

Gabrielle Azevedo Rizzato

Evolução molecular do *enhancer* HACNS1 em primatas

Molecular evolution of HACNS1 gene enhancer in primates

São Paulo

2020

Gabrielle Azevedo Rizzato

Evolução molecular do *enhancer* HACNS1 em primatas

Molecular evolution of HACNS1 gene enhancer in primates

Dissertação apresentada ao Instituto de Biociências da Universidade de São Paulo para a obtenção do Título de Mestre em Ciências, na Área de Biologia/Genética. Dissertação corrigida, o original encontra-se disponível no Instituto de Biociências da Universidade de São Paulo.

Orientadora: Prof^a Dr^a Tábita Hünemeier

São Paulo

2020

Ficha Catalográfica

Rizzato, Gabrielle Azevedo
Evolução molecular do *enhancer*
HACNS1 em primatas
78 páginas

Dissertação (Mestrado) - Instituto de
Biotecnologia da Universidade de São Paulo.
Departamento de Genética e Biologia
Evolutiva.

1. HACNS1 2. *Enhancer* 3. Primatas
4. Fatores de transcrição I. Universidade
de São Paulo. Instituto de Biotecnologia.
Departamento de Genética e Biologia
Evolutiva.

Comissão Julgadora:

Prof(a). Dr(a).

Prof(a). Dr(a).

Prof(a). Dr(a). Tábata Hünemeier

À minha mãe Neli,
minha maior inspiração,
por sempre acreditar em
mim e me apoiar em
todos os momentos.

*“O mais bonito da Ciência é que ela nunca
tem fim”*

Francisco Mauro Salzano

Agradecimentos

À minha orientadora Prof^a Dr^a Tábita Hünemeier pela orientação e paciência durante esses anos.

À minha coorientadora Prof^a Dr^a Vanessa Paixão-Cortês e ao Dr. Renan Barbosa Lemes pelas orientações no decorrer do trabalho.

Aos demais colegas de laboratório Cainã, Marcos, Tiago e Maíra pelo companheirismo e apoio.

À Prof^a Dr^a Sônia Andrade pelo espaço proporcionado para a realização das análises laboratoriais.

A CAPES pela bolsa de mestrado, à FAPESP e o CNPq pela ajuda financeira.

À minha mãe Neli que me apoiou em cada momento e decisão, que me sempre me incentivou, sempre acreditou no meu potencial e nunca desistiu de mim.

À Yu, a Cecili e a Paola as melhores amigas que eu poderia ter me acompanhando nessa fase do início ao fim, me apoiando sempre, me ouvindo, me aconselhando e me ajudando de diversas formas. Amo vocês!

À Isabella, por ser a melhor roomie dessa SP! Obrigada pelas conversas até tarde da noite, pelas viagens juntas, pela casa compartilhada, e pela Vamp.

Ao Guilherme, pelo apoio, pelo carinho e por ter me ajudado a atravessar o pior momento de todos.

A todos os meus amigos que participaram dessa jornada, não irei citar nomes pra não esquecer ninguém, mas se você esteve presente na minha vida nesses últimos três anos, saiba que você está aqui e no meu coração.

À minha psicóloga Gabriela e à minha psiquiatra Marina, pela ajuda profissional pra superar as diversas adversidades que se apresentaram pelo caminho.

A todos os meus alunos, por terem me lembrado do quanto eu amo ensinar.

Ao Thiago por ter aparecido na minha vida nessa reta final e ter deixado tudo mais leve e feliz com nosso amor.

Por último eu agradeço a meus Orixás por terem me guiado e não me deixado desistir dos meus sonhos.

Índice Geral

1. INTRODUÇÃO	11
1.1 Contextualização	11
1.2. Elementos regulatórios	12
1.3. Estratégias para descoberta de motivos	15
1.4. Ordem Primates.....	17
2. OBJETIVOS GERAIS	21
2.1. Objetivos Específicos	21
3. MATERIAIS E MÉTODOS.....	22
3.1. Amostras <i>in silico</i>	22
3.2. Amostras de DNA.....	23
3.3. Análises Laboratoriais	25
3.4. Análise de Seleção	26
3.5. Análise de TFBS	28
4. RESULTADOS	30
4.1. Variabilidade Molecular	30
4.2. Análise de Seleção	31
4.3. Análise de TFBS	32
5. DISCUSSÃO	35
5.1. MYRF.....	37
5.2. CAMTA1	40
5.3. PTX1.....	44
6. CONCLUSÃO	48
7. RESUMO	50
8. <i>ABSTRACT</i>	51
7. REFERÊNCIAS BIBLIOGRÁFICAS.....	52
8. ANEXOS.....	60

Índice de Figuras

Figura 1. Alinhamento do HACNS1 com sequências ortólogas de outros genomas de vertebrados, com enfoque na região de 81-pb que contém as 13 substituições específicas de humanos, destacadas em vermelho. Pontos representam as posições nos genomas não humanos idênticas ao de humanos.....	14
Figura 2. Padrões de expressão obtidos com o gene repórter do <i>enhancer</i> HACNS1 em ortólogos de chimpanzé e rhesus em embriões de ratos. Setas indicam a germinação dos membros anteriores onde o enhancer está ativo ou inativo expressando o gene repórter.....	15
Figura 3. Exemplo da apresentação de um logo de um motivo hipotético	17
Figura 4. Esquematização da Ordem Primates	19

Índice de Tabelas

Tabela 1. Espécies de primatas com genoma completo disponível.	22
Tabela 2. Amostras de primatas no novo mundo empregadas para amplificação e sequenciamento da região genômica HACNS1.	24
Tabela 3. Espécimes utilizados para sequenciamento do enhancer HACNS1 e do exon 11 do gene AGAP1-201, e seus respectivos códigos no banco de amostras do Laboratório de Evolução Humana e Molecular da Universidade Federal do Rio Grande do Sul.	26
Tabela 4. Resultados da análise de seleção dos modelos neutros e seus respectivos modelos Alternate 1 e Alternate 2, com e sem especificação de ramos.	32
Tabela 5. Principais TFBS identificados como ganhos ou perdidos no HACNS1. “Do” e “ao” se referem ao número do nucleotídeo em que os TFBS começam e terminam, respectivamente, nas sequências do Anexo II.	34

Índice de Abreviaturas

BEB: *Bayes Empirical Bayes*;

CAMTA: *Calmodulin-binding Transcription Activator*; ativador da transcrição com ligação à calmodulina.

CREs: *Cis-Regulatory Elements*; elementos *cis*-regulatórios.

DNA: *Deoxyribonucleic Acid*; ácido desoxirribonucleico.

dNTPs: *Deoxyribonucleotide Triphosphate*; desoxirribonucleotídeos trifosfatados

HARs: *Human Accelerated Regions*; regiões humanas aceleradas.

LRT: *Likelihood Ratio Test*; teste de razão de verossimilhança.

MHCII: *Major Histocompatibility*; complexo principal de histocompatibilidade.

mRNA: RNA (ácido ribonucleico) mensageiro.

MTFs: *Membrane Bound Transcription Factor*; fatores de transcrição ligados à membrana.

MYRF: *Myelin regulatory factor*; fator de transcrição de regulação da mielina.

NTM: *N-terminal-most domain*; domínio mais N-terminal.

OLs: Oligodendrócitos

OPCs: Células Progenitoras de Oligodendrócitos

pb: pares de base

PCR: *Polymerase Chain Reaction*; reação em cadeia da polimerase.

PTX1: *Pituitary Homeobox 1*; fator de transcrição *homeobox* da pituitária.

PWMs: *Position Weight Matrix*; matriz de peso posicional.

SNC: Sistema Nervoso Central.

SNPs: *Single Nucleotide Polymorphism*; polimorfismo de nucleotídeo único.

TF: *Transcriptional Factors*; fatores de transcrição.

TFBS: *Transcriptional Factors Binding Sites*; sítios de ligação de fatores de transcrição.

1. INTRODUÇÃO

1.1 Contextualização

Por milhões de anos, os processos evolutivos vêm moldando a variabilidade biológica existente. Dentre todas as mudanças evolutivas, talvez nenhuma tenha sido mais estudada e discutida do que as mudanças em nível morfológico. Apesar de séculos de registros e de classificação dessas diferenças, nós ainda estamos longe de responder algumas questões básicas sobre suas origens, tais como: como novas morfologias surgem? Quais são suas bases moleculares? (POPADIC *et al.*, 1998). Apesar do sequenciamento completo ou quase completo de diversos genomas de seres procariotos e eucariotos, incluindo os humanos, a função de muitos genes permanece desconhecida (BALL, 2013).

Ao estudar a evolução biológica, é necessário combinar a informação de diversos campos do conhecimento, como biologia do desenvolvimento e genética molecular, dado que é preciso haver alguma mudança em um ou mais genes para que haja mudanças fenotípicas (FUTUYMA, 1998).

Até recentemente, a teoria evolucionista explicava grandes diferenças fenotípicas como o resultado de numerosas substituições nucleotídicas em sequências codificadoras de proteína, cada qual com um efeito relativamente pequeno (FUTUYMA, 1998), sendo geralmente assumido que traços espécie-específicos resultavam de proteínas espécie-específicas (WITTKOPP & KALAY, 2011). Embora mudanças em funções de proteínas sejam responsáveis por divergência fenotípica em alguns casos (WITTKOPP & KALAY, 2011), estudos conduzidos em genes que regulam processos de desenvolvimento sugerem que poucas mutações nesses genes podem ter um efeito maior e mais instantâneo na morfologia (POPADIC *et al.*, 1998).

Muitas proteínas homólogas têm funções altamente conservadas entre espécies, mas apresentam diferenças em seus padrões de expressões que contribuem para divergência fenotípica (WITTKOPP & KALAY, 2011). Como exemplo temos os primeiros estudos com humanos e chimpanzés, nos quais pesquisadores concluíram que a pequena divergência molecular observada não poderia explicar as diferenças anatômicas e comportamentais entre as duas espécies. King & Wilson (1975) então propuseram que as mudanças evolutivas eram mais frequentemente causadas por mudanças nos mecanismos de controle de expressão gênica, do que mudanças nas sequências das proteínas.

De fato, atualmente sabemos que a grande maioria de todas as mudanças genômicas que ocorreram desde o ancestral humano-chimpanzé está em regiões não codificadoras, consistente com a hipótese de King & Wilson de que mudanças regulatórias influenciaram as diferenças entre nossas espécies. Em retrospectiva, a importância da regulação de genes na evolução humana é lógica já que existem muito mais bases de DNA nas regiões reguladoras do que nos genes codificadores de proteínas, tornando-as um alvo maior para a evolução (FRANCHINI & POLLARD, 2017).

A regulação da transcrição é a forma mais comum de controle da expressão gênica, com os fatores de transcrição (TF, *transcriptional factors*) tendo um papel crucial na regulação de todos os aspectos do ciclo de vida de um organismo (BOUCHÉ *et al.*, 2002; WITTKOPP & KALAY, 2011). Os fatores de transcrição são proteínas que possuem domínios que se ligam à regiões específicas do DNA, e ao responderem à diversos sinais, permitem uma regulação específica de genes durante o desenvolvimento (HAKIM & USMANI, 2014).

De maneira geral, genes que codificam fatores de transcrição envolvidos nos estágios iniciais do desenvolvimento são altamente conservados (VOGEL & CHOTHIA, 2006; NEI, 2007). Entretanto, mutações que levam a alterações na expressão ou ainda na estrutura de uma proteína regulatória podem afetar de maneira indireta, porém determinante, uma rede de outros genes (sujeitos a regulação pela proteína alterada) que não apresentam nenhuma modificação em sua estrutura molecular original (CARROLL, 2003). Contudo, compreender de que forma mudanças na regulação gênica podem alterar o fenótipo é consideravelmente mais complexo do que a investigação de mudanças gênicas estruturais (WRAY, 2006), e requer uma metodologia clara para a identificação de genes candidatos, bem como o completo entendimento das redes e rotas responsáveis pelo desenvolvimento de características complexas (HONEYCUTT, 2008).

1.2. Elementos regulatórios

Dois dos mais importantes elementos funcionais em qualquer genoma são os fatores de transcrição e os sítios no DNA onde os quais eles se ligam, os sítios de ligação de fatores de transcrição (TFBS, *transcriptional factors binding sites*). Essas interações entre proteínas e DNA controlam muitos processos importantes, como passos críticos no desenvolvimento e respostas a estresses ambientais, e defeitos nos mesmos podem contribuir para a progressão de diversas doenças (BULYK, 2003). Não surpreendentemente, uma alta proporção do

genoma eucarioto codifica fatores de transcrição, com uma estimativa de aproximadamente 8% do genoma humano (LAMBERT *et al.*, 2018).

Mutações afetando sequências *cis*-regulatórias – em oposição às *trans*-regulatórias que codificam os fatores de transcrição – são consideradas a causa mais prevalente de divergência fenotípica (WITTKOPP & KALAY, 2011). Sequências *cis*-regulatórias, também chamadas de elementos *cis*-regulatórios (CREs, *cis-regulatory elements*), são compostos de DNA, (tipicamente DNA não codificante), contendo sítios de ligação para fatores de transcrição e/ou outras moléculas regulatórias que são necessárias para ativar ou manter a transcrição. Os promotores e os *enhancers* são os tipos de CREs mais compreendidos. A maioria dos genes eucarióticos contém um único promotor localizado perto ao sítio de início da transcrição, embora alguns genes contenham promotores alternativos que, sob condições específicas, ativam a transcrição em diferentes posições do genoma (WITTKOPP & KALAY, 2011).

Promotores são necessários para a transcrição em eucariotos, mas sozinhos produzem apenas níveis basais de mRNA. Por se ligarem a um conjunto básico de reguladores transcricionais altamente utilizados e amplamente conservados, não parece que eles sejam a força primária de divergência *cis*-regulatória. Comparado aos promotores, *enhancers* tendem a serem mais variáveis entre espécies; sendo o tipo de CRE que é mais frequentemente considerado o responsável pela divergência *cis*-regulatória (WITTKOPP & KALAY, 2011).

Enhancers são classicamente definidos como CREs que quando ligados a fatores de transcrição, aumentam a transcrição de um gene associado. Eles geralmente funcionam independentemente da orientação, também podendo se encontrar a distâncias variáveis de seu gene (ou genes) alvo, sendo possível encontrá-los *upstream* (5') ou *downstream* (3') dos genes, ou até mesmo dentro de íntrons (PENNACCHIO *et al.*, 2013). A localização genômica de *enhancers* funcionalmente homólogos é frequentemente, mas não sempre, conservada entre espécies (WITTKOPP & KALAY, 2011). Além disso, eles não agem necessariamente no promotor mais próximo, mas podem pular seus genes vizinhos para regular genes que estejam mais adiante no cromossomo. Em alguns casos foram encontrados *enhancers* que regulam múltiplos genes. A atividade dos *enhancers* pode ser restrita a um tipo particular de tecido ou célula, um momento da vida, ou a condições específicas de caráter fisiológico, patológico ou ambiental (PENNACCHIO *et al.*, 2013).

Atualmente podemos assumir que há no genoma humano pelo menos de cinco a dez vezes mais regiões de DNA com função regulatória do que codificadores de proteínas (PENNACCHIO *et al.*, 2013), e estudos com sequências não-codificadoras conservadas de humanos sugerem que uma grande fração desses genes regulatórios pode atuar como

enhancers, resultando em um grande alvo em potencial para a evolução (PENNACCHIO *et al.*, 2006).

Acredita-se que mudanças evolutivas na sequência de DNA de *enhancers* tenham desencadeado a evolução de fenótipos específicos, de insetos a humanos (PENNACCHIO *et al.*, 2013), incluindo mudanças no desenvolvimento humano que o diferenciam dos chimpanzés e outros grandes primatas. Por exemplo, 13 substituições nucleotídicas em *enhancers* em espécies divergentes são responsáveis por diferenças em atividade *cis*-regulatória entre espécies de *Drosophila* que estão associadas com a perda de tricomas em larvas (FRANKEL *et al.*, 2011) e pigmentação divergente (JEONG *et al.*, 2008; WILLIAMS *et al.*, 2008).

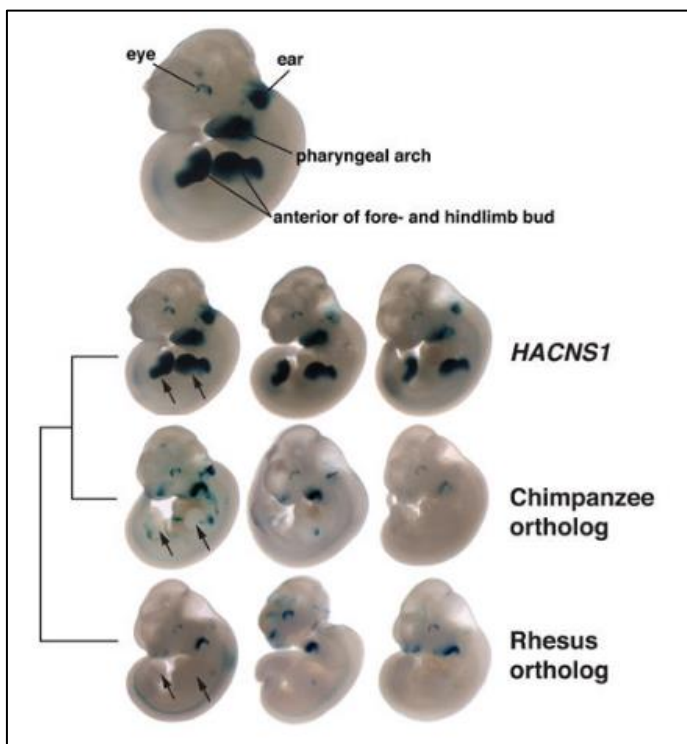
Similarmente, estudos em uma região genômica de 546 pb que atua como um *enhancer*, denominada HACNS1 (*Human Accelerated Non-Coding Sequence 1*), mostraram que ela é altamente conservada no genoma de todos os vertebrados terrestres, mas que acumulou 16 mudanças específicas em humanos nos ~6 milhões de anos desde a divergência de humanos e chimpanzés, com 13 dessas 16 mutações dentro de um segmento de 81 pb dessa região (Fig. 1). O HACNS1 tem um papel específico em humanos, possivelmente relacionado ao desenvolvimento da destreza das mãos e bipedalismo (Fig. 2) (PRABHAKAR *et al.*, 2008).

Figura 1. Alinhamento do HACNS1 com sequências ortólogas de outros genomas de vertebrados, com enfoque na região de 81-pb que contém as 13 substituições específicas de humanos, destacadas em vermelho. Pontos representam as posições nos genomas não humanos idênticas ao de humanos.

Human	GCAGCCTTGGGTT	CCGCAAAATAGGGCA	CCCACAGT	AACCGTGTGGCGCG	ACCCCGCCGTGC	GCAATCGGGGCTTTATAC
Chimpanzee	A . . . G T T A A T . A . TA AT G					
Rhesus	A . . . G T T A A T . A . TA AT G					
Mouse GT C T T A CA T . A . TA AT C . G					
Rat GT T T A CA T . A . TA AT C . G					
Dog GT T T . G A A A . TA T AT C . G					
Chicken GG . G TA T A A T . A . TA T AT A . T G					

Fonte: PRABHAKAR *et al.*, 2008

Figura 2. Padrões de expressão obtidos com o gene repórter do *enhancer* HACNS1 em ortólogos de chimpanzé e rhesus em embriões de ratos. Setas indicam a germinação dos membros anteriores onde o enhancer está ativo ou inativo expressando o gene repórter.



