU, T. Chromosome banding patterns of *Saimiri vanzolinii* Ayres, 1985 (Primates, Cebidae). **Papéis Avulsos de Zoologia,** v. 36, p. 165-168, 1985.