Fonte: PRABHAKAR *et al.*, 2008

O HACNS1 de humanos se encontra localizado no cromossomo 2, dentro do íntron 9 do gene AGAP1-202, e pertence a uma série de sequências curtas de DNA denominadas regiões humanas aceleradas (HARs, *Human Accelerated Regions*), que são altamente conservadas em vertebrados, mas que mostram uma evolução acelerada apenas em humanos ao adquirirem um número significativamente maior de substituições nucleotídicas do que esperado na linhagem humana desde a divergência dos chimpanzés (SUMIYAMA & SAITOU, 2011; FRANCHINI & POLLARD, 2017). As HARs estão entre as sequências mais interessantes para estudo, pois elas ajudam a elucidar o aparecimento do *Homo sapiens* como uma espécie ao analisar questões diversas, como a origem da fala (ENARD *et al.*, 2002), o desenvolvimento do cérebro humano (LEVCHENKO *et al.*, 2018), e no caso do HACNS1, como já dito, destreza manual (PRABHAKAR *et al.*, 2008, HÜNEMEIER *et al.*, 2010).

1.3. Estratégias para descoberta de motivos

Elucidar os mecanismos que controlam a transcrição de genes é uma das maiores

metas da biologia molecular. Uma das abordagens atuais é determinar se as sequências regulatórias possuem uma afinidade maior que a esperada para com uma proteína regulatória cujos motivos de ligação de DNA são conhecidos. Diz-se que tal motivo está “enriquecido” no conjunto de sequências. As proteínas reguladoras cujos motivos são enriquecidos no conjunto de sequências reguladoras são candidatos a reguladores transcricionais (MCLEAY & BAILEY, 2010).

Um motivo é um padrão, ou um sinal, que ocorre em sequências relacionadas de DNAs ou proteínas. Por exemplo, alguns ou todos os promotores de genes co-expressos e/ou ortólogos podem conter sítios de ligação (o “sinal”) para o mesmo fator de transcrição (BAILEY *et al.*, 2006).

Há uma grande variedade de TFs em uma célula. Atualmente, mais de 1600 TFs de humanos são conhecidos (LAMBERT *et al.*, 2018) e um total de aproximadamente 1700 a 1900 são estimados no genoma humano (VAQUERIZAS *et al.*, 2009). Ser capaz de prever os TFBSs potencialmente funcionais é um importante passo inicial em análises de TF (CARTHARIUS *et al.*, 2005).

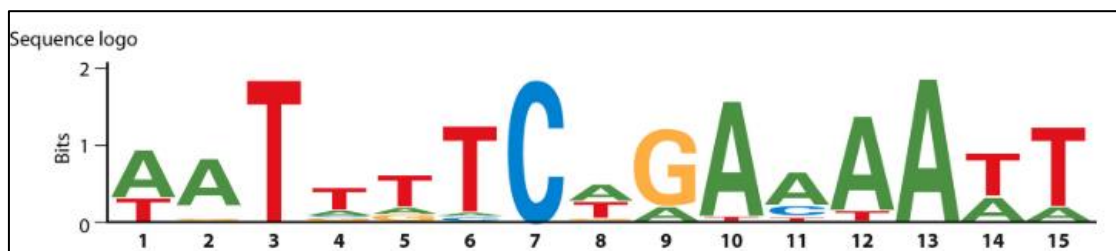
Descobrir TFBS em um conjunto de sequências de DNA, por exemplo, em um *enhancer*, é uma tarefa difícil devido à tendência dos sítios de ligação de serem curtos e degenerados, e o algoritmo procura por um conjunto de sequências curtas familiares em um *set* de sequências bem mais longas (BAILEY *et al.*, 2006). TFBS para um específico TF podem se diferenciar em sequência nucleotídica e de composição, mas eles podem ser representados por uma sequência consenso, ou seja, a representação da variabilidade do alvo do TF (VAN HIJUM *et al.*, 2009)

Uma grande questão na descoberta de motivos em TFBS é a maneira como eles são representados. Provavelmente a maneira mais simples de representação de motivos é o uso da sequência consenso de nucleotídeos (A, C, G e T), a qual pode ser representada de uma maneira estrita, no caso representando apenas a sequência ótima, ou degenerada, onde temos, por exemplo: R é purina, Y é pirimidina, S é *strong*, W é *weak*, K é *keto*, M é *amino*, e N é qualquer nucleotídeo de acordo com a nomenclatura da IUPAC. Neste caso, uma quantidade limitada de informação pode ser representada nas proporções de nucleotídeos nas dadas posições (VAN HIJUM *et al.*, 2009).

A matriz de peso posicional (PWMs, *position weight matrix*) é atualmente o modelo mais comum para identificar motivos de TFBS e são mais precisos que representações baseadas em consenso. Em PWMs, o nucleotídeo observado em uma posição é assumido ser independente dos nucleotídeos em outras posições. Por conveniência os motivos são

visualizados por uma sequência denominada *logo* consistindo de uma pilha de letras ordenadas, na qual a letra de maior tamanho indica a quantidade de informação que o motivo contém em dada posição (Fig. 3). Algoritmos computacionais são então usados para identificar os motivos que podem ser responsáveis pela ligação dos TFs (VAN HIJUM *et al.*, 2009).

Figura 3. Exemplo da apresentação de um logo de um motivo hipotético



Fonte: VAN HIJUM *et al.*, 2009

1.4. Ordem Primates

Quando analisamos a Ordem Primates como um todo, as classificações morfológicas tradicionais identificam duas subordens: Strepsirrhini (“nariz fendido”) e Haplorrhini (“nariz simples”) (JOBILING *et al.*, 2014), sendo a principal distinção entre as subordens a estrutura do nariz. Na primeira subordem o nariz está ligado ao lábio superior que por sua vez se encontra fundido à gengiva, e há a presença do rinário, a superfície úmida e sem pelos ao redor do nariz (BEARD, 1988; MITTERMEIER *et al.*, 2013). Na segunda subordem o nariz possui uma membrana ao redor das narinas, não existe o rinário, e o lábio superior não está diretamente conectado ao nariz ou fundido à gengiva (MITTERMEIER *et al.*, 2013).

A subordem Strepsirrhini inclui três infraordens, os Chiromyformes, os Lemuriformes e os Lorisiformes, enquanto a subordem Haplorrhini engloba os Tarsiformes e Simiiformes (Anthropoidea) (PERELMAN *et al.*, 2011). Estudos genéticos demonstraram que Chiromyformes foi a primeira infraordem a divergir, há cerca de 66 milhões de anos, e atualmente a mesma é composta por uma única espécie, *Daubentonia madagascariensis* (aye-aye). Tal espécie possui várias características peculiares, como: dentes incisivos de crescimento contínuo (como em roedores), grandes pavilhões auditivos, dentição reduzida e um dedo médio muito longo usado para extração de larvas de insetos em árvores (MITTERMEIER *et al.*, 2013).

As infraordens Lemuriformes, Lorisiformes, Tarsiformes e Chiromyformes são

classificados, como prosímios. Outras características apresentadas e estudos moleculares mostram que os Tarsiformes compartilham um ancestral comum com os Anthropeoidea (POUX & DOUZERY, 2004; HARTIG *et al.*, 2013), sendo que sua divergência ocorreu cerca 87 milhões de anos atrás, logo após a divergência de Haplorrhinis e Strepsirrhinis (PERELMAN *et al.*, 2011; MITTERMEIER *et al.*, 2013).

A infraordem Simiiformes é dividida em duas parvordens: Platyrrhini que engloba os chamados macacos do novo mundo, e Catarrhini que engloba a superfamília Cercopithecoidea, que são os macacos do velho mundo, e a superfamília Hominoidea que abrange os “grandes primatas” (gibões, orangotangos, gorilas, chimpanzés e humanos) (JOBLING *et al.*, 2014). A divergência entre essas duas superfamílias ocorreu entre 45 e 37 milhões de anos atrás, de acordo com observações fósseis (MASTERS, 2009) e moleculares (PERELMAN *et al.*, 2011). Os Platyrrhini são assim chamados por possuírem um nariz largo e achatado e suas narinas serem distantes entre si e voltadas para os lados enquanto os Catarrhini possuem narinas mais estreitas e próximas umas das outras. Os Platyrrhini têm várias outras características distintas como: três dentes pré-molares, ausência de ossos internos do ouvido, e a sínfise dos ossos zigomático e parietal. Os Catarrhini possuem dois dentes pré-molares, tuba auditiva, e os ossos zigomático e parietal não se fundem, estando separados pela sínfise entre os ossos frontal e esenoide (MITTERMEIER *et al.*, 2013).

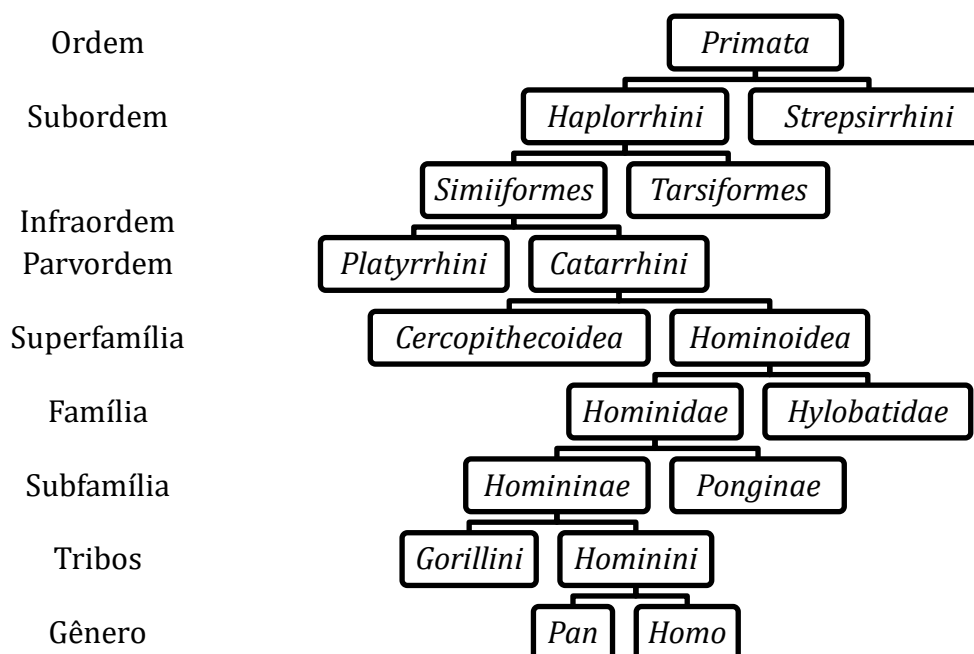
Apesar dos Platyrrhini e dos Cercopithecoidea (macacos do novo e do velho mundo respectivamente) compartilharem numerosos traços ancestrais, como por exemplo cauda e membros anteriores ligeiramente mais curtos, um grande número de caracteres derivados, incluindo a orientação da narina, aspectos comuns da dentição e evidências moleculares, indica que os macacos do velho mundo e os hominoides compartilham um ancestral comum mais recente (JOBLING *et al.*, 2014). Os macacos do velho mundo possuem cauda não preênsil, enquanto os hominoides, ou grandes macacos, não possuem cauda. Dados moleculares indicam que os Cercopithecoidea divergiram dos Hominoidea entre 29 e 23 milhões de anos atrás (RAAUM *et al.*, 2005).

A superfamília Hominoidea é dividida em duas famílias: Hylobatidae (gibões) e Hominidae (orangotangos, gorilas, chimpanzés e humanos). A divergência entre essas duas famílias teria ocorrido há 16,8 milhões de anos. A família Hominidae divide-se em duas subfamílias: Homininae e Ponginae. A divergência entre essas duas subfamílias teria ocorrido há cerca de 14 milhões de anos. A subfamília Ponginae é composta pelos orangotangos (RAAUM *et al.*, 2005).

Ao longo da década de 1960, Morris Goodman usou uma variedade de testes

imunológicos de proteínas séricas para investigar o parentesco genético entre humanos, chimpanzés e gorilas. Baseado em seus resultados moleculares, chimpanzés e gorilas foram agrupados junto com os homens na subfamília Homininae (HAGEN, 2010). Tal família divide-se em duas tribos: Gorillini (gorilas) e Hominini (chimpanzés e humanos). A tribo Hominini inclui os chimpanzés (gênero *Pan*) e os humanos (gênero *Homo*), incluindo espécies humanas extintas como o Denisova e Neandertais. Estudos moleculares têm concluído que os gêneros *Pan* e *Homo* têm antepassados comuns que viveram até 6 a 7 milhões de anos atrás (RAAUM *et al.*, 2005) (Figura 4).

Figura 4. Esquemática da Ordem Primates



Fonte: Gabrielle A. Rizzato

Apesar de todas as diferenças morfológicas, uma característica em comum entre todos primatas é possuir mãos preênsis, ou seja, possuir a capacidade de segurar objetos. Isso se dá graças a suas mãos com longas falanges em relação aos metacarpos. Quando comparamos os humanos com outros primatas, principalmente com seus parentes mais próximos (como bonobos e chimpanzés), humanos possuem uma combinação de características “únicas”, incluindo uma forma peculiar de locomoção (bipedalismo), um cérebro grande e mãos com grande destreza (ALMÉCIJA & SHERWOOD, 2017).

Humanos e chimpanzés (e bonobos) diferem consideravelmente nas capacidades cognitivas e linguísticas, apesar de possuírem apenas aproximadamente 1% de diferença na

sequência de DNA genômico (GU & GU, 2003). Um desafio de longa data na biologia evolucionária é descobrir mudanças genômicas que sejam responsáveis pelos atributos únicos da espécie humana.

Mudanças genômicas que alteraram a maquinaria molecular do desenvolvimento provavelmente facilitaram a evolução dessas características humanas únicas. Embora essas modificações genéticas continuem em grande parte não identificadas, tem se levado em consideração que elas incluam mudanças na expressão gênica resultantes de substituições nucleotídicas direcionadas por seleção positiva que levaram à alteração na atividade de CREs como o HACNS1 (PRABHAKAR *et al.*, 2008).

Enquanto estudos com o HACNS1 mostram as mutações que levam a diferenças dos humanos em relação aos chimpanzés, não há estudos abrangentes envolvendo o HANCS1 em espécies de macacos do velho mundo e grandes macacos, bem como nenhum estudo em macacos do novo mundo. É importante considerar que a divergência entre Platyrrhini e Catarrhini ocorreu entre 45 e 37 milhões de anos e é possível que tenham ocorrido outras mutações nesse grupo que possibilitaram o desenvolvimento de outras características próprias a cada grupo.

Como exemplo de características únicas dos primatas do novo mundo temos que dentro da família Cebidae os exemplares do gênero *Cebus* possuem polegar oponível e cauda semipreênsil, os do gênero *Mico* possuem garras em todos os dígitos menos no hálux, os do *Leontopithecus* possuem dedos alongados e o *Callimico* apresenta polegar não oponível, hálux oponível e garras em todos os dígitos menos no hálux. Já na família Atelidae, os exemplares do gênero *Brachyteles* possuem polegares vestigiais ou inexistentes, inclusive sendo a principal diferença entre as espécies, *B. arachnoides* tendo polegar vestigial enquanto *B. hypoxanthus* tendo polegar ausente (REIS *et al.*, 2008).

2. OBJETIVOS GERAIS

Considerando que o HACNS1 está implicado na formação do polegar e possivelmente relacionado com a destreza manual, e que essas características são variadas em primatas do novo e velho mundo e grandes macacos, o presente estudo visou à caracterização molecular do *enhancer* HACNS1 em 22 espécies de primatas do novo mundo, 10 espécies do velho mundo, 5 espécies de grandes macacos e 4 espécies de prossímios, quanto à existência de seleção natural e o ganho e perda de sítios de motivo de ligação para fatores de transcrição.

2.1. Objetivos Específicos

- Verificar a variabilidade nucleotídica do *enhancer* HACNS1 nas 10 espécies do velho mundo, 5 espécies de grandes macacos, 4 espécies de prossímios e 4 espécies de novo mundo de primatas já sequenciadas e disponíveis em banco de dados.

- Amplificar, por meio de PCRs (*Polymerase Chain Reaction*) utilizando *primers* específicos desenhados para este estudo, e sequenciar a região do HACNS1 nas 18 espécies de macacos de novo mundo disponíveis no banco de DNA do Laboratório de Evolução Humana e Molecular da Universidade Federal do Rio Grande do Sul

- Amplificar, por meio de PCRs utilizando *primers* específicos desenhados para este estudo, e sequenciar um segmento de um éxon próximo ao íntron onde se localiza o HACNS1, de tamanho 173pb, para utilizar como região codificadora nas análises de seleção

- Comparar as taxas de substituições em regiões não codificantes em relação às taxas de substituições sinônimas nas regiões codificadoras adjacentes utilizando o programa EvoNC para determinar se o padrão de variação do *enhancer* apresenta sinais de seleção positiva.

- Realizar uma predição de sítios de ligação de fatores de transcrição com os *softwares* DiAlign TF e MatInspector para identificar fatores de transcrição que ativam genes de desenvolvimento diferentes famílias de primatas.

3. MATERIAIS E MÉTODOS

3.1. Amostras *in silico*

Foram utilizadas nesse estudo 24 sequências do HACNS1 obtidas *in silico* no banco de dados Ensembl (HUNT *et al.*, 2018) (Tabela 1).

Tabela 1. Espécies de primatas com genoma completo disponível.

Espécie	Nome Comum	Identificação Taxonômica	Assembly
<i>Microcebus murinus</i>	Rato-lemuriano	30608	Mmur3.0
<i>Otolemur garnettii</i>	Bushbaby	30611	OtoGar3
<i>Propithecus coquereli</i>	Sifaka-de-coquerel	379532	Pcoq_1.0
<i>Carlito syrichta</i>	Társio	1868482	Tarsius_syrichta-2.0.1
<i>Callithrix jacchus</i>	Sagui-de-tufo-branco	9483	ASM275486v1
<i>Saimiri boliviensis</i>	Saimiri	39432	SaiBol1.0
<i>Cebus capucinus imitator</i>	Capuchinho de cara branca	1737458	Cebus_imitator-1.0
<i>Aotus nancymae</i>	Macaco da noite	37293	Anan_2.0
<i>Cercocebus atys</i>	Mangabei	9531	Caty_1.0
<i>Macaca fascicularis</i>	Macaco carangueijo	9541	Macaca_fascicularis_5.0
<i>Macaca mulatta</i>	Rhesus	9544	Mmul_8.0.1
<i>Macaca nemestrina</i>	Macaco-rabo-de-porco	9545	Mnem_1.0
<i>Papio Anubis</i>	Babuíno	9555	Panu_3.0
<i>Mandrillus leucophaeus</i>	Mandrill	9568	Mleu.le_1.0
<i>Chlorocebus sabaues</i>	Macaco verde	60711	ChlSab1.1
<i>Colobus angolensis palliatus</i>	Colobus	336983	Cang.pa_1.0
<i>Rhinopithecus bieti</i>	Macaco-preto-de-nariz-arrebitado	61621	ASM169854v1

Espécie	Nome Comum	Identificação Taxonômica	Assembly
<i>Rhinopithecus roxellana</i>	Macaco-dourado	61622	Rrox_v1
<i>Pongo abeli</i>	Orangotango	9601	PPYG2
<i>Gorilla gorilla gorilla</i>	Gorila	9595	gorGor4
<i>Nomascus leucogenys</i>	Gibão	61853	Nleu3.0
<i>Pan paniscus</i>	Bonobo	9597	panpan1.1
<i>Pan troglodytes</i>	Chimpanzee	9598	Pan_tro_3.0
<i>Homo sapiens</i>	Human	9606	GRCh38

Fonte: Gabrielle A. Rizzato

3.2. Amostras de DNA

Foram utilizadas 18 amostras de DNA de primatas do novo mundo (Tabela 2) do banco de DNA do Laboratório de Evolução Humana e Molecular da Universidade Federal do Rio Grande do Sul, provenientes do Centro de Primatologia do Rio de Janeiro (CPRJ) para a obtenção da sequência do HACNS1 por PCR e sequenciamento.

Tabela 2. Amostras de primatas no novo mundo empregadas para amplificação e sequenciamento da região genômica HACNS1.

Espécie	Nome Comum	Habitat	Identificação Taxonômica
<i>Brachyteles arachnoides</i>	Muriqui	Mata Atlântica	30594
<i>Cacajao melanocephalus</i>	Ucari-preto	Amazônia	70825
<i>Callibella humilis</i>	Sagui-anão	Amazônia	666519
<i>Callicebus coimbrai</i>	Guigó	Mata Atlântica	867333
<i>Callimico goeldii</i>	Mico-de-goeldi	Amazônia	9495
<i>Callithrix geoffroyi</i>	Sagui-de-cara-branca	Mata Atlântica ES, MG e RJ	52231
<i>Cebuella pygmaea</i>	Sagui-leãozinho	Amazônia	9493
<i>Cebus robustus</i>	Macaco-prego	Mata Atlântica e Cerrado	867367
<i>Cebus xanthosternos</i>	Macaco-prego-do-peito-amarelo	Mata Atlântica	174599
<i>Chiropotes albinasus</i>	Cuxiú-de-nariz-branco	Sul da Amazônia	198627
<i>Chiropotes utahickae</i>	Cuxiú	Amazônia	280160
<i>Leontopithecus chrysomelas</i>	Mico-leão-da-cara-dourada	Mata Atlântica do Sul da BA	57374
<i>Leontopithecus rosalia</i>	Mico-leão-dourado	Mata Atlântica	30588
<i>Mico humeralifer</i>	Sagui-de-santarém	Amazônia	52232
<i>Mico melanura</i>	Sagui-do-cerrado	Amazônia, Cerrado e Chaco	1090896
<i>Saguinus bicolor</i>	Sagui	Amazônia	37588
<i>Saguinus martinsi</i>	Sagui	Amazônia	356666
<i>Saguinus niger</i>	Sagui	Amazônia	356665

Fonte: Gabrielle A. Rizzato

3.3. Análises Laboratoriais

As amostras de DNA foram amplificadas via PCR, utilizando um par de *primers* específicos desenhados para este estudo para a região do HACNS1 (F-5'TCAGCAGTGAGGTAGGAATAGA3'; R-5'GCCGGTTCTGAATCTTTAACCA3'). O volume total da reação foi de 30 µl, sendo que por reação foram utilizados 0,3 µM de cada primer, 1,5 µM de MgCl₂, 0,2 mM de dNTPs e 0,2 U de *DreamTaq Hot Start DNA Polymerase*. As reações foram colocadas em termociclador *Veriti Thermal Cycler* (Applied Biosystems®) com temperatura inicial de 95°C por 1min para desnaturação inicial, seguido de 30 ciclos de 95°C por 30s para desnaturação, 50°C por 30s para anelamento dos *primers* e 72°C por 1min para a extensão, por último um período final de extensão de 72°C por 10 min, de acordo com as orientações do fabricante.

As reações foram verificadas através de eletroforese em gel de agarose 1% com *Safer Dye*, e a seguir, enviadas para purificação e sequenciamento em um sequenciador *3730xl DNA Analyzer* (Applied Biosystems®). Foram amplificadas e sequenciadas amostras de 18 indivíduos, apresentados na Tabela 3.

Os cromatogramas foram analisados manualmente com o programa CodonCode Aligner (<https://www.codoncode.com/>) para avaliar a confiabilidade dos picos (Q-score > 30) e confirmar se as mutações encontradas eram reais. Cada mutação encontrada na sequência teve ambas as fitas (*forward* e *reverse*) analisadas para checar a veracidade da informação.

As sequências ortólogas dos outros primatas foram obtidas nos bancos de dados do Ensembl (<https://www.ensembl.org/>) e NCBI (<https://www.ncbi.nlm.nih.gov/>), e os alinhamentos foram feitos com BioEdit (HALL, 1997).

Para a região codificadora foi utilizado outro par *primers* desenhados para este estudo (F-5'CACAGTGGAAGGGTCATAG3'; R-5'GTGATAGTTACTGTCTCCGC3'), específico para a região do éxon 11 do gene *AGAPI-201* (*transcript* ID: ENST00000336665.9), por ser o éxon mais próximo e de tamanho similar do HACNS1, o qual se localiza no íntron 9 do mesmo gene. O volume total da reação foi de 30 µl, sendo que por reação foram utilizados 0,2 µM de cada primer, 1,3 µM de MgCl₂, 0,2 mM de dNTPs e 0,2 U de *PlatinumTaq DNA Polymerase High Fidelity*. As reações foram colocadas em termociclador *Veriti Thermal Cycler* (Applied Biosystems®) com temperatura inicial de 94°C por 1min para desnaturação inicial, seguido de 40 ciclos de 94°C por 15s para desnaturação, 51,4°C por 30s para

anelamento dos *primers* e 72°C por 1min para a extensão, e por último um período final de extensão de 72°C por 10 min, de acordo com as orientações do fabricante.

As reações foram verificadas através de eletroforese horizontal em gel de agarose 2% com *Safer Dye*, e a seguir, enviadas para purificação e sequenciamento. Foram amplificadas e sequenciadas as amostras dos mesmos 18 indivíduos do HACNS1, apresentados na tabela 3.

Os cromatogramas foram analisados manualmente com o programa CodonCode Aligner (<https://www.codoncode.com/>) para avaliar a confiabilidade dos picos (Q-score > 30) e confirmar se as mutações encontradas eram reais.

As seqüências ortólogas dos outros primatas foram obtidas nos bancos de dados do Ensembl e NCBI, e os alinhamentos foram feitos com BioEdit (HALL, 1997).

Tabela 3. Espécimes utilizados para sequenciamento do enhancer HACNS1 e do exon 11 do gene AGAP1-201, e seus respectivos códigos no banco de amostras do Laboratório de Evolução Humana e Molecular da Universidade Federal do Rio Grande do Sul.

Número	Espécie
M2160	<i>Brachyteles arachnoides</i>
M2072	<i>Cacajao melanocephalus</i>
M2201	<i>Callibella humilis</i>
M2275	<i>Callicebus coimbrai</i>
M2033	<i>Callimico goeldii</i>
M2217	<i>Callithrix geoffroyi</i>
M2074	<i>Cebuella pygmaea</i>
M2209	<i>Cebus robustus</i>
M0474	<i>Cebus xanthosternos</i>
M2094	<i>Chiropotes albinasus</i>
M2095	<i>Chiropotes uthaika</i>
M2186	<i>Leontopithecus chrysomelas</i>
M1620	<i>Leontopithecus rosalia</i>
M1799	<i>Mico humeralifer</i>
M1942	<i>Mico melanura</i>
M2043	<i>Saguinus bicolor</i>
M1446	<i>Saguinus martinsi</i>
M2011	<i>Saguinus niger</i>

Fonte: Gabrielle A. Rizzato

3.4. Análise de Seleção

Diversos métodos podem ser usados para identificar seleção atuando em regiões codificadoras, sendo uma das abordagens mais comuns a estimativa de ômega (ω), calculado

pela taxa de substituição não-sinônima (d_N) sobre a taxa de substituição sinônima (d_S), ou seja: ω (d_N/d_S) (GOLDMAN & YANG, 1994).

Considerando que as taxas de mutação em sítios sinônimos e não-sinônimos dentro de um gene devem ser as mesmas, ou ao menos similares, e que por ser um processo inteiramente randômico essas mutações deveriam ter a mesma chance de fixação, as diferenças nas taxas de substituição entre sítios sinônimos e não-sinônimos pode ser atribuída à diferenças na intensidade da seleção purificadora entre esses dois tipos de sítios. Mutações que resultam em uma substituição de aminoácido tem uma maior chance de causar efeitos deletérios na função da proteína do que mutações sinônimas. Consequentemente, a maioria das mutações não-sinônimas vai ser eliminada das populações por seleção purificadora. O resultado vai ser uma redução na taxa de substituição em sítios não-sinônimos. Em contraste, mutações sinônimas tem uma maior chance de serem neutras, e uma maior proporção delas será fixada na população. Porém, existe uma chance que mutações não-sinônimas melhorem a função da proteína. Dessa forma, se uma seleção positiva estiver atuando na evolução das proteínas, a taxa de substituição não-sinônima deve exceder a de substituição sinônima (GRAUR & LI, 2000).

Logo, quando um sítio está sob seleção purificadora, substituições sinônimas ocorrem mais rapidamente que substituições não-sinônimas e $\omega < 1$. No entanto, quando não há seleção ocorrendo a taxa de substituições sinônimas é igual à taxa de substituições não-sinônimas e $\omega = 1$, por outro lado, se está ocorrendo seleção positiva, novas mutações são benéficas e $\omega > 1$ (WONG & NIELSEN, 2004).

Porém, regiões não codificadoras não se traduzem em aminoácidos, logo não é possível pensar em termos do que é uma substituição sinônima ou não-sinônima considerando apenas essas regiões. Para conseguir isso, Wong e Nielsen (2004) modelaram a evolução em regiões codificadoras e não codificadoras em conjunto, assumindo que a taxa de substituição sinônima é constante em ambas as regiões em um mesmo gene.

Com base nisso é calculado o parâmetro zeta (ζ) para modelar a evolução na região não codificadora, sendo a taxa de substituição de nucleotídeos na região não codificadora normalizada pela taxa de substituição sinônima na região codificadora. Sendo assim, o programa EvoNC (WONG & NIELSEN, 2004) compara as taxas de substituições em regiões não codificadoras em relação às taxas de substituições sinônimas nas regiões codificadoras adjacentes. Logo, se um sítio está sujeito à seleção purificadora $\zeta < 1$, similarmente $\zeta = 1$ indica evolução neutra e $\zeta > 1$ sugere a presença de seleção positiva, sendo a interpretação de

ζ similar à interpretação de ω em modelos de evolução em regiões codificadoras (WONG & NIELSEN, 2004).

Para a análise de seleção positiva desse estudo foi utilizado o *software* HyPhy (<http://www.hyphy.org/>) no qual é possível usar um arquivo *batch file* (<https://github.com/ofedrico/TestForPositiveSelection>) para testar assinaturas de seleção positiva em sequências não codificadoras. O método é uma versão aprimorada do EvoNC (WONG & NIELSEN, 2004).

Assim como o EvoNC esse *script* permite usar substituições sinônimas de regiões codificadoras próximas como *proxies* neutros para o teste. Também é possível observar a seleção positiva em qualquer ramo e executar o algoritmo *Bayes Empirical Bayes* (BEB) (YANG *et al.*, 2005) para detectar quais sítios estão sob seleção positiva, seleção negativa ou evolução neutra.

Nesse *script* três modelos são testados, o modelo neutro, o modelo de duas categorias e o modelo de três categorias. No modelo neutro é assumido que não há seleção positiva, logo ζ pode apenas ser ≤ 1 . No modelo de duas categorias (*Alternate 1*) ζ pode ser ≤ 1 (H_0 = seleção negativa/evolução neutra) e pode ser > 1 . No modelo de três categorias (*Alternate 2*), podemos ter $\zeta < 1$, $\zeta = 1$ e $\zeta > 1$.

Para essas análises de seleção foram utilizadas as sequências do *enhancer* HACNS1 (173 pb) mais o exon 11 do gene *AGAPI-201* (166 pb) como as regiões não codificadora e codificadora adjacente, respectivamente. Estes alinhamentos em conjunto com uma árvore filogenética dos primatas analisados (Anexo I) criada pelo *software* FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>) foram então utilizados para identificar o padrão evolutivo do *enhancer* HACNS1 nos primatas.

Foram testados os modelos *Alternate 1* e *Alternate 2* contra o modelo nulo, com e sem especificidade de ramo (com especificidade foi testado se o ramo dos primatas do novo mundo estava sob seleção positiva enquanto o resto da árvore não). No *Alternate 2* foi executado o algoritmo BEB (YANG *et al.*, 2005) que infere a probabilidade posterior de cada sítio estar sob seleção positiva.

3.5. Análise de TFBS

A predição de sítios de transcrição foi feita com os *software* MatInspector (QUANDT

et al., 1995) e DiAlign TF (CARTHARIUS *et al.*, 2005), ambos do *Genomatix Software Suit* (www.genomatix.de).

MatInspector é uma ferramenta que utiliza uma grande biblioteca de matrizes para TFBS para localizar correspondências nas sequências de DNA e retorna uma lista de potenciais TFBSs com base em uma correspondência estatística entre uma região na sequência e essas matrizes (MASTON *et al.*, 2006). Sua primeira versão foi descrita por Quandt *et al.* (1995) onde foi introduzido um conjunto de ferramentas para a geração de matrizes (MatInd) e a detecção de potenciais TFBS (MatInspector). O MatInspector usa as matrizes para varrer as sequências de nucleotídeos em busca de correspondências a esse padrão, enquanto o DiAlignTF é uma combinação do MatInspector com o programa de alinhamento múltiplo DiAlign (MORGENSTERN *et al.*, 1998) e detecta TFBS em regiões de sequências ortólogas (CARTHARIUS *et al.*, 2005)

Os recursos mais importantes dos programas de previsão do TFBS são a cobertura mais alta possível de sítios conhecidos e a qualidade das descrições de materiais na biblioteca. A biblioteca de matrizes MatInspector foi projetada para representar o melhor dos conhecimentos atuais em termos de especificidade e sensibilidade em relação ao TFBS e dessa forma pode encontrar os possíveis locais de ligação de vários ativadores e repressores que se ligam a sequências reguladoras específicas de DNA (CARTHARIUS *et al.*, 2005) e por isso foi o software escolhido para realizar as análises deste estudo.

4. RESULTADOS

4.1. Variabilidade Molecular

Tanto a obtenção das sequências em bancos de dados online quanto o sequenciamento do HACNS1 em todos os primatas de novo mundo disponíveis no banco de DNA foram bem-sucedidos. Quanto ao éxon 11 do *AGAP-201*, só não foram possíveis obter as sequências do *Brachyteles arachnoides* e do *Callicebus coimbrai*, e por esse motivo essas duas espécies não foram incluídas na análise de seleção. As sequências completas obtidas se encontram nos Anexo II e Anexo III. Os resultados do alinhamento do HACNS1 com o BioEdit se encontram no Anexo IV.

Quatro substituições foram observadas em todos os primatas do novo mundo analisados. Desta, uma substituição é específica de primatas do novo mundo, a adenina na posição 27 da sequência (Anexo IV). Interessantemente, é no mesmo local onde acontece uma das substituições específicas de humanos (PRABHAKAR *et al.*, 2008). Outra mutação específica de primatas de novo mundo é uma timina encontrada em três espécies na posição 7 da sequência (Anexo IV), sendo que ela está presente em *Mico humeralifer*, *Saguinus martinsi* e *Saguinus niger*, mas não está presente em *Mico melanura* e *Saguinus bicolor*, ambos os gêneros pertencentes à família Cebidae. A última mutação específica de primatas do novo mundo encontrada nos sequenciamentos desse estudo foi uma citosina na posição 135 da sequência de *Callimico goeldii* (Anexo IV).

Considerando as sequências dos macacos do novo mundo obtidas no banco de dados Ensembl, foram observadas duas substituições em *Saimiri boliviensis*, uma citosina na posição 72 e uma adenina na posição 79 (Anexo IV). *Aotus nancymae* apresentou também duas substituições, uma citosina na posição 50 e uma guanina na posição 56 (Anexo IV), sendo que o mais interessante dessas substituições é que elas são idênticas as que foram identificadas especificamente em humanos por Prabhakar *et al.* (2008).

Além das substituições específicas de primatas do novo mundo, foram identificadas quatro outras substituições em um grande número de espécies, indicando estarem conservadas. Uma timina na posição 109 foi identificada em todos os taxa menos na Superfamília Hominoidea (e em *Microcebus murinus*), uma guanina na posição 165 também foi identificada em todos os taxa menos na Subfamília Homininae e uma timina na posição 168 foi identificada na Subordem Strepsirrhini, Infraordem Tarsiforme, Parvordem Platyrrhini

e em uma espécie da Superfamília Cercopithecoidea, o *P. coquereli* (Anexo IV). A presença dessas substituições em todos esses taxa sugere que esses nucleotídeos provavelmente sejam o estado ancestral. Uma adenina na posição 6 da sequência foi identificada em toda a Superfamília Cercopithecoidea, menos em *Propithecus coquereli* (Anexo IV).

4.2. Análise de Seleção

Nossas análises de seleção positiva obtiveram os seguintes resultados: na comparação do modelo nulo, tanto com o modelo *Alternate 1* (com e sem especificação de ramo dos primatas do novo mundo), quanto com o modelo *Alternate 2* (com especificação do ramo dos primatas do novo mundo), os resultados dos LRTs (*likelihood ratio test*) foram não significativos. O LRT é usado para determinar se o modelo testado é significativamente melhor do que o modelo nulo. Como o resultado obtido dos LRTs foi um p-valor de aproximadamente = 1.0, isso significa que o algoritmo não detectou que nenhum desses modelos seja mais provável de explicar os dados do que o modelo nulo.

O único modelo que apresentou um p-valor significativo foi o *Alternate 2*, o modelo de três categorias, sem especificação de ramo dos primatas do novo mundo, como mostrado na Tabela 4, mostrando que o modelo alternativo que admite valores de ζ diferentes é o que possui a maior probabilidade posterior de explicar a evolução do HACNS1.

Tabela 4. Resultados da análise de seleção dos modelos neutros e seus respectivos modelos Alternate 1 e Alternate 2, com e sem especificação de ramos.

Modelo	Parâmetros estimados	<i>l</i>	P-value
Modelo neutro		-1034.987357655307	
<i>Alternate 1</i> (sem especificação de ramos)		-1034.987357432572	0.9994674639
Modelo neutro		-1034.987357403621	
<i>Alternate 2</i> (sem especificação de ramos)	$\zeta[0] = 0.169065$, $p[2] = 0,83236$ $\zeta[1] = 1$, $p[1] = 0$ $\zeta[2] = 0.024328$, $p[2] = 0,16763$	-1014.430394721923	0.0000000255
Modelo neutro		-1034.987357410729	
<i>Alternate 1</i> (com especificação de ramos)		-1034.987357598781	1
Modelo neutro		-1013.728003216408	
<i>Alternate 2</i> (com especificação de ramos)		-1013.727096390003	0.9660307563

Fonte: Gabrielle A. Rizzato

Apesar do p-valor significativo que sugere que o modelo *Alternate 2*, com variação do ζ entre os sítios, seja o que melhor representa a evolução da sequência, não foi encontrado valor de $\zeta[2] > 1$ (Tabela 4). Esse resultado, junto com o resultado do BEB, que indica a probabilidade posterior de cada sítio se encontrar em cada uma das posições $\zeta[0]$, $\zeta[1]$ ou $\zeta[2]$ (Anexo V) sugere que não há seleção positiva.

4.3. Análise de TFBS

As análises de predição de TFBS sugerem que houve um grande número de locais que foram ganhos e perdidos no HACNS1 durante a evolução dos primatas. Prabhakar *et al.* (2008) também fizeram uma predição de TFBS e um sítio de ligação foi predito como ganho se foi observada a sua presença na sequência humana, mas não na dos outros primatas ou vertebrados analisados, e predito como perdido se não observado em humano mas observado

nos outros primatas e vertebrados analisados. O mesmo critério foi utilizado no presente estudo, tanto para humanos quanto para os outros primatas.

Quanto aos ganhos de novos motivos nos primatas do novo mundo, esses foram: um sítio de ligação do fator de transcrição MYT1 (*Myelin transcription factor 1*) em *Callimico goeldii*; um sítio de ligação para cada um dos fatores de transcrição AML3 (*Runt-related transcription factor 2*) e ZKSCAN3 (*Zinc finger transcription factor*) em *Saimiri boliviensis*; um sítio de ligação para cada um dos fatores de transcrição ZSCAN4 (*Zinc finger and SCAN domain containing 4*), PLAGL1 (*Pleomorphic adenoma gene-like 1*) e RFX5 em *Aotus nancymae*.

A presença de um TFBS não significa que ele seja funcional, e após uma extensa revisão bibliográfica obteve-se os principais resultados apresentados no Anexo VI e resumidos na Tabela 5. Foi identificado uma mutação pontual de uma guanina para uma citosina na posição 165 do *Homo sapiens* e seus parentes mais próximos, *Pan troglodytes* e *Pan paniscus*, e *Gorilla gorilla* (Anexo IV) que levou ao ganho de um motivo para ligação para um fator de transcrição de regulação da mielina, o MYRF (*Myelin regulatory factor*), com um *score* de 0,974 na matriz de similaridade, do nucleotídeo 155 ao 167 na fita *reverse*. Também foi identificado a mutação pontual de uma adenina para uma guanina na posição 86 da sequência do HACNS1 de *H. sapiens* (Anexo IV) que levou ao ganho de dois motivos para ligação para um ativador da transcrição com ligação à calmodulina, o CAMTA (*Calmodulin-binding transcription activator*), ambos com *score* de aproximadamente 0,842 na matriz de similaridade, indo do nucleotídeo 78 ao 86, um na fita *forward* e outro na fita *reverse*. Finalmente, foram identificadas quatro mutações pontuais, de timina pra guanina na posição 68, de adenina pra guanina na posição 70, de timina pra citosina na posição 72 e de adenina pra guanina na posição 73 na sequência do HACNS1 de *H. sapiens*. Essas quatro mutações levaram à perda de um TFBS para o fator de transcrição PTX1 (*Pituitary Homeobox 1*), indo do nucleotídeo 67 ao 83, na fita *forward* com *score* de 0,945 na matriz de similaridade.

Tabela 5. Principais TFBS identificados como ganhos ou perdidos no HACNS1. “Do” e “ao” se referem ao número do nucleotídeo em que os TFBS começam e terminam, respectivamente, nas sequências do Anexo II.

TFBS	Matriz detalhada	do	ao	Fita	Matriz sim.	Sequência
CAMTA	<i>Calmodulin-binding transcription activator</i>	78	86	(+)	0.842	cgccg TGCG
CAMTA	<i>Calmodulin-binding transcription activator</i>	78	86	(-)	0.841	cgcac GGCG
PTX1	<i>Pituitary Homeobox 1</i>	67	83	(+)	0.945	gtca CTAA ccccgcgt
MYRF	<i>Myelin regulatory factor</i>	155	167	(-)	0.974	ttgtg CCAG ggtg

Fonte: Gabrielle A. Rizzato

5. DISCUSSÃO

Prabhakar *et al.* (2008) investigaram a funcionalidade do HACNS1 em humanos, e evidenciariam sua rápida divergência quando comparado outras espécies, nas quais essa região é fortemente conservada. Devido à quase totalidade das substituições serem GC (14 de 16 substituições específicas de humanos) os autores examinaram se haveria conversão gênica enviesada, que é considerado um processo neutro detectado em *hotspots* de evolução acelerada no genoma nos quais a taxa local de substituição de AT pra GC é aumentada. Ao comparar a taxa de substituições AT-GC em uma janela de 9kb que engloba o HACNS1, com a taxa de substituições AT-GC específica de humanos na mesma área eles inferiram que esta última é muito maior na região de 81 pb do que o esperado sob o modelo neutro, sugerindo seleção positiva. Hünemeier *et al.* (2010) realizaram testes adicionais utilizando dados populacionais e DNA antigo de Neandertal para checar se a variação selecionada poderia ser estendida para todas as populações humanas e para outras espécies do gênero *Homo* mais antigas. Os autores utilizaram o programa EvoNC (WONG & NIELSEN, 2004) para testar se o padrão de variação no *enhancer* HACNS1 poderia ser explicado por um modelo neutro ou não. Comparações entre os modelos mostraram uma evidência significativa de seleção positiva quando as sequências de *H. sapiens*/Neandertal são comparadas com as de outros primatas, incluindo chimpanzé (*Pan troglodytes*). O teste indicou que todos os 13 sítios reconhecidos como exclusivos na linhagem *Homo* mostram sinais de seleção positiva com probabilidade posterior maior que 0,99, evidenciando a ocorrência de seleção positiva não só em humanos, mas já no gênero *Homo*.

O foco inicial desse trabalho era caracterizar o *enhancer* HACNS1 principalmente nos macacos de novo mundo, estendendo o trabalho realizado anteriormente por Hünemeier et al. (2010) Analisando os testes de seleção percebemos que não há indícios que as mutações encontradas nas sequências desses primatas tenham sido fixadas por seleção positiva, nem mesmo a adenina da posição 23, detectada em todos os primatas de novo mundo. O que encontramos de interessante nas análises de seleção foi que todas as seis substituições (nas posições 68, 70, 72, 73, 86 e 165) que levaram aos principais ganhos e perda de TFBS, apresentaram uma probabilidade posterior maior de estarem na classe do $\zeta[2] = 0.024328$ (Anexo V). Esse valor de ζ em conjunto com as probabilidades posteriores encontradas sugerem que esses sítios são mais conservados do que o esperado para uma região não codificadora.

Sobre os fatores de transcrição que tiveram TFBS ganhos nos primatas do novo mundo temos: em *Callimico goeldii*, o fator de transcrição MYT1 (*Myelin transcription factor 1*), envolvido em neurogênese primária, regulação de diferenciação neuronal, e que possui participação na ativação e repressão de diversos outros genes (MANUKYAN *et al.*, 2018; MELHUIH *et al.*, 2018); em *Saimiri boliviensis*, os fatores de transcrição AML3 (*Runt-related transcription factor 2*, também conhecido por RUNX2), relacionado com diferenciação de osteoblastos (GILBERT *et al.*, 2002; BARNES *et al.*, 2003), e ZKSCAN3 (*Zinc finger transcription factor*), um fator de transcrição envolvido na regulação de diversos outros genes associados com a presença de diversos tipos de câncer (YANG *et al.*, 2011; ZHANG *et al.*, 2012; CHAUHAN *et al.*, 2013; CHI *et al.*, 2018); e finalmente em *Aotus nancymae* os fatores de transcrição ZSCAN4 (*Zinc finger and SCAN domain containing 4*), envolvido na manutenção da telomerase (ZALZMAN *et al.*, 2010; DAN *et al.*, 2017), PLAGL1 (*Pleomorphic adenoma gene-like 1*), que funciona como um controle de proliferação celular (VEGA-BENEDETTI *et al.*, 2018), e RFX5 (*Regulatory factor X 5*) um fator de transcrição associado ao complexo MHCII (ZHAO *et al.*, 2017).

A citosina que no alinhamento (Anexo V) se encontra na posição 50 de *A. nancymae* foi identificada manualmente por Prabhakar *et al.* (2008) como fazendo parte de um *half-site* (sítio de ligação para um monômero) para o fator de transcrição ZNF423 (GCACCC[A/T]), o qual está relacionada com diversos papéis no desenvolvimento, podendo atuar também como um repressor envolvido na neurogênese olfativa participando do mecanismo de desenvolvimento que regula a transição da diferenciação para a maturação em neurônios receptores do olfato. A presença dessa mesma substituição em *A. nancymae* sugere uma possível atividade desse mesmo fator de transcrição nessa espécie.

Nenhum desses TFBS previstos como ganhos nos primatas de novo mundo sugerem qualquer relação com as características únicas encontradas nos diferentes polegares dos mesmos, por exemplo, *Callimico* que apresenta polegar não oponível, hálux oponível e garras em todos os dígitos menos no hálux possui apenas um ganho para um TFBS relacionado à neurogênese. Nem a adenina na posição 27 de todos os primatas de novo mundo, nem a timina encontrada na posição 7 de *Mico humeralifer*, *Saguinus martinsi* e *Saguinus niger* (Anexo IV), levaram ao ganho de nenhum TFBS nessas espécies. Já as três espécies de *Cebus* e o *B. arachnoides*, que também possuem características únicas em seus polegares, não apresentaram nenhuma mutação específica.

Programas de previsão de TFBS como o MatInspector podem inferir o potencial de ligação, embora não a funcionalidade de um sítio, a qual pode ser mostrada apenas por experimentos funcionais em laboratório, principalmente porque os potenciais locais de ligação podem ser funcionais em certas células, tecidos ou estágios de desenvolvimento e não funcionais sob diferentes condições. Embora exista consenso de que a inspeção de um TFBS único nas sequências promotoras não é suficiente para entender completamente a regulação gênica, continuará sendo um dos primeiros passos cruciais na cadeia de eventos analíticos para determinar quais sítios tem a maior probabilidade de serem funcionais (CARTHARIUS *et al.*, 2005).

Apesar das descobertas em primatas de novo mundo terem sido bastante limitadas, o fato de analisarmos em conjunto todos os outros primatas disponíveis, nos deu a chance de realizar uma nova análise nos ganhos e perdas de *Homo sapiens* com novos dados que não estavam disponíveis para Prabhakar há dez anos, e dessa maneira obtivemos novas descobertas muito interessantes. Dentre os TFBS ganhos em humanos os que mais sugerem um ganho funcional são o para os fatores de transcrição CAMTA1 e o MYRF (que é um ganho não só de humanos, mas também de *P. troglodytes*, *P. paniscus* e *G. gorilla*). Além disso, a perda de um sítio para o fator de transcrição PTX1 também sugere uma alteração funcional em *H. sapiens*.

5.1. MYRF

No sistema nervoso central (SNC) dos vertebrados, os axônios são firmemente envolvidos pelas bainhas de mielina (KIM *et al.*, 2017). A mielina é uma membrana celular em espiral que envolve e isola os axônios promovendo a rápida propagação de potenciais de ação por condução saltatória ao longo dos axônios. Dessa forma, a mielina é crítica para a rápida propagação de sinais elétricos por aumentar em muito a velocidade da comunicação entre os neurônios, e conseqüentemente, o poder computacional do cérebro, sendo essencial para a realização de funções cerebrais poderosas como as humanas (EMERY *et al.*, 2009; MCKENZIE *et al.*, 2014; KIM *et al.*, 2017).

A mielina do SNC dos vertebrados é sintetizada por oligodendrócitos (OLs) que se originam de células progenitoras (OPCs), as quais surgem de regiões germinais do SNC, proliferam e migram por todo o SNC antes de se diferenciarem em OLs pré-mielinizantes, as

quais subsequentemente aumentam a regulação dos genes da mielina e envolvem os axônios adjacentes (EMERY *et al.*, 2009).

Assim, as bainhas de mielina se desenvolvem à medida que os OLs sofrem diferenciação terminal. A diferenciação de OPCs em OLs e sua subsequente mielinização são processos cruciais e altamente regulados na mielinização do SNC (BUJALKA *et al.*, 2013; MCKENZIE *et al.*, 2014; KIM *et al.*, 2017)

A mielinização é um processo complexo e requer uma intensa coordenação e organização dos processos regulatórios e metabólicos dos genes envolvidos. Por exemplo, os genes que codificam para as proteínas da mielina já são expressos nos estágios de OPCs durante o desenvolvimento, seguidos por sua síntese proteica durante a maturação do OL, e a produção de lipídios da membrana é induzida rapidamente. Muitos reguladores do processo de mielinização foram identificados, incluindo diversos fatores de transcrição, sendo um dos principais o MYRF (CANTONE *et al.*, 2019)

O fator regulador do gene da mielina (MYRF) é considerado um dos principais fatores de transcrição para a produção de mielina por oligodendrócitos no sistema nervoso central, é expresso apenas na diferenciação terminal de oligodendrócitos e controla diretamente a expressão gênica da mielina (EMERY *et al.*, 2009). Emery *et al.* (2009) foram os primeiros a identificar o MYRF como necessário para a maturação de OLs que expressam o gene da mielina no SNC, além de mostrarem que em camundongos nos quais MYRF é excluído, a expressão de muitas proteínas relacionadas à mielina não pode ser detectada e os animais apresentam desmielinização grave, estabelecendo o MYRF como um regulador transcricional vital para a mielinização do SNC.

Depois dele, McKenzie *et al.* (2014) demonstraram que a mielinização ativa, governada pelo MYRF, durante a idade adulta, é essencial para o aprendizado de novas habilidades motoras em camundongos adultos e que o aprendizado motor aumenta a produção de oligodendrócitos. Logo, a exclusão do MYRF impede a geração de novos OLs e prejudica o aprendizado motor. Além disso, distúrbios do desenvolvimento na produção de mielina ou comprometimento das bainhas de mielina são características típicas de doenças neurodegenerativas em humanos (CANTONE *et al.*, 2019), como esclerose múltipla (COMPSTON & COLES, 2008), leucodistrofia (AUBOURG, 1993) e o já citado defeito na capacidade de aprendizado motor (MCKENZIE *et al.*, 2014).

O MYRF pode controlar a expressão dos genes da mielina diretamente, ligando-se aos seus promotores ou, alternativamente, pode induzir indiretamente a transição para um estado

mielinizante, bloqueando a expressão de outros fatores de transcrição. Portanto, é possível que o MYRF induza uma coorte de genes durante a maturação do OL que trabalham juntos para montar e envolver as bainhas de mielina do SNC. (EMERY *et al.*, 2009) Recentemente, Kurahashi *et al.* (2018) mostraram que a introdução de uma única substituição num resíduo altamente conservado no sítio de ligação à DNA do MYRF resultou em uma diminuição significativa na atividade transcricional da região N-terminal.

O MYRF pertence a uma classe de fatores de transcrição ligados à membrana (MTFs), que são gerados inicialmente como proteínas integrais da membrana. Mediante sinalizações, elas passam por processamento autoproteolítico, e o trímero N-terminal, contendo o domínio de ligação ao DNA, é liberado da membrana do retículo endoplasmático permitindo a sua translocação para o núcleo onde funciona como um fator de transcrição controlando a expressão gênica e ativando a expressão dos genes da mielina (LI *et al.*, 2013).

Geralmente, fatores de transcrição interagem com o DNA como monômeros, dímeros ou tetrâmeros, enquanto fatores de transcrição homo-triméricos são raros (ZHEN *et al.*, 2017). O trímero N-terminal do MYRF pode ser dividido em três partes - o domínio mais N-terminal (NTM), o sítio de ligação ao DNA e o motivo da homo-trimerização. Dado esse arranjo de domínio e o fato de a região N-terminal de MYRF funcionar como um ativador de transcrição é provável que o domínio NTM funcione como um domínio de transativação. Choi *et al.* (2018) elucidaram esse domínio de transativação, descobrindo que o NTM do MYRF é necessário para a atividade transcricional. Além disso, mostrou que a homo-trimerização não é crítica para a função de transativação do domínio NTM diferentemente dos domínios de ligação ao DNA, cuja função depende da homo-trimerização.

Kim *et al.* (2017) descobriram que a homo-trimerização é crítica para a ligação ao DNA, por conferir especificidade de transcrição ao distinguir entre pontos genuínos ou não, e por permitir uma ligação firme aos locais de destino. Os autores especularam que a especificidade incomum conferida pela homo-trimerização é o que distingue MYRF de outros fatores de transcrição do OL, que são muito mais simples em comparação e se ligam a muito mais regiões genômicas e que isso possa ser devido a esse aumento da especificidade de interação entre os fragmentos N-terminais de MYRF e o DNA.

Bujalka *et al.* (2013) mostraram que o MYRF se liga aos elementos cis-regulatórios de vários genes específicos de oligodendrócitos envolvidos na mielinização estimulando sua expressão, sendo muitos desses locais de ligação *enhancers* anteriormente não caracterizados para esses genes da mielina.

O MYRF também pode cooperar com outros fatores de transcrição para controlar a expressão do gene da mielina nos oligodendrócitos, ao induzir uma rede de genes que trabalham juntos para envolver as bainhas de mielina do axônio no sistema nervoso central. MYRF possui a região de transativação que pode interagir diretamente com outros fatores transcripcionais, o homo-trímero MYRF pode ser estabilizado ou desestabilizado por essas interações para regular a expressão dos genes da mielina (ZHEN *et al.*, 2017).

A ligação do MYRF ao DNA ocorre por meio de uma sequência consenso CTGGYAC (onde Y = C ou T), identificada primeiramente por Bujalka *et al.* (2013), no entanto, a sequência CTGG[C/T]AC não condizia com uma ligação homo-trimérica. Em 2017, Kim *et al.* descobriram um novo motivo de DNA homo-trimérico, o qual incorpora o motivo anterior descoberto por Bujalka e que apoia fortemente a ideia de que os fragmentos N-terminais do MYRF se ligam ao DNA como homo-trímeros.

Miller *et al.* (2012) constataram que a mielinização neocortical (que inclui o córtex pré-frontal) é prolongada no desenvolvimento em humanos em comparação com os chimpanzés. Nos chimpanzés, a densidade dos axônios mielinizados aumentou de maneira constante até a idade adulta. Ao mesmo tempo, humanos exibiram mielinização mais lenta durante a infância, caracterizada por um período de maturação tardio que se estendeu até o final da adolescência. Smaers *et al.* (2017) ao estudar o córtex pré-frontal dos humanos, comparativamente com outros primatas, concluíram que o *H. sapiens* apresenta uma excepcional expansão do pré-frontal, provavelmente devido aos processos diferenciais de mielinização. Além disso, os autores também encontraram um aumento dessa região que desvia significativamente do esperado pela alometria nos grandes macacos (chimpanzé e gorila). Os autores inferem que as prováveis causas para tal expansão dessa região sejam as mesmas em humanos e grandes macacos, ou seja, mudanças neurodesenvolvimentais associadas com a reorganização cortical numa escala macroevolutiva. Neste contexto, nossos resultados poderiam aportar uma base molecular para dar subsídio a essa hipótese.

5.2. CAMTA1

Nas células de mamíferos, vários fatores estão envolvidos na regulação da transcrição de genes, como o Ca^{2+} que desempenha um papel fundamental na regulação da transcrição de genes, sendo um dos mais versáteis mensageiros na ativação ou inibição da transcrição (BOUCHÉ *et al.*, 2002).

Os mecanismos de regulação da transcrição dependente de Ca^{2+} são numerosos e vão desde a ligação e regulação direta dos fatores de transcrição, até vários transdutores de sinal, como a superfamília de proteínas de ligação *EF-hand*, a qual inclui a calmodulina, que regulam indiretamente a transcrição através da fosforilação e desfosforilação dos fatores de transcrição em resposta a um aumento de Ca^{2+} na célula (BOUCHÉ *et al.*, 2002; IKURA *et al.*, 2002; FINKLER *et al.*, 2007).

O ativador de transcrição de ligação à calmodulina 1 (CAMTA1) é um membro da família de ativadores de transcrição com ligação à calmodulina (CAMTAs) que se mostra conservado em uma ampla gama de eucariotos multicelulares e está envolvido nas funções cognitivas (HUENTELMAN *et al.*, 2007; THEVENON *et al.*, 2012). O gene *CAMTA1* foi identificado pela primeira vez em um estudo com cérebro humano, onde é altamente expresso (HUENTELMAN *et al.*, 2007).

Pouco se sabe sobre a maquinaria molecular envolvida nos intrincados processos que permitem a formação e recuperação de memórias e os polimorfismos genéticos correspondentes na população humana que levam ao desempenho diferencial da memória, mas Huentelman *et al.* em seu estudo de 2007, identificaram SNPs no gene *CAMTA1* que foram significativamente associados ao desempenho da memória, encontrando um enriquecimento da transcrição do *CAMTA1* em regiões cerebrais humanas relacionadas à memória episódica.

Memória episódica é um sistema neuro-cognitivo que permite ao ser humano se lembrar de experiências passadas, envolvendo a recuperação consciente dessas experiências anteriores (TULVING, 2002). As regiões do lobo temporal medial, incluindo o hipocampo, geralmente são consideradas como desempenhadoras de papéis críticos na formação e recuperação da memória episódica, e a patologia nessas regiões (por exemplo, em pacientes com doença de Alzheimer) está tipicamente associada à perda de memória (FRANKLAND & BONTEMP, 2005).

Polimorfismos de *CAMTA1* em seres humanos se correlacionam com níveis de desempenho em testes psicológicos (MILLER *et al.*, 2011), e Thevenon *et al.* (2012) obtiveram evidências que a perda de função de *CAMTA1* em humanos é responsável por ataxia, com ou sem déficit intelectual, e identificou uma série de manifestações em pacientes adultos, como falta de processamento na memória visual episódica, resultados que convergem com os de desempenho diferencial de memória episódica descritos por Huentelman *et al.* (2007).

Em camundongos, o *knockdown* de *CAMTA1* leva a comprometimentos na formação da memória de longo prazo (BAS-ORTH *et al.*, 2016), e causa severa ataxia acompanhada de degeneração das células de Purkinje, diminuição do tamanho cerebelar e interrupção da expressão de genes neuronais (LONG *et al.*, 2014).

O estudo de Bas-Orth *et al.* (2016) foi baseado na hipótese de que os CAMTAs, como ativadores transcricionais, regulam a transcrição de genes induzidos pela aprendizagem para controlar a formação da memória de longo prazo. De acordo com isso, seus experimentos comportamentais forneceram evidências robustas do papel do CAMTA1 como um regulador na formação da memória de longo prazo ao regular um conjunto de genes que controlam a transmissão sináptica e a excitabilidade neuronal.

Todos os CAMTAs compartilham uma organização de domínio semelhante, com um tipo de sítio de ligação ao DNA específico da sequência (designado CG-1). Esse domínio pode se ligar diretamente ao DNA e ativar a transcrição, ou interagir com outros fatores de transcrição agindo assim como um co-ativador da transcrição (FINKLER *et al.*, 2007). Os resultados de Long *et al.*, (2014) sugeriram que a sequência de consenso de ligação ao DNA do CAMTA1 poderia ser (T/C)GCANTGCG, em que "N" representa qualquer nucleotídeo.

Tulving (2002) um dos pioneiros no estudo da memória episódica humana, acredita que não há evidências de que animais não-humanos, inclusive outros primatas, consigam pensar no que chamamos de tempo subjetivo. A memória episódica se diferencia de outros sistemas de memória, como memória semântica (memória para fatos) ou memória lexical (memória para palavras) porque envolve essa experiência subjetiva de mentalmente viajar no tempo, possuindo uma autoconsciência do próprio passado, permitindo que se reexperimente, através dessa autoconsciência, as próprias experiências anteriores. A evidência neuropsicológica também apoia a ideia de memória episódica como um sistema distinto e Nyberg *et al.* (1996) e Tulving (1989) mostraram que diferentes áreas do cérebro estão ativas durante a recuperação da memória semântica e a recuperação da memória episódica.

Outros autores como Schwartz & Evans (2001) tentaram buscar evidências que apoiassem a ideia de que outros primatas não-humanos possuem memória episódica, e argumentam que os grandes primatas possuem uma memória episódica-like. Essa é uma questão particularmente complicada, devido à memória episódica ser pensada como um sistema neuro-cognitivo distinto que armazena eventos e fornece sua recuperação, mas sua assinatura são as experiências mentais que acompanham essa recuperação, a saber,

sentimentos de passado e sentimentos de confiança. Além disso, a maneira usual de avaliar a memória episódica em humanos é com testes verbais.

Lewis *et al.* (2019) mostrou que grandes primatas possuem memórias notáveis. Os autores demonstraram que esses primatas se lembram de eventos passados quando sondados por recursos que são diagnósticos do mesmo, consistentes com a recuperação involuntária de memória em humanos. Porém os autores reconhecem as limitações do trabalho, e que não podem afirmar se os macacos lembram conscientemente um evento passado (porque eles não fornecem descrições verbais). No entanto, as descobertas apontam para um reavivamento comportamental de uma memória (uma reencenação de um evento passado) que é desencadeada de maneira consistente com as memórias involuntárias em humanos. Esses achados parecem indicar importantes semelhanças entre memórias de eventos passados em macacos e humanos, sugerindo que a capacidade de recordar espontaneamente eventos passados pode ser compartilhada entre espécies.

Porém essas descobertas não vão totalmente contra os argumentos de Tulving (2002) que diz que animais não-humanos têm mentes, estão conscientes de seu mundo e confiam tanto em aprendizado e memória para adquirir as habilidades necessárias para a sobrevivência quanto nós, mas que por mais semelhanças que existam, não parecem ter o mesmo tipo de capacidade que os humanos têm para viajar de volta no tempo em suas próprias mentes, provavelmente porque não precisam. Tulving (2002) sugere que o senso de tempo subjetivo não é uma necessidade biológica, e que se os humanos o têm, é uma novidade evolutiva, necessário para viagens mentais no tempo. Sem noção de tempo subjetivo, sem viagem mental no tempo. Com isso, Tulving (2002) sugere que a memória episódica é um sistema recentemente desenvolvido, com desenvolvimento tardio e deterioração precoce, mais vulnerável do que outros sistemas de memória à disfunção neuronal e provavelmente exclusiva para os seres humanos.

Nesse contexto nossos resultados estão de acordo com as ideias de Tulving. Como mostrado aqui, o *CAMTA1* está associado à memória episódica em humanos, e o ganho de dois TFBS para o fator de transcrição CAMTA1 no *enhancer* HACNS1 de humanos apoia a hipótese de a memória episódica ser uma novidade evolutiva em humanos.

5.3. PTX1

Genes *homeobox* codificam fatores de transcrição que são essenciais para um funcionamento correto do organismo (LANCTÔT *et al.*, 1997). Proteínas *homeobox* são uma classe de proteínas conservadas evolutivamente que são essenciais no desenvolvimento de organismos adultos. Embora essas proteínas possuam especificidades de ligação ao DNA similares, suas funções são altamente diversas e dependentes de contexto (MIU *et al.*, 2019). PTX1 é membro de uma família de fatores de transcrição *homeobox* que possuem especificidade de ligação ao DNA (LANCTÔT *et al.*, 1997), tendo sido primeiramente identificado como um forte ativador de muitos promotores específicos da glândula pituitária (TREMBLAY *et al.*, 2000) possuindo um grande número de genes alvos *downstream* (LANCTÔT *et al.*, 1997).

Membros anteriores e posteriores de vertebrados possuem a mesma origem embrionária, sendo estruturas homólogas seriais. Embora os brotos dos quais eles derivam sejam modelados por sinais comuns durante o desenvolvimento embrionário, eles acabam formando estruturas morfologicamente distintas. Uma questão que surge disso é como as células expostas a sinais comuns podem responder diferentemente e dar origem a morfologias distintas (MINGUILLON *et al.*, 2005). Além disso, membros de vertebrados exibem uma variação estrutural extensa em animais adaptados a diferentes ambientes. Em um conjunto filogeneticamente diverso de vertebrados, incluindo as linhagens de répteis, anfíbios, mamíferos marinhos e peixes, os membros posteriores são reduzidos ou ausentes. O aparecimento de distintos membros posteriores no decorrer da evolução provavelmente requereu um novo conjunto de sinais e fatores de transcrição para diferenciar os membros posteriores dos anteriores (SHAPIRO *et al.*, 2004).

Poucos genes possuem expressão específica nos membros: a expressão precoce de *PTX1* em todo o mesoderma da placa lateral da metade posterior do embrião resulta na sua expressão nos membros posteriores, mas em nenhum momento *PTX1* é expresso no mesênquima dos membros anteriores ou nos membros anteriores em si. Essa expressão específica nos membros posteriores foi o que primeiro sugeriu que *PTX1* podia estar envolvido na determinação dos mesmos (LANCTÔT *et al.*, 1997).

Experimentos funcionais com camundongos deficientes em *PTX1* deram origem a indivíduos severamente afetados em seus membros posteriores e mandíbula, evidenciando uma atuação na identidade dos mesmos (LANCTÔT *et al.*, 1999). Camundongos com

knockout de *PTX1* são prontamente reconhecidos pelo encurtamento da mandíbula, apresentando micrognatia severa, fenda palatina e língua bifurcada. Os membros anteriores são normais, mas os posteriores possuem truncamentos graves em diversas estruturas, como o *ilium* ausente, tíbia menor, e os fêmures têm uma grande redução de tamanho, mostrando que esse fator de transcrição possui papéis fundamentais no crescimento, direcionamento e especificação da morfologia dos membros posteriores de vertebrados sendo expresso durante o desenvolvimento dos mesmos (LANCTÔT *et al.*, 1999; WANG *et al.*, 2018).

Mutações no *PTX1* afetam as estruturas mandibulares cujas funções tiveram mudanças nas articulações durante a transição de uma mandíbula de réptil para mamífero. Os defeitos nos membros posteriores podem também exemplificar uma perda de função adquirida durante a evolução dos membros. Alguns fósseis de cobras e algumas pítons atuais possuem membros vestigiais que lembram aqueles dos camundongos com *knockout* de *PTX1*. A associação de membros posteriores e deficiências mandibulares em camundongos *knockout* para o *PTX1* pode refletir um paralelo evolutivo dessas estruturas (LANCTÔT *et al.*, 1999).

O papel do *PITX1* na identidade dos membros posteriores foi posteriormente apoiado por experimentos de inativação de genes em camundongos que resultaram em membros posteriores mostrando características de membros anteriores (MARCIL *et al.*, 2003). Esses estudos foram apoiados por experimentos de ganho de função em embriões de galinha que resultou em asas com feições parciais de pernas (LOGAN & TABIN, 1999; MARCIL *et al.*, 2003) enquanto nos experimentos de inativação em camundongos o resultado foi a perda de características específicas de membros posteriores e sua substituição por características remanescentes de membros anteriores, mostrando que a expressão total do *PTX1* parecer ser o parâmetro mais importante no crescimento do broto dos membros posteriores (LANCTÔT *et al.*, 1999; MARCIL *et al.*, 2003). O trabalho de Marcil *et al.* (2003) e Minguillon *et al.* (2005) indicaram que o *PTX1* é um fator chave na regulação da identidade dos membros posteriores, possuindo papel essencial na padronização, crescimento e especificação da morfologia dessas estruturas, e também identificando o *PTX1* como o gene mais *upstream* em uma cascata de genes de especificação para identidade dos membros posteriores (MARCIL *et al.*, 2003).

Em 2004, Shapiro *et al.* identificaram uma grande região cromossômica controlando a perda de estruturas pélvicas em populações naturais de *Gasterosteus aculeatus*, um peixe conhecido pelo nome de esgana-gato. Embora o gene *PTX1* desses esgana-gatos com redução pélvica não mostre alterações no código de proteínas, em comparação com o de peixes marinhos ancestrais, sua expressão na região pélvica em desenvolvimento é quase

completamente perdida. Chan *et al.* (2010) mostraram que essa perda pélvica ocorreu devido a mutações em áreas reguladoras que excluíram um *enhancer* tecido-específico do *PTX1*.

Domyan *et al.* (2016) observaram que o *PTX1* é expresso em níveis mais baixos no desenvolvimento das pernas de raças de pombos com pés emplumados do que em uma raça com escamas nos pés. Em aves, plumas são uma característica de membros anteriores e escamas uma característica de membros posteriores, demonstrando que o desenvolvimento dos membros posteriores é sensível à dosagem de *PTX1*.

Esses estudos ilustram como grandes mudanças morfológicas podem surgir em populações naturais a partir de saltos mutacionais únicos, por meio de alterações regulatórias em um gene-chave de controle do desenvolvimento (CHAN *et al.*, 2010). Portanto, *PTX1* não é apenas importante para a formação dos membros posteriores, mas alterações na expressão de *PTX1* nos membros posteriores contribuíram para a evolução de diferenças na morfologia dos mesmos (WANG *et al.*, 2018).

Análises comparativas de sequências de genoma de camundongo e humano mostraram que o gene *PTX1* é flanqueado por grandes regiões de possíveis sequências reguladoras, incluindo mais de 300 kb de DNA flanqueador que é altamente enriquecido em sequências não codificadoras conservadas. Essas sequências provavelmente incluem múltiplos elementos reguladores que controlam a expressão do gene *PTX1* em diferentes tecidos (SHAPIRO *et al.*, 2004). A sequência do *PTX1* possui três domínios ricos em prolina e dois ricos em serina que funcionam como domínios trans-ativadores (LAMONERIE *et al.*, 1996). *PTX1* atua como um ativador transcricional ao se ligar a sequências com motivos TAA(T/G)CC que está presente na região promotora de diversos genes alvos (LANCTOT *et al.*, 1997).

Os membros dos primatas são morfologicamente diversos, e grande parte dessa variação na anatomia musculoesquelética se correlaciona com a função locomotora. Em *Pan*, os membros posteriores servem funções em diferentes tipos de locomoção (por exemplo, bipedalismo terrestre, escalada vertical) e em atividades manipulativas, incluindo agarrar galhos. Por outro lado, os membros posteriores de humanos estão envolvidos principalmente na propulsão terrestre e perderam a capacidade de desempenhar funções como agarrar (DIOGO *et al.*, 2018).

As capacidades locomotoras são determinadas por diversos fatores, incluindo a morfologia do esqueleto ósseo e as propriedades da unidade músculo-tendão (PAYNE *et al.*, 2006). O *PTX1* é um fator de transcrição expresso durante o desenvolvimento dos membros posteriores, onde tem um papel crítico direcionando o seu crescimento. Ele não é importante

apenas para a formação, mas mudanças na expressão do PTX1 provavelmente contribuíram para a evolução de diferenças em morfologias, e a perda de um motivo de ligação do PTX1 no HACNS1 de humanos, um *enhancer* ativo no desenvolvimento pode levar a justamente mudanças na expressão do gene e conseqüentemente a diferenças na morfologia dos membros posteriores de humanos.

6. CONCLUSÃO

A origem do ser humano foi acompanhada pelo surgimento de novas funções comportamentais e cognitivas, como a linguagem e formas especializadas de representação abstrata e uma forma diferencial de locomoção, o bipedalismo (WANG *et al.*, 2012), e de destreza manual, o polegar humano que possibilitou que humanos desenvolvessem movimentos finos. As mudanças evolutivas na expressão gênica que levaram a uma progressão em direção a um maior repertório cognitivo em humanos podem ter sido causadas por alterações nas regiões reguladoras do DNA, como promotores e *enhancers* (HARDINGHAM *et al.*, 2018).

Embora a ordem dos fatores causais ainda não esteja totalmente esclarecida, existe um consenso geral de que a evolução das capacidades manipulativas e do bipedalismo em humanos estão inter-relacionados. Neste estudo, encontramos evidências moleculares da perda do motivo PTX1 em humanos, associado em diversos estudos com mudanças nos membros posteriores de mamíferos. As mãos liberadas de uma função na locomoção podem ter proporcionado uma oportunidade para a evolução do aumento da destreza manipulativa, mas essa liberação da função locomotora não seria suficiente por si só. Outras mudanças no controle neural dos movimentos da mão também seriam necessárias (ALMÉCIJA & SHERWOOD, 2017).

Estudos que compararam perfis de expressão gênica no córtex cerebral de humanos e chimpanzés encontraram um padrão no cérebro esquerdo humano distinto do de primatas não humanos. Os resultados indicaram que muitas mudanças na expressão dos genes de função no córtex humano envolviam predominantemente aumento da expressão, e que muitos dos genes regulados em humanos podem estar relacionados a níveis mais altos de atividade neuronal (CÁCERES *et al.*, 2003). Nesse estudo identificamos o ganho de motivos para o CAMTA1, um TFBS relacionado com funções cognitivas, e o MYRF, que está envolvido na mielinização do sistema nervoso central.

Estudos anteriores sugerem que o HACNS1 age como um robusto *enhancer* durante o desenvolvimento e ganhos de motivos sugerem um aumento de expressão desses genes durante o desenvolvimento. O HACNS1 também foi previamente relacionado com a formação do polegar em humano (PRABHAKAR *et al.*, 2008), característica que provê uma alta destreza manual, e que também está envolvida com extensa ativação em ambos os hemisférios cerebrais e conseqüentemente uma maior cognição (ALMÉCIJA &

SHERWOOD, 2017). Hünemeier *et al.* (2010) mostrou que todos os 13 sítios do HACNS1 possuem sinais de seleção positiva em humanos, o que sugere uma grande importância no que está sendo regulado por esse *enhancer*.

Considerando o exposto, além da destreza manual, nós sugerimos um papel do HACNS1 na evolução do bipedalismo e das funções cognitivas humanas. É a tríade que separa os humanos dos outros grandes primatas, e são características que estão intimamente relacionadas entre si: maior destreza, bipedalismo e grandes habilidades cognitivas.

7. RESUMO

Análises moleculares sempre foram um componente importante no estudo da evolução dos primatas. Elementos cis-regulatórios (CREs) são sequências de DNA não codificantes que contêm sítios de ligação para fatores de transcrição ou outras moléculas reguladoras necessárias para ativar e manter a transcrição, sendo os *enhancers* os CREs que, quando ligados a um fator de transcrição, aumentam a transcrição de um gene associado. O HACNS1 é um *enhancer* altamente conservado no genoma de todos os vertebrados terrestres, mas que acumulou 16 substituições humano-específicas desde a divergência dos chimpanzés, com 13 dessas 16 mutações sendo encontradas dentro de um segmento de 81-pb dessa região e tendo um papel específico nos seres humanos, possivelmente relacionado ao desenvolvimento da destreza das mãos. A divergência entre os macacos de velho e novo mundo ocorreu cerca de 45 a 37 milhões de anos atrás, e é possível que tenha acontecido outras mutações entre esses grupos que levaram ao desenvolvimento de outras características específicas. Dessa forma, este estudo objetiva a caracterização molecular do *enhancer* HACNS1 quanto à previsão da presença de sítios de ligação ao fator de transcrição (TFBS) em 22 espécies do novo mundo e 20 espécies de macacos do velho mundo. As sequências HACNS1 foram obtidas em 18 novos macacos do mundo por sequenciamento sanger, enquanto as sequências dos outros macacos e do *Homo sapiens* foram obtidas do banco de dados NCBI. A previsão do TFBS foi feita através do software Genomatix. A análise computacional dos TFBS previsto nos ortólogos do HACNS1 sugeriu que vários locais foram ganhos e perdidos nesse *enhancer* durante a evolução dos primatas. Os principais ganhos previstos no TFBS incluem alguns ganhos específicos para o ser humano, como o CAMTA1, um fator de transcrição de ligação à calmodulina associado ao desempenho da memória episódica e cuja ausência está relacionada à deficiência intelectual. Outra descoberta foi o ganho do TFBS para o fator de transcrição MYRF em grandes primatas. MYRF é um fator regulador da mielina que participa da regulação da mielinização do sistema nervoso central. Como o HACNS1 está relacionado à destreza manual, o ganho deste TFBS também pode estar relacionado ao mesmo processo. Em relação às perdas, encontramos a perda de uma cópia do PTX1 em humanos, um fator de transcrição que é expresso durante o desenvolvimento dos membros posteriores e tem um papel crítico no direcionamento do crescimento e na especificação da morfologia dos mesmos.

8. ABSTRACT

Molecular analyses were always an important component when studying primate evolution. Cis-regulatory elements (CREs) are non-coding DNA sequences containing binding sites for transcription factors or other regulatory molecules that are needed to activate and sustain transcription, with enhancers being the ones CREs that when binded to a transcription factor increase the transcription of an associated gene. HACNS1 is a highly constrained enhancer in the genome of all terrestrial vertebrates, but it has accumulated 16 human-specific changes after the human chimpanzee split, with 13 of these 16 mutations being found within an 81-bp segment of this region and having a specific role in humans, possibly related to the development of hand dexterity. The divergence between old and new world monkeys happened about 45 to 37 million of years ago, and is possible that there were other mutations between these groups that lead to the development of other specific characteristics. Hereby, this study aims at the molecular characterization of the HACNS1 enhancer as to the presence of predicted transcription factor binding sites (TFBS) in 22 species of new world and 20 species of old world monkeys. The HACNS1 sequences were obtained in 18 new world monkeys through sanger sequencing, whereas the sequences from other monkeys and from *Homo sapiens* were obtained from NCBI database. Prediction of TFBS was made through the software Genomatix. Computational analysis of predicted TFBS in HACNS1 orthologs suggested that multiple sites have been gained and lost in this enhancer during primate evolution. Main predicted TFBS gains include some human-specific gains such as CAMTA1, a calmodulin-binding transcription factor associated with episodic memory performance and whose absence is related to intellectual disability. Another finding was the gain of MYRF TFBS in great apes. MYRF is a myelin regulatory factor that participates in the regulation of the myelination of central nervous system. Since HACNS1 is related to hand dexterity, the gain of MYRF TFBS can also be related to the same process. Regarding losses, we found the loss of one copy of PTX1 in humans, a transcription factor that is expressed during hindlimb development and has a critical role in directing hindlimb growth and the specification of hindlimb morphology.

7. REFERÊNCIAS BIBLIOGRÁFICAS

- ALMÉCIJA, S.; SHERWOOD, C. C. **Hands, Brains, and Precision Grips: Origins of Tool Use Behaviors**. Second Ed, Elsevier, 2017.
- AUBOURG, P. The leukodystrophies: A window to myelin. **Nature Genetics**, v. 5, n. 2, p. 105–106, 1993.
- BAILEY, T. L. *et al.* MEME: discovering and analyzing DNA and protein sequence motifs. **Nucleic Acids Research**, v. 34, p. W369–W373, 2006.
- BALL, P. DNA: Celebrate the unknowns. **Nature**, v. 496, n. 7446, p. 419–420, 2013.
- BARNES, G. L. *et al.* Osteoblast-related transcription factors Runx2 (Cbfa1/AML3) and MSX2 mediate the expression of bone sialoprotein in human metastatic breast cancer cells 1. **Cancer Research**, v. 63, p. 2631–2637, 2003.
- BAS-ORTH, C. *et al.* The calmodulin-binding transcription activator CAMTA1 is required for long-term memory formation in mice. **Learning & Memory**, v. 23, n. 6, p. 313–21, 2016.
- BEARD, K. C. The phylogenetic significance of strepsirrhinism in Paleogene primates. **International Journal of Primatology**, v. 9, n. 2, p. 83–96, 1988.
- BOUCHÉ, N. *et al.* A novel family of calmodulin-binding transcription activators in multicellular organisms. **The Journal of Biological Chemistry**, v. 277, n. 24, p. 21851–61, 2002.
- BUJALKA, H. *et al.* MYRF is a membrane-associated transcription factor that autoproteolytically cleaves to directly activate myelin genes. **PLoS Biology**, v. 11, n. 8, p. 1–15, 2013.
- BULYK, M. L. Computational prediction of transcription-factor binding site locations. **Genome Biology**, v. 5, n. 201, 2003.
- CÁCERES, M. *et al.* Elevated gene expression levels distinguish human from non-human primate brains. **PNAS**, v. 100, n. 22, p. 13030–13035, 2003.
- CANTONE, M. *et al.* A gene regulatory architecture that controls region-independent dynamics of oligodendrocyte differentiation. **Glia**, v. 67, n. 5, p. 825–843, 2019.
- CARROLL, S. B. Genetics and the making of *Homo sapiens*. **Nature**, v. 422, n. 6934, p. 849–857, 2003.
- CARTHARIUS, K. *et al.* MatInspector and beyond: promoter analysis based on transcription factor binding sites. **Bioinformatics**, v. 21, n. 13, p. 2933–2942, 2005.
- CHAN, Y. F. *et al.* Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a Pitxl enhancer. **Science**, v. 327, n. 5963, p. 302–305, 2010.

CHAUHAN, S. *et al.* ZKSCAN3 is a master transcriptional repressor of autophagy. **Molecular Cell**, v. 50, n. 1, p. 16–28, 2013.

CHI, Y. *et al.* ZKSCAN3 promotes breast cancer cell proliferation, migration and invasion. **Biochemical and Biophysical Research Communications**, v. 503, n. 4, p. 2583–2589, 2018.

CHOI, J. *et al.* Elucidating the transactivation domain of the pleiotropic transcription factor Myrf. **Scientific Reports**, v. 8, p. 13075, 2018.

COMPSTON, A.; COLES, A. Multiple sclerosis. **The Lancet**, v. 372, p. 1502–1517, 2008.

DAN, J. *et al.* Zscan4 inhibits maintenance DNA methylation to facilitate telomere elongation in mouse embryonic stem cells. **Cell reports**, v. 20, n. 8, p. 1936–1949, 2017.

DIOGO, R. *et al.* First anatomical network analysis of fore-and hindlimb musculoskeletal modularity in bonobos, common chimpanzees, and human. **Scientific Reports**, v. 8, n. 6885, p. 1–10, 2018.

DOMYAN, E. T. *et al.* Molecular shifts in limb identity underlie development of feathered feet in two domestic avian species. **eLife**, v. 5, p. 1–21, 2016.

EMERY, B. *et al.* Myelin gene regulatory factor is a critical transcriptional regulator required for CNS myelination. **Cell**, v. 138, n. 1, p. 172–185, 2009.

ENARD, W. *et al.* Molecular evolution of FOXP2, a gene involved in speech and language. **Nature**, v. 418, n. 6900, p. 869–872, 2002.

FINKLER, A. *et al.* CAMTAs: Calmodulin-binding transcription activators from plants to human. **FEBS Letters**, v. 581, n. 21, p. 3893–3898, 2007.

FRANCHINI, L. F.; POLLARD, K. S. Human evolution: The non-coding revolution. **BMC Biology**, v. 15, n. 89, 2017.

FRANKEL, N. S. *et al.* Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. **Nature**, v. 474, n. 7353, p. 598–603, 2011.

FRANKLAND, P. W.; BONTEMP, B. The organization of recent and remote memories. **Nature Reviews**, v. 6, p. 119–130, 2005.

FUTUYMA, D. J. **Evolutionary Biology**. 3rd Ed. ed. Sunderland, Massachusetts: Sinauer Associates, Inc., 1998.

GILBERT, L. *et al.* Expression of the osteoblast differentiation factor RUNX2 (Cbfa1/AML3/Pebp2A) is inhibited by tumor necrosis factor. **The Journal of Biological Chemistry**, v. 277, n. 4, p. 2695–2701, 2002.

GOODMAN, M. Serological analysis of the systematics of recent hominoids. **Human Biology**, v. 35, n. 3, p. 377–436, 1963.

GOLDMAN, N.; YANG, Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. **Molecular Biology and Evolution**, v. 11, n. 5, p. 725–736, 1994.

GRAUR, D.; LI, W. H. **Fundamentals of Molecular Evolution**. Second ed. Sunderland, Massachusetts: Sinauer Associates, Inc., 2000.

GU, J.; GU, X. Induced gene expression in human brain after the split from chimpanzee. **Trends in Genetics**, v. 19, n. 2, p. 63–65, 2003.

HAGEN, J. B. Waiting for Sequences: Morris Goodman, Immunodiffusion Experiments, and the Origins of Molecular Anthropology. **Journal of the History of Biology**, v. 43, n. 4, p. 697–725, 2010.

HAKIM, A.; USMANI, O. S. Transcription Factors. In: **Reference Module in Biomedical Sciences**. Elsevier, 2014.

HALL, T. A. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. **Nucleic Acids Symposium Series**, v. 41, p. 95–98, 1997.

HARDINGHAM, G. E. *et al.* Lineage divergence of activity-driven transcription and evolution of cognitive ability. **Nature Reviews Neuroscience**, v. 19, p. 9–15, 2018.

HARTIG, G. *et al.* Retrophylogenomics place tarsiers on the evolutionary branch of anthropoids. **Scientific Reports**, v. 3, 2013.

HONEYCUTT, R. L. Small changes, big results: Evolution of morphological discontinuity in mammals. **Journal of Biology**, v. 7, n. 3, 2008.

HUENTELMAN, M. J. *et al.* Calmodulin-binding transcription activator 1 (CAMTA1) alleles predispose human episodic memory performance. **Human molecular genetics**, v. 16, n. 12, p. 1469–77, 2007.

HÜNEMEIER, T. *et al.* Brief communication: Population data support the adaptive nature of HACNS1 sapiens/neandertal-chimpanzee differences in a limb expression domain. **American Journal of Physical Anthropology**, v. 143, n. 3, p. 478–481, 2010.

HUNT, S. E. *et al.* Ensembl variation resources. **Database**, v. 2018, p. 1–12, 2018.

IKURA, M. *et al.* The role of calcium-binding proteins in the control of transcription: structure to function. **BioEssays**, v. 24, n. 7, p. 625–636, 2002.

JEONG, S. *et al.* The evolution of gene regulation underlies a morphological difference between two Drosophila sister species. **Cell**, v. 132, n. 5, p. 783–793, 2008.

JOBLING, M. A. *et al.* **Human Evolutionary Genetics: Origins, Peoples and Disease**. 2nd Ed. New York: Garland Science, 2014.

KIM, D. *et al.* Homo-trimerization is essential for the transcription factor function of Myrf for

oligodendrocyte differentiation. **Nucleic Acids Research**, v. 45, n. 9, p. 5112–5125, 2017.

KING, M. C.; WILSON, A. C. Evolution at two levels in humans and chimpanzees. **Science**, v. 188, n. 4184, p. 107-116, 1975.

KURAHASHI, H. *et al.* MYRF is associated with encephalopathy with reversible myelin vacuolization. **Annals of Neurology**, v. 83, n. 1, p. 98–106, 2018.

LAMBERT, S. A. *et al.* The Human Transcription Factors. **Cell**, v. 172, n. 4, p. 650–665, 2018.

LAMONERIE, T. *et al.* Ptx1, a bicoid-related homeo box transcription factor involved in transcribption of the pro-opiomelanocortin gene. **Genes & Development**, n. 10, p. 1284–1295, 1996.

LANCTÔT, C. *et al.* The bicoid-related homeoprotein Ptx1 defines the most anterior domain of the embryo and differentiates posterior from anterior lateral mesoderm. **Development**, v. 124, n. 14, 1997.

LANCTÔT, C. *et al.* Hindlimb patterning and mandible development require the Ptx1 gene. **Development**, v. 126, n. 9, p. 1805–1810, 1999.

LEVCHENKO, A. *et al.* Human accelerated regions and other human-specific sequence variations in the context of evolution and their relevance for brain development. **Genome Biology and Evolution**, v. 10, n. 1, p. 166–188, 2018.

LEWIS, A. *et al.* Long-term memory of past events in great apes. **Current Directions in Psychological Science**, v. 28, n. 2, p. 117–123, 2019.

LI, Z. *et al.* A bacteriophage tailspike domain promotes self-cleavage of a human membrane-bound transcription factor, the myelin regulatory factor MYRF. **PLoS Biology**, v. 11, n. 8, 2013.

LOGAN, M.; TABIN, C. J. Role of Pitx1 upstream of Tbx4 in specification of hindlimb identity. **Science**, v. 283, n. 5408, p. 1736–1739, 1999.

LONG, C. *et al.* Ataxia and Purkinje cell degeneration in mice lacking the CAMTA1 transcription factor. **Proceedings of the National Academy of Sciences of the United States of America**, v. 111, n. 31, p. 11521–6, 2014.

MANUKYAN, A. *et al.* Analysis of transcriptional activity by the Myt1 and Myt11 transcription factors. **Journal of Cellular Biochemistry**, v. 119, n. 6, p. 4644–4655, 2018.

MARCIL, A. *et al.* Pitx1 and Pitx2 are required for development of hindlimb buds. **Development**, v. 130, n. 1, p. 45–55, 2003.

MASTERS, J. Primate Evolution. In: MINELLI, A.; CONTRAFATTO, G. (Eds.). . **Biological Science Fundamentals and Systematics**. Oxford, UK: Eolss

Publishers/UNESCO, 2009. v. IVp. 346.

MASTON, G. A. *et al.* Transcriptional Regulatory Elements in the Human Genome. **Annual Review of Genomics and Human Genetics**, v. 7, n. 1, p. 29–59, 2006.

MCKENZIE, I. A. *et al.* Motor skill learning requires active central myelination. **Science**, v. 346, n. 6207, p. 318–322, 2014.

MCLEAY, R. C.; BAILEY, T. L. Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data. **BMC Bioinformatics**, v. 11, n. 1, p. 165, 2010.

MELHUISE, T. A. *et al.* Myt1 and Myt11 transcription factors limit proliferation in GBM cells by repressing YAP1 expression. **Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms**, v. 1861, n. 11, p. 983–995, 2018.

MILLER, D. J. *et al.* Prolonged myelination in human neocortical evolution. **PNAS**, v. 109, n. 41, p. 16480–16485, 2012.

MILLER, L. A. *et al.* CAMTA1 T polymorphism is associated with neuropsychological test performance in older adults with cardiovascular disease. **Psychogeriatrics**, v. 11, n. 3, p. 135–140, 2011.

MINGUILLON, C. *et al.* Tbx5 and Tbx4 are not sufficient to determine limb-specific morphologies but have common roles in initiating limb outgrowth. **Developmental Cell**, v. 8, n. 1, p. 75–84, 2005.

MITTERMEIER, R. A. *et al.* **Handbook of the Mammals of the World: Primates**. Barcelona, Spain: Lynx Ediciones, 2013.

MIU, K. K. *et al.* Genetics and genomics of early gonad development. In: **Human Reproductive and Prenatal Genetics**. Elsevier, 2019.

MORGENSTERN, B. *et al.* DIALIGN: Finding local similarities by multiple sequence alignment. **Bioinformatics**, v. 14, n. 3, p. 290–294, 1998.

NEI, M. The new mutation theory of phenotypic evolution. **Proceedings of the National Academy of Sciences**, v. 104, n. 30, p. 12235–12242, 2007.

NYBERG, L. *et al.* PET studies of encoding and retrieval: The HERA model. **Psychonomic Bulletin and Review**, v. 3, n. 2, p. 135–148, 1996.

PAYNE, R. C. *et al.* Morphological analysis of the hindlimb in apes and humans. I. Muscle architecture. **Journal of Anatomy**, v. 208, p. 709–724, 2006.

PENNACCHIO, L. A. *et al.* In vivo enhancer analysis of human conserved non-coding sequences. **Nature Letters**, v. 444, p. 499–502, 2006.

PENNACCHIO, L. A. *et al.* Enhancers: Five essential questions. **Nature Reviews Genetics**,

v. 14, n. 4, p. 288-295, 2013.

PERELMAN, P. *et al.* A molecular phylogeny of living primates. **PLoS Genetics**, v. 7, n. 3, 2011.

POPADIC, A. *et al.* Understanding the genetic basis of morphological evolution: The role of homeotic genes in the diversification of the arthropod bauplan. **International Journal of Developmental Biology**, v. 42, n. 3, p. 453–461, 1998.

POUX, C.; DOUZERY, E. J. P. Primate phylogeny, evolutionary rate variations, and divergence times: A contribution from the nuclear gene IRBP. **American Journal of Physical Anthropology**, v. 124, n. 1, p. 01–16, 2004.

PRABHAKAR, S. *et al.* Human-specific gain of function in a developmental enhancer. **Science**, v. 321, n. 5894, p. 1346–1350, 2008.

QUANDT, K. *et al.* MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. **Nucleic Acids Research**, v. 23, n. 23, p. 4878–4884, 1995.

RAAUM, R. L. *et al.* Catarrhine primate divergence dates estimated from complete mitochondrial genomes: Concordance with fossil and nuclear DNA evidence. **Journal of Human Evolution**, v. 48, n. 3, p. 237–257, 2005.

REIS, N. R. *et al.* **Primatas Brasileiros**. Londrina: Technical Books Edition, 2008.

SCHWARTZ, B. L.; EVANS, S. Episodic memory in primates. **American Journal of Primatology**, v. 55, n. 2, p. 71–85, 2001.

SHAPIRO, M. D. *et al.* Genetic and developmental basis of evolutionary pelvic reduction in threespine sticklebacks. **Nature**, v. 428, n. 6984, p. 717–723, 2004.

SMAERS, J. B. *et al.* Exceptional evolutionary expansion of prefrontal cortex in great apes and humans. **Current Biology**, v. 27, n. 5, p. 714–720, 2017.

SUMIYAMA, K.; SAITOU, N. Loss-of-function mutation in a repressor module of human-specifically activated enhancer HACNS1. **Molecular Biology and Evolution**, v. 28, n. 11, p. 3005–3007, 2011.

THEVENON, J. *et al.* Intragenic CAMTA1 rearrangements cause non-progressive congenital ataxia with or without intellectual disability. **Journal of Medical Genetics**, v. 49, p. 400–408, 2012.

TREMBLAY, J. J. *et al.* Transcriptional properties of Ptx1 and Ptx2 isoforms. **Neuroendocrinology**, v. 71, n. 5, p. 277–286, 2000.

TULVING, E. Memory: Performance, knowledge, and experience. **European Journal of**

- Cognitive Psychology**, v. 1, n. 1, p. 3–26, 1989.
- TULVING, E. EPISODIC MEMORY: From Mind to Brain. **Annual Review of Psychology**, v. 53, p. 1–25, 2002.
- VAN HIJUM, S. A. F. T. *et al.* Mechanisms and evolution of control logic in prokaryotic transcriptional regulation. **Microbiology And Molecular Biology Reviews**, v. 73, n. 3, p. 481–509, 2009.
- VAQUERIZAS, J. M. *et al.* A census of human transcription factors: Function, expression and evolution. **Nature Reviews Genetics**, v. 10, n. 4, p. 252–263, 2009.
- VEGA-BENEDETTI, A. F. *et al.* PLAGL1 gene function during hepatoma cells proliferation. **Oncotarget**, v. 9, n. 67, p. 32775–32794, 2018.
- VOGEL, C.; CHOTHIA, C. Protein family expansions and biological complexity. **PLoS Computational Biology**, v. 2, n. 5, p. 370–382, 2006.
- WANG, J. S. *et al.* PITX1 promotes chondrogenesis and myogenesis in mouse hindlimbs through conserved regulatory targets. **Developmental Biology**, v. 434, n. 1, p. 186–195, 2018.
- WANG, L. *et al.* Signal Transducer and Activator of Transcription 2 (STAT2) metabolism coupling postmitotic outgrowth to visual and sound perception network in human left cerebrum by biocomputation. **Journal of Molecular Neuroscience**, v. 47, n. 3, p. 649–658, 2012.
- WILLIAMS, T. M. *et al.* The regulation and evolution of a genetic switch controlling sexually dimorphic traits *Drosophila*. **Cell**, v. 134, n. 4, p. 610–623, 2008.
- WITTKOPP, P. J.; KALAY, G. Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. **Nature Reviews Genetics**, v. 13, p. 59–69, 2011.
- WONG, W. S. W.; NIELSEN, R. Detecting selection in noncoding regions of nucleotide sequences. **Genetics Society of America**, v. 167, n. 2, p. 949–958, 2004.
- WRAY, G. A. Evolution: Spot on (and off). **Nature**, v. 440, n. 7087, p. 1001–1002, 2006.
- YANG, L. *et al.* Evidence of a role for the novel zinc-finger transcription factor ZKSCAN3 in modulating Cyclin D2 expression in multiple myeloma. **Oncogene**, v. 30, n. 11, p. 1329–40, 2011.
- YANG, Z. *et al.* Bayes Empirical Bayes Inference of Amino Acid Sites Under Positive Selection. **Molecular Biology and Evolution**, v. 22, n. 4, p. 1107–1118, 2005.
- ZALZMAN, M. *et al.* Zscan4 regulates telomere elongation and genomic stability in ES cells.

Nature, v. 464, n. 7290, p. 858–863, 2010.

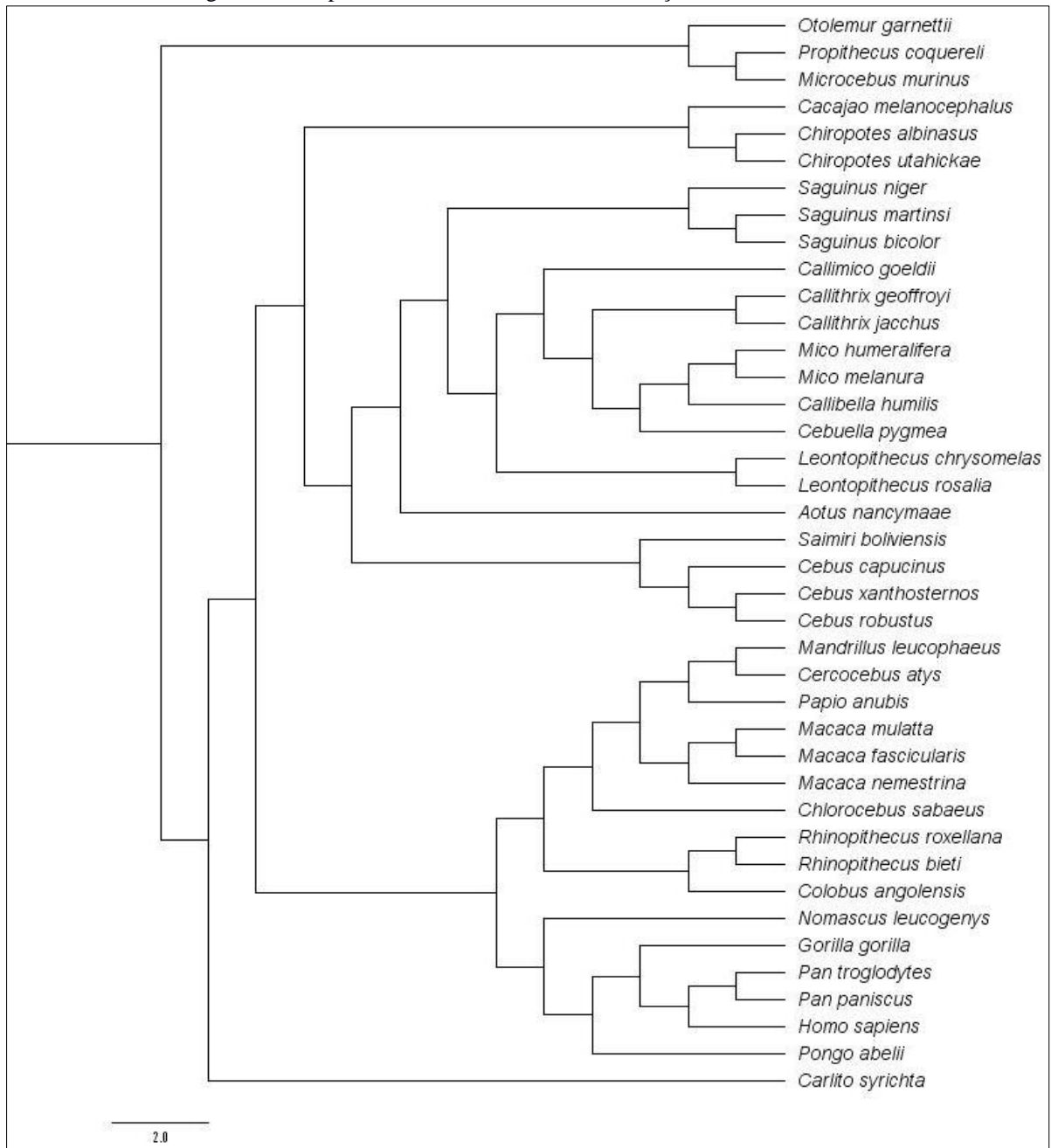
ZHANG, X. *et al.* The zinc finger transcription factor ZKSCAN3 promotes prostate cancer cell migration. **The International Journal Of Biochemistry & Cell Biology**, v. 44, n. 7, p. 1166–73, 2012.

ZHAO, Y. *et al.* The transcription factor RFX5 is a transcriptional activator of the TPP1 gene in hepatocellular carcinoma. **Oncology Reports**, v. 37, n. 1, p. 289–296, 2017.

ZHEN, X. *et al.* Crystal structure of the DNA-binding domain of Myelin-gene Regulatory Factor. **Scientific Reports**, v. 7, n. 3696, 2017.

8. ANEXOS

Anexo I – Árvore filogenética dos primatas utilizada na análise de seleção.



Fonte: Gabrielle A. Rizzato

Anexo II – Sequências completas do HACNS1 dos primatas de velho e novo mundo

Espécies	HACNS1
<i>Homo sapiens</i>	GGTGAGCTGACCTGTGTGTGGCGCAGCCTTGGGTTCCGCAAATAGGGCACCCACAGTAACACGCTGTGGCGCCGACCCCGCCGT GCGCAATCGGGGCTTTATACGATTGCTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTGCCCTGCAACCTGGCACA ACGACAG
<i>Pan troglodytes</i>	GGTGAGCTGACCTGTGTGTGGCACAGGCTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGCTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCACA ACGACAG
<i>Pan paniscus</i>	GGTGAGCTGACCTGTGTGTGGCACAGGCTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCTGCCGT GCATAATCGGGGCTTTATAGGATTGCTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCACA ACGACAG
<i>Gorilla gorilla</i>	GGTGAGCTGACCTGTGTGTGGCACAGGCTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGCTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCACA ACGACAG
<i>Pongo abelii</i>	GGTGAGCTGACCTGTGTGTGGCACAGGCTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGCTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTAGTACCCTGCAACCTGGCAGA ACGACAG
<i>Nomascus leucogenys</i>	GGTGAGCTGACCTGTGTGTGGCACAGGCTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGCTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ACGACAG
<i>Chlorocebus sabaues</i>	GGTGAACCTGACCTGTGTGTGGCACAGGCTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ACGACAG

<i>Macaca mulatta</i>	GGTGAACTGACCTGTGTGTGGCACAGGCTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ACGACAG
<i>Macaca fascicularis</i>	GGTGAACTGACCTGTGTGTGGCACAGGCTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ACGACAG
<i>Macaca nemestrina</i>	GGTGAACTGACCTGTGTGTGGCACAGGCTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ACGACAG
<i>Papio anubis</i>	GGTGAACTGACCTGTGTGTGGCACAGGCTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ACGACAG
<i>Rhinopithecus bieti</i>	GGTGAACTGACCTGTGTGTGGCACAGGCTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ACGACAG
<i>Rhinopithecus roxellana</i>	GGTGAACTGACCTGTGTGTGGCACAGGCTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ACGACAG
<i>Cercocebus atys</i>	GGTGAACTGACCTGTGTGTGGCACAGGCTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ACGACAG
<i>Colobus</i>	GGTGAACTGACCTGTGTGTGGCACAGGCTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT

<i>angolensis</i>	GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ACGACAG
<i>Mandrillus leucophaeus</i>	GGTGAACCTGACCTGTGTGTGGCACAGGCTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ACGACAG
<i>Propithecus coquereli</i>	GGTGAGCTGACCTGTGTGTGGCACAGGTTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTACAACCTGGCAGA ATGACAG
<i>Otolemur garnettii</i>	GGTGAGCTGACCTGTGTGTGGCACAGGTTTCGGGTTCCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTACAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTACAACCTGGCAGA ATGACAG
<i>Microcebus murinus</i>	GGTGAGCTGACCTGTGTGTGGCACAGGTTTGGGCTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTACAGGATTGCTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTACAACCTGGCAGA ATGACAG
<i>Carlito syrichta</i>	GGTGAGCTGACCTGTGTGTGGCACAGGTTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTACAACCTGGCAGA ATGACAG
<i>Callithrix jacchus</i>	GGTGAGCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ATGACAG
<i>Saimiri boliviensis</i>	GGTGAGCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA

	ATGACAG
<i>Aotus nancymae</i>	GGTGAGCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCACCCACAGTAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ATGACAG
<i>Cebus capucinus</i>	GGTGAGCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ATGACAG
<i>Cebus xanthosternos</i>	GGTGAGCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ATGACAG
<i>Cebus robustus</i>	GGTGAGCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ATGACAG
<i>Leontopithecus chrysomelas</i>	GGTGAGCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ATGACAG
<i>Leontopithecus rosalia</i>	GGTGAGCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ATGACAG
<i>Callimico goeldii</i>	GGTGAGCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ATGACAG

<i>Callibella humilis</i>	GGTGAGCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ATGACAG
<i>Mico melanura</i>	GGTGAGCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ATGACAG
<i>Mico humeralifera</i>	GGTGATCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ATGACAG
<i>Saguinus martinsi</i>	GGTGATCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ATGACAG
<i>Saguinus niger</i>	GGTGATCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ATGACAG
<i>Saguinus bicolor</i>	GGTGAGCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ATGACAG
<i>Cebuella pygmaea</i>	GGTGAGCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ATGACAG
<i>Callithrix</i>	GGTGAGCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT

<i>geoffroyi</i>	GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ATGACAG
<i>Chiropotes albinasus</i>	GGTGAGCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ATGACAG
<i>Chiropotes utahickae</i>	GGTGAGCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ATGACAG
<i>Cacajao melanocephalus</i>	GGTGAGCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ATGACAG
<i>Callicebus coimbrai</i>	GGTGAGCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ATGACAG
<i>Brachyteles arachnoides</i>	GGTGAGCTGACCTGTGTGTGGCACAGACTTGGGTTTCGCAAATAGGGCATCCACAATAACAAGTGTGTCACTAACCCCGCCGT GCATAATCGGGGCTTTATAGGATTGTTGGAGATGCTGACAGCTCAATTAAAGTGTAACCCTTTGTACCCTGCAACCTGGCAGA ATGACAG

Fonte: Gabrielle A. Rizzato

Espécies	Exon 11 do <i>AGAP-201</i>
<i>Homo sapiens</i>	GATTACATGCAGAATGTTTCATGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG AGCCACGTCAGCCTGCGCACCCATCTCCAGCCCTAAAACCAATGGCCTATCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Pan troglodytes</i>	GATTACATGCAGAACGTTTCATGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG AGCCACGTCAGCCTGTGCACCCATCTCCAGCCCTAAAACCAATGGCCTATCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Pan paniscus</i>	GATTACATGCAGAACGTTTCATGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG AGCCACGTCAGCCTGTGCACCCATCTCCAGCCCTAAAACCAATGGCCTATCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Gorilla gorilla</i>	GATTACATGCAGAATGTTTCATGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG AGCCACGTCAGCCTGCGCACCCATCTCCAGCCCTAAAACCAATGGCCTATCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Pongo abelii</i>	GATTACATGCAGAATGTTTCATGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG AGCCACGTCGGCCTGCGCGCCCATCTCCAGCCCTAAAACCAATGGCCTATCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Nomascus leucogenys</i>	GATTACATGCAGAATGTTTCATGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG AGCCACGTCGGCCTGCGCGCCCATCTCCAGCCCTAAAACCAATGGCCTATCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Chlorocebus</i>	GATTACATGCAGAATGTTTCATGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG

<i>sabaeus</i>	AGCCACGTCGGCCTGCGCGCCCATCTCCAGCCCTAAAACCAATGGCCTGTCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
	GATTACATGCAGAATGTTTCATGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG
<i>Macaca mulatta</i>	AGCCACGTCGGCCTGCGCGCCCATCTCCAGCCCTAAAACCAATGGCCTGTCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
	GATTACATGCAGAATGTTTCATGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG
<i>Macaca fascicularis</i>	AGCCACGTCGGCCTGCGCGCCCATCTCCAGCCCTAAAACCAATGGCCTGTCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
	GATTACATGCAGAATGTTTCATGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG
<i>Macaca nemestrina</i>	AGCCACGTCGGCCTGCGCGCCCATCTCCAGCCCTAAAACCAATGGCCTGTCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
	GATTACATGCAGAATGTTTCATGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG
<i>Papio anubis</i>	AGCCACGTCGGCCTGCGCGCCCATCTCCAGCCCTAAAACCAATGGCCTGTCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
	GATTACATGCAGAATGTTTCATGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG
<i>Rhinopithecus bieti</i>	AGCCACGTCGGCCTGCGCGCCCATCTCCAGCCCTAAAACCAATGGCCTGTCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
	GATTACATGCAGAATGTTTCATGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG
<i>Rhinopithecus roxellana</i>	AGCCACGTCAGCCTGCGCGCCCATCTCCAGCCCTAAAACCAATGGCCTGTCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Cercocebus</i>	GATTACATGCAGAATGTTTCATGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG

<i>atys</i>	AGCCACGTCGGCCTGCGCGCCCATCTCCAGCCCTAAAACCAATGGCCTGTCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Colobus angolensis</i>	GATTACATGCAGAATGTTTCATGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG AGCCACGTCGGCCTGCGCGCCCATCTCCAGCCCTAAAACCAATGGCCTGTCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Mandrillus leucophaeus</i>	GATTACATGCAGAATGTTTCATGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG AGCCACGTCGGCCTGCGCACCCATCTCCAGCCCTAAAACCAATGGCCTGTCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Propithecus coquereli</i>	GATTACATGCAGAATGTTTCATGGTAAGGAGATTGATCTGCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG AGCCACGTCGGCCTGCGCGCCCATCTCCAGCCCTAAAACCAATGGCCTGTCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Otolemur garnettii</i>	GATTACATGCAGAATGTCCATGGTAAGGAGATTGATCTGCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCGCCCCG AGCCACGTCCGCCTGCGCGCCCATCTCCAGCCCTAAAACCAATGGCCTGTCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Microcebus murinus</i>	GATTACATGCAGAATGTTTCATGGTAAGGAGATTGATCTGCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG AGCCACCTCAGCCTGCGCGCCCATCTCCAGCCCTAAAACCAATGGCCTGTCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Carlito syrichta</i>	GATTACATGCAGAATGTTTCATGGTAAGGAGATTGATCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCTCG AGCCACGTCGGCCTGCGCACCTATCTCCAGCCCTAAAACCAATGGCCTATCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Callithrix</i>	GATTACATGCAGAATGTTTCACGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG

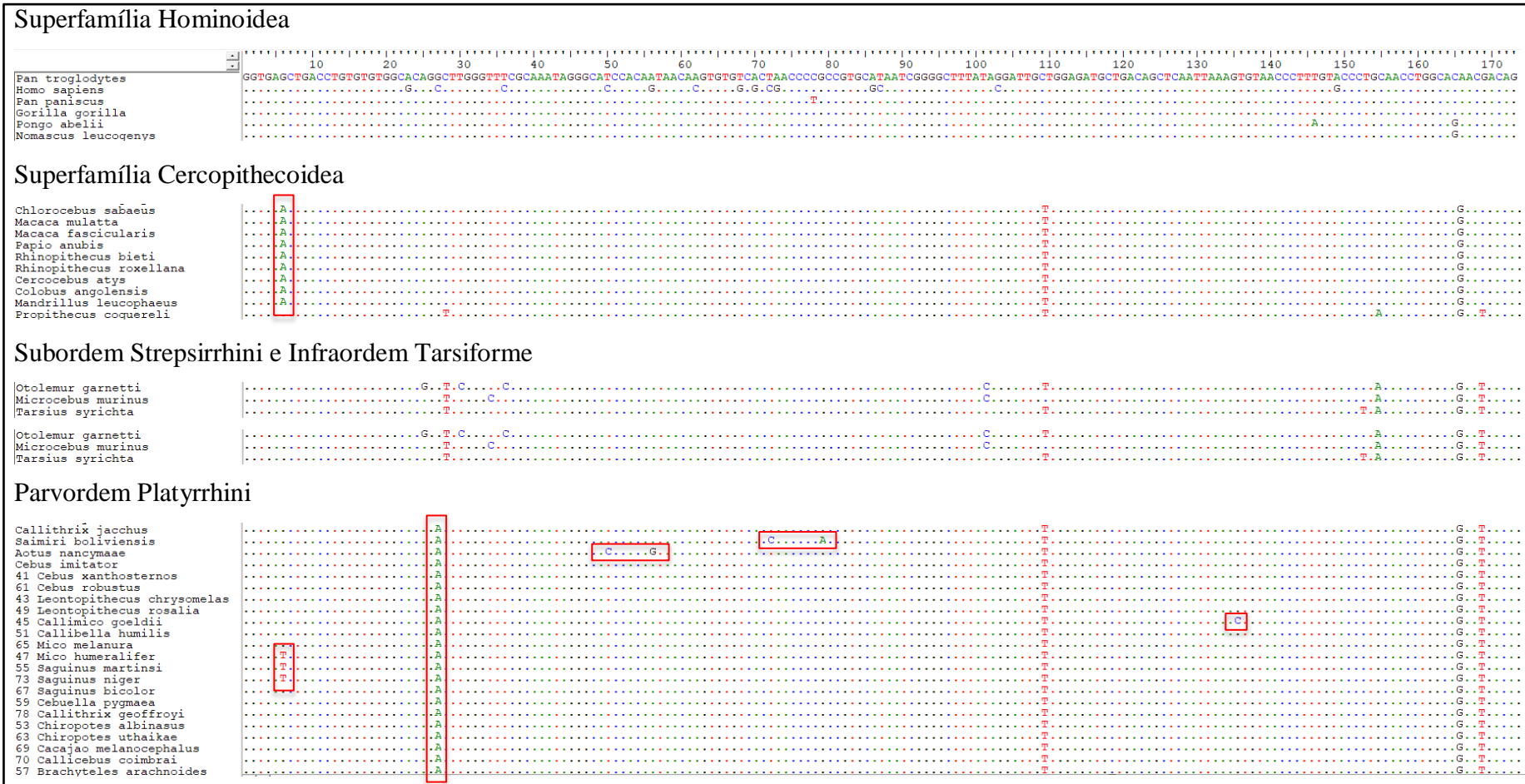
<i>jacchus</i>	AGCCACGTCAGCCTGCGCTCCCATCTCCAGCCCTAAAACCAATGGCCTATCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Saimiri boliviensis</i>	GATTACATGCAGAATGTTACGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCGCCCG AGCCACGTCAGCCTGCGCTCCCATCTCCAGCCCTAAAACCAATGGCCTGTCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Aotus nancymae</i>	GATTACATGCAGAATGTTACGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCGCCCG AGCCACGTCAGCCTGCGCTCCCATCTCCAGCCCTAAAACCAATGGCCTATCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Cebus capucinus</i>	GATTACATGCAGAATGTTACGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCGCCCG AGCCACGTCAGCCTGCGCTCCCATCTCCAGCCCTAAAACCAATGGCCTGTCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Cebus xanthosternos</i>	GATTACATGCAGAATGTTACGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCGCCCG AGCCACGTCAGCCTGCGCTCCCATCTCCAGCCCTAAAACCAATGGCCTATCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Cebus robustus</i>	GATTACATGCAGAATGTTACGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCGCCCG AGCCACGTCAGCCTGCGCTCCCATCTCCAGCCCTAAAACCAATGGCCAATGCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Leontopithecus chrysomelas</i>	GATTACATGCAGAAGGTTACGGTAACGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG AGCCACGTCAGCCTGCGCTCCCATCTCCAGCCCTAAAACCAATGGCCTATCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Leontopithecus</i>	GATTACATGCAGAATGTTACGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG

<i>rosalia</i>	AGCCACGTCAGCCTGCGCTCCCATCTCCAGCCCTAAAACCAATGGCCTATCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Callimico goeldii</i>	GATTACATGCAGAATGTTACGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG AGCCACGTCAGCCTGCGCTCCCATCTCCAGCCCTAAAACCAATGGCCTATCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Callibella humilis</i>	GATTACATGCAGAATGTTACGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG AGCCACGTCAGCCTGCGCTCCCCTCTCCAGCCCTAAAACCAATGGCCTATCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Mico melanura</i>	GATTACATGCAGAATGTTACGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG AGCCACGTCAGCCTGCGCTCCCATCTCCAGCCCTAAAACCAATGGCCTATCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Mico humeralifera</i>	GATTACATGCAGAATGTTACGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCGCCCCG AGCCACGTCAGCCTGCGCTCCCATCTCCAGCCCTAAAACCAATGGCCTATCCAAGGACATGAGCAGTTTACACCTCTCAC CCAATT
<i>Saguinus martinsi</i>	GATTACATGCAGAATGTTACGGTAAGGAGATCGACCTTCTGAGAACCACTGTGAAAGTCCCAGGAAAGAGGCCGCCCCG AGCCACATCAGCCTGCGCTCCCATCTCCAGCCCTAAAACCAATGGCCTGTCCAAGGACATAAGCAGTTTACACATCTCAC CCAATT
<i>Saguinus niger</i>	GATTACATGCAGAATGTTACGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG AGCCACGTCAGCCTGCGCTCCCATCTCCAGCCCTAAAACCAATGGCCTATCCAAGGACATGAGCAGTTTACACCTCTCAC CCAATT
<i>Saguinus</i>	GATTACATGCAGAATGTTACGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG

<i>bicolor</i>	AGCCACGTCAGCCTGCGCTCCCATCTCCAGCCCTAAAACCCATGGCCTATCCCAGGACATGAGGAGTTTACACATCTCAC CCCATT
<i>Cebuella pygmaea</i>	GATTACATGCAGAATGTTACGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG AGCCACGTCAGCCTGCGCTCCCCTCTCCAGCCCTAAAACCAATGGCCTATCCAAGGACATGAGCAGTTTACTCATCTCAC CCAATT
<i>Callithrix geoffroyi</i>	GATTACATGCAGAATGTTACGGTAAGGAGATTGACCTTCTGAGAACCACTGTGAAAGTCCCAGGGAAGAGGCCACCCCG AGCCACGTCAGCCTGCGCTCCCATCTCCAGCCCTAAAACCAATGGCCTATCCAAGGACATGAGCAGTTTACACATCTCAC CCAATT
<i>Chiropotes albinasus</i>	GATTACATGCAGAATGTTACGGTAAGGAGATCGACCTTCTGAGAACCACTGTGAAAGTCCCAGGAAAGAGGCCGCCCCCG AGCCACATCAGCCTGCACTCCCATCTCCAGCCCTAAAACCAATGGCCTGTCCAAGGACATAAGCAGTTTACACATCTCGC CCAATT
<i>Chiropotes utahickae</i>	GATTACATGCAGAATGTTACGGTAAGGAGATCGACCTTCTGAGAACCACTGTGAAAGTCCCAGGAAAGAGGCCGCCCCCG AGCCACATCAGCCTGCACTCCCATCTCCAGCCCTAAAACCAATGGCCTGTCCAAGGACATAAGCAGTTTACACATCTCGC CCAATT
<i>Cacajao melanocephalus</i>	GATTACATGCAGAATGTTACGGTAAGGAGATCGACCTTCTGAGAACCACTGTGAAAGTCCCAGGAAAGAGGCCGCCCCCG AGCCACATCAGCCTGCACTCCCATCTCCAGCCCTAAAACCAATGGCCTGTCCAAGGACATAAGCAGTTTACACATCTCAC CCAATT

Fonte: Gabrielle A. Rizzato

Anexo IV. Alinhamento do HACNS1 de primatas de velho e novo mundo feito no BioEdit. Retângulos vermelhos mostram substituições específicas.



Fonte: Gabrielle A. Rizzato

Anexo V. Resultado do BEB do modelo *Alternate 2*.

NON-BRANCH SPECIFIC TEST.

NULL=NEGATIVE SELECTION AND NEUTRAL EVOLUTION.

ALTERNATE= CLASS 1: NEGATIVE, CLASS 2: NEUTRAL EVOLUTION AND CLASS 3:

POSITIVE SELECTION.

ZETA0: 0.1690654854523994

ZETA1: 1

ZETA2: 0.02432873312756848

LK NULL = -1034.987357403621 LK ALT = -1014.430394721923

LRT = 41.11392536339463

LRT P-VALUE = 0.0000000255

CALCULATE BEB

SITE	BEB		
	NEG	NEUTRAL	Pos
1	0.885412*	0.040715	0.073872
2	0.885412*	0.040715	0.073872
3	0.882165*	0.041800	0.076034
4	0.885412*	0.040715	0.073872
5	0.887263*	0.040099	0.072638
6	0.001590	0.259683	0.738727*
7	0.885378*	0.040733	0.073889
8	0.882165*	0.041800	0.076034
9	0.885412*	0.040715	0.073872
10	0.887263*	0.040099	0.072638
11	0.885378*	0.040733	0.073889
12	0.885378*	0.040733	0.073889
13	0.882165*	0.041800	0.076034
14	0.885412*	0.040715	0.073872
15	0.882165*	0.041800	0.076034
16	0.885412*	0.040715	0.073872
17	0.882165*	0.041800	0.076034
18	0.885412*	0.040715	0.073872
19	0.882165*	0.041800	0.076034
20	0.885412*	0.040715	0.073872
21	0.885412*	0.040715	0.073872
22	0.885378*	0.040733	0.073889
23	0.313599	0.231433	0.454968*
24	0.885378*	0.040733	0.073889
25	0.313606	0.231366	0.455028*
26	0.885412*	0.040715	0.073872
27	0.073362	0.282947	0.643691*
28	0.308293	0.232461	0.459246*
29	0.882165*	0.041800	0.076034
30	0.306998	0.232723	0.460279*
31	0.885412*	0.040715	0.073872

32	0.885412*	0.040715	0.073872
33	0.885412*	0.040715	0.073872
34	0.306997	0.232718	0.460285*
35	0.882165*	0.041800	0.076034
36	0.077145	0.283838	0.639017*
37	0.885378*	0.040733	0.073889
38	0.885412*	0.040715	0.073872
39	0.885378*	0.040733	0.073889
40	0.887263*	0.040099	0.072638
41	0.887263*	0.040099	0.072638
42	0.887263*	0.040099	0.072638
43	0.882165*	0.041800	0.076034
44	0.887263*	0.040099	0.072638
45	0.885412*	0.040715	0.073872
46	0.885412*	0.040715	0.073872
47	0.885412*	0.040715	0.073872
48	0.885378*	0.040733	0.073889
49	0.887263*	0.040099	0.072638
50	0.076936	0.283769	0.639296*
51	0.885378*	0.040733	0.073889
52	0.885378*	0.040733	0.073889
53	0.887263*	0.040099	0.072638
54	0.885378*	0.040733	0.073889
55	0.887263*	0.040099	0.072638
56	0.080539	0.284505	0.634956*
57	0.882165*	0.041800	0.076034
58	0.887263*	0.040099	0.072638
59	0.887263*	0.040099	0.072638
60	0.885378*	0.040733	0.073889
61	0.887263*	0.040099	0.072638
62	0.302186	0.233710	0.464104*
63	0.885412*	0.040715	0.073872
64	0.882165*	0.041800	0.076034
65	0.885412*	0.040715	0.073872
66	0.882165*	0.041800	0.076034
67	0.885412*	0.040715	0.073872
68	0.296237	0.234850	0.468913*
69	0.885378*	0.040733	0.073889
70	0.313599	0.231433	0.454968*
71	0.885378*	0.040733	0.073889
72	0.077159	0.283820	0.639021*
73	0.313599	0.231433	0.454968*
74	0.887263*	0.040099	0.072638
75	0.885378*	0.040733	0.073889
76	0.885378*	0.040733	0.073889
77	0.885378*	0.040733	0.073889
78	0.311370	0.231849	0.456781*
79	0.311332	0.231806	0.456863*
80	0.885378*	0.040733	0.073889
81	0.885378*	0.040733	0.073889
82	0.885412*	0.040715	0.073872

83	0.882165*	0.041800	0.076034
84	0.885412*	0.040715	0.073872
85	0.885378*	0.040733	0.073889
86	0.313599	0.231433	0.454968*
87	0.307156	0.232747	0.460097*
88	0.887263*	0.040099	0.072638
89	0.887263*	0.040099	0.072638
90	0.882165*	0.041800	0.076034
91	0.885378*	0.040733	0.073889
92	0.885412*	0.040715	0.073872
93	0.885412*	0.040715	0.073872
94	0.885412*	0.040715	0.073872
95	0.885412*	0.040715	0.073872
96	0.885378*	0.040733	0.073889
97	0.882165*	0.041800	0.076034
98	0.882165*	0.041800	0.076034
99	0.882165*	0.041800	0.076034
100	0.887263*	0.040099	0.072638
101	0.076750	0.283651	0.639599*
102	0.887263*	0.040099	0.072638
103	0.300011	0.234119	0.465870*
104	0.885412*	0.040715	0.073872
105	0.887263*	0.040099	0.072638
106	0.882165*	0.041800	0.076034
107	0.882165*	0.041800	0.076034
108	0.885412*	0.040715	0.073872
109	0.077261	0.283724	0.639015*
110	0.882165*	0.041800	0.076034
111	0.885412*	0.040715	0.073872
112	0.885412*	0.040715	0.073872
113	0.887263*	0.040099	0.072638
114	0.885412*	0.040715	0.073872
115	0.887263*	0.040099	0.072638
116	0.882165*	0.041800	0.076034
117	0.885412*	0.040715	0.073872
118	0.885378*	0.040733	0.073889
119	0.882165*	0.041800	0.076034
120	0.885412*	0.040715	0.073872
121	0.887263*	0.040099	0.072638
122	0.885378*	0.040733	0.073889
123	0.887263*	0.040099	0.072638
124	0.885412*	0.040715	0.073872
125	0.885378*	0.040733	0.073889
126	0.882165*	0.041800	0.076034
127	0.885378*	0.040733	0.073889
128	0.887263*	0.040099	0.072638
129	0.887263*	0.040099	0.072638
130	0.882165*	0.041800	0.076034
131	0.882165*	0.041800	0.076034
132	0.887263*	0.040099	0.072638
133	0.887263*	0.040099	0.072638

134	0.887263*	0.040099	0.072638
135	0.310580	0.231954	0.457466*
136	0.882165*	0.041800	0.076034
137	0.885412*	0.040715	0.073872
138	0.882165*	0.041800	0.076034
139	0.887263*	0.040099	0.072638
140	0.887263*	0.040099	0.072638
141	0.885378*	0.040733	0.073889
142	0.885378*	0.040733	0.073889
143	0.885378*	0.040733	0.073889
144	0.882165*	0.041800	0.076034
145	0.882165*	0.041800	0.076034
146	0.306349	0.232835	0.460817*
147	0.885412*	0.040715	0.073872
148	0.882165*	0.041800	0.076034
149	0.313599	0.231433	0.454968*
150	0.885378*	0.040733	0.073889
151	0.885378*	0.040733	0.073889
152	0.310460	0.232036	0.457504*
153	0.882165*	0.041800	0.076034
154	0.309690	0.232135	0.458175*
155	0.885378*	0.040733	0.073889
156	0.887263*	0.040099	0.072638
157	0.887263*	0.040099	0.072638
158	0.885378*	0.040733	0.073889
159	0.885378*	0.040733	0.073889
160	0.882165*	0.041800	0.076034
161	0.885412*	0.040715	0.073872
162	0.885412*	0.040715	0.073872
163	0.885378*	0.040733	0.073889
164	0.887263*	0.040099	0.072638
165	0.309067	0.232278	0.458655*
166	0.887263*	0.040099	0.072638
167	0.887263*	0.040099	0.072638
168	0.290645	0.235942	0.473414*
169	0.885412*	0.040715	0.073872
170	0.887263*	0.040099	0.072638
171	0.885378*	0.040733	0.073889
172	0.887263*	0.040099	0.072638
173	0.885412*	0.040715	0.073872

Fonte: Gabrielle A. Rizzato

Anexo VI. TFBS preditos como ganhos e perdidos no HACNS devido a substituições específicas (em amarelo). Flechas pra direita indicam fita (+) e esquerda fita (-)

	PTX1 →			CAMTA1												← MYRF																				
<u>H. sapiens</u>	TG	GC	CCG	AC	CCG	CCG	TGC	G	CAAT	CGGG	CTTT	TAT	CGA	TTG	CTG	GAGA	TG	CTG	GAC	AGC	TCA	ATT	AAAG	TG	TAA	CCCT	TGT	GC	CCCT	G	AAC	CT	GG	C	A	ACG
<u>P. troglodytes</u>	TG	CA	CT	AAC	CCC	GCC	TGC	A	TAAT	CGGG	CTTT	TAT	AGGA	TTG	CTG	GAGA	TG	CTG	GAC	AGC	TCA	ATT	AAAG	TG	TAA	CCCT	TGT	ACC	CTG	AAC	CT	GG	C	A	ACG	
<u>P. paniscus</u>	TG	CA	CT	AAC	CCT	GCC	TGC	A	TAAT	CGGG	CTTT	TAT	AGGA	TTG	CTG	GAGA	TG	CTG	GAC	AGC	TCA	ATT	AAAG	TG	TAA	CCCT	TGT	ACC	CTG	AAC	CT	GG	C	A	ACG	
<u>G. gorilla</u>	TG	CA	CT	AAC	CCC	GCC	TGC	A	TAAT	CGGG	CTTT	TAT	AGGA	TTG	CTG	GAGA	TG	CTG	GAC	AGC	TCA	ATT	AAAG	TG	TAA	CCCT	TGT	ACC	CTG	AAC	CT	GG	C	A	ACG	
<u>P. abelii</u>	TG	CA	CT	AAC	CCC	GCC	TGC	A	TAAT	CGGG	CTTT	TAT	AGGA	TTG	CTG	GAGA	TG	CTG	GAC	AGC	TCA	ATT	AAAG	TG	TAA	CCCT	AG	TAC	CTG	AAC	CT	GG	C	A	ACG	
<u>N. leucogenys</u>	TG	CA	CT	AAC	CCC	GCC	TGC	A	TAAT	CGGG	CTTT	TAT	AGGA	TTG	CTG	GAGA	TG	CTG	GAC	AGC	TCA	ATT	AAAG	TG	TAA	CCCT	TGT	ACC	CTG	AAC	CT	GG	C	A	ACG	
<u>C. sabaesus</u>	TG	CA	CT	AAC	CCC	GCC	TGC	A	TAAT	CGGG	CTTT	TAT	AGGA	TTG	TG	TGAGA	TG	CTG	GAC	AGC	TCA	ATT	AAAG	TG	TAA	CCCT	TGT	ACC	CTG	AAC	CT	GG	C	A	ACG	
<u>M. mulata</u>	TG	CA	CT	AAC	CCC	GCC	TGC	A	TAAT	CGGG	CTTT	TAT	AGGA	TTG	TG	TGAGA	TG	CTG	GAC	AGC	TCA	ATT	AAAG	TG	TAA	CCCT	TGT	ACC	CTG	AAC	CT	GG	C	A	ACG	
<u>M. fascicularis</u>	TG	CA	CT	AAC	CCC	GCC	TGC	A	TAAT	CGGG	CTTT	TAT	AGGA	TTG	TG	TGAGA	TG	CTG	GAC	AGC	TCA	ATT	AAAG	TG	TAA	CCCT	TGT	ACC	CTG	AAC	CT	GG	C	A	ACG	
<u>M. nemestrina</u>	TG	CA	CT	AAC	CCC	GCC	TGC	A	TAAT	CGGG	CTTT	TAT	AGGA	TTG	TG	TGAGA	TG	CTG	GAC	AGC	TCA	ATT	AAAG	TG	TAA	CCCT	TGT	ACC	CTG	AAC	CT	GG	C	A	ACG	
<u>P. anubis</u>	TG	CA	CT	AAC	CCC	GCC	TGC	A	TAAT	CGGG	CTTT	TAT	AGGA	TTG	TG	TGAGA	TG	CTG	GAC	AGC	TCA	ATT	AAAG	TG	TAA	CCCT	TGT	ACC	CTG	AAC	CT	GG	C	A	ACG	
<u>R. bieiti</u>	TG	CA	CT	AAC	CCC	GCC	TGC	A	TAAT	CGGG	CTTT	TAT	AGGA	TTG	TG	TGAGA	TG	CTG	GAC	AGC	TCA	ATT	AAAG	TG	TAA	CCCT	TGT	ACC	CTG	AAC	CT	GG	C	A	ACG	
<u>R. roxellana</u>	TG	CA	CT	AAC	CCC	GCC	TGC	A	TAAT	CGGG	CTTT	TAT	AGGA	TTG	TG	TGAGA	TG	CTG	GAC	AGC	TCA	ATT	AAAG	TG	TAA	CCCT	TGT	ACC	CTG	AAC	CT	GG	C	A	ACG	
<u>C. atys</u>	TG	CA	CT	AAC	CCC	GCC	TGC	A	TAAT	CGGG	CTTT	TAT	AGGA	TTG	TG	TGAGA	TG	CTG	GAC	AGC	TCA	ATT	AAAG	TG	TAA	CCCT	TGT	ACC	CTG	AAC	CT	GG	C	A	ACG	
<u>C. angolensis</u>	TG	CA	CT	AAC	CCC	GCC	TGC	A	TAAT	CGGG	CTTT	TAT	AGGA	TTG	TG	TGAGA	TG	CTG	GAC	AGC	TCA	ATT	AAAG	TG	TAA	CCCT	TGT	ACC	CTG	AAC	CT	GG	C	A	ACG	
<u>M. leucophaeus</u>	TG	CA	CT	AAC	CCC	GCC	TGC	A	TAAT	CGGG	CTTT	TAT	AGGA	TTG	TG	TGAGA	TG	CTG	GAC	AGC	TCA	ATT	AAAG	TG	TAA	CCCT	TGT	ACC	CTG	AAC	CT	GG	C	A	ACG	
<u>P. coquereli</u>	TG	CA	CT	AAC	CCC	GCC	TGC	A	TAAT	CGGG	CTTT	TAT	AGGA	TTG	TG	TGAGA	TG	CTG	GAC	AGC	TCA	ATT	AAAG	TG	TAA	CCCT	TGT	ACC	TAC	AAC	CT	GG	C	A	AATG	
<u>O. garnettii</u>	TG	CA	CT	AAC	CCC	GCC	TGC	A	TAAT	CGGG	CTTT	TAC	AGGA	TTG	TG	TGAGA	TG	CTG	GAC	AGC	TCA	ATT	AAAG	TG	TAA	CCCT	TGT	ACC	TAC	AAC	CT	GG	C	A	AATG	
<u>M. murinus</u>	TG	CA	CT	AAC	CCC	GCC	TGC	A	TAAT	CGGG	CTTT	TAC	AGGA	TTG	CTG	TGAGA	TG	CTG	GAC	AGC	TCA	ATT	AAAG	TG	TAA	CCCT	TGT	ACC	TAC	AAC	CT	GG	C	A	AATG	
<u>T. syrichta</u>	TG	CA	CT	AAC	CCC	GCC	TGC	A	TAAT	CGGG	CTTT	TAT	AGGA	TTG	TG	TGAGA	TG	CTG	GAC	AGC	TCA	ATT	AAAG	TG	TAA	CCCT	TGT	ACC	TAC	AAC	CT	GG	C	A	AATG	

