

PAX9: Uma ferramenta evolutiva?

Vanessa Rodrigues Paixão-Côrtes

Dissertação submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de Mestre.

Orientadora: Profa. Dra. Maria Cátira Bortolini

Porto Alegre, Março de 2008.

Este trabalho foi realizado nas instalações do Laboratório de Evolução Molecular e Genética de Populações Humanas do Departamento de Genética do Instituto de Biociências da Universidade Federal do Rio Grande do Sul e no Laboratório de Biologia Evolutiva e Conservação de Vertebrados (Labec) do Instituto de Biociências da USP, Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan e foi subvencionado pelo programa Institutos do Milênio e Programa de Apoio a Núcleos de Excelência, pelo Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), pela CAPES e pela Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS).

***" O que não se sabe,
Aprende-se..."***

AGRADECIMENTOS

- ✓ À Dra. Maria Cátira Bortolini pela orientação, confiança e incentivo;
- ✓ Ao Dr. Francisco Mauro Salzano, pelo fato de estar sempre presente e disposto a ajudar. Um exemplo como cientista e pessoa;
- ✓ À Dra. Vera Gayesky, ao Dr. Aldo Mellender, à Dra. Karen Haag, à Dra. Sídia Callegari-Jacques e a todos os professores do Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS, que contribuíram muito para o meu aperfeiçoamento;
- ✓ Ao Dr. Diogo Meyer pela valiosa assistência e colaboração;
- ✓ Aos componentes do Laboratório de Biologia Evolutiva e Conservação de Vertebrados (Labec), principalmente à Márcia Pincerati, à Kelly Nunes, Rodrigo Ramalho e Rodrigo Francisco por me receberem e por me ajudarem com os programas Phred, Phrap e Consed;
- ✓ Ao Prof. Dr. Fabrício Santos e ao Dr. Rodrigo Redondo do banco de DNA da Universidade Federal de Minas Gerais pela colaboração e pelo empréstimo das amostras;
- ✓ Ao Dr. Alcides Pissinatti veterinário responsável pelas coletas de amostras de Macacos do Novo Mundo;
- ✓ Ao Centro de Primatologia do Rio de Janeiro (CPRJ-FEEMA), à Estação Ecológica Estadual Paraíso (ESEC/ FEEMA) por disponibilizar material científico e orientação na execução da pesquisa;
- ✓ Ao Ministério Público Federal e Estadual, ao Instituto Brasileiro do Meio Ambiente e Recursos Naturais Renováveis (IBAMA), à Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro – FAPERJ (Proc. N° E-26/171.573/2000 e E-26/171.185/2004) ao *Greater Los Angeles Zoo Association* (GLAZA), *The Zoological Society of Philadelphia*, *American Society of Primatologist* (ASP) e a *Conservation Internacional* (CI), pela constante cooperação com o programa de reprodução de primatas do neotrópico.

- ✓ Ao GEMARS (Grupo de Estudos de Mamíferos Aquáticos do Rio Grande do Sul), principalmente a Larissa S. Heinzemann pelo suporte e atenção;
- ✓ Ao Projeto Tuco-Tuco, e ao Dr. Thales de Freitas (Dep. Genética-UFRGS) pelas amostras cedidas.
- ✓ Ao Dr. Fengtang Yang (*The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Inglaterra*), que disponibilizou as amostras da espécie *Manis javanica* (pangolin de Java).
- ✓ Aos estudantes Lin Huang e Jun Che do *Laboratory of Cellular and Molecular Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming, Yunnan*. Por fazerem as reações de PCR e o seqüenciamento da amostra de pangolin;
- ✓ A Dra. Eliane Bandinelli, pelo empréstimo do termociclador em um momento crítico;
- ✓ Ao Steph Mazières pela colaboração com o mapa sintético;
- ✓ Ao Elmo e à Ellen por ajudarem sempre;
- ✓ À Lúcia e à Laci pela companhia e o café;
- ✓ À Andrea que me ensinou quando iniciei em meu estágio;
- ✓ Aos companheiros de Sala e Laboratório Lu, Tábita, Victor, Vanderlei, Bisso e Isabel pela força e pela paciência de me escutarem;
- ✓ Aos meus colegas;
- ✓ À Karen, ao Dennis, à Cláudia e a todos os meus amigos;
- ✓ A toda a minha família, em especial à Vó, Mãe e Aline: obrigado por cuidarem de mim;
- ✓ A família que agora é minha também a Débora, o Bolívar, a Maria Luíza, o Dinho;
- ✓ **E ao Walter.... “Por Tudo, e mais um pouco !!!!!”.**

SUMÁRIO

RESUMO	8
ABSTRACT	9
CAPÍTULO I – INTRODUÇÃO	10
I-1. CONSIDERAÇÕES GERAIS	10
I-2. CAIXA DE FERRAMENTAS	11
I-3. EVOLUÇÃO DOS MAMÍFEROS	13
I-4. ORDEM PRIMATA	15
I-5. BIOLOGIA EVOLUTIVA DO DESENVOLVIMENTO	16
I-6. OS GENES PAX	17
I-7. PAX9 -UMA PEQUENA PROTEÍNA COM GRANDES FUNÇÕES	24
I-7.1 O GENE PAX9 E O DESENVOLVIMENTO DENTÁRIO	28
I-7.2 EVOLUÇÃO MOLECULAR DO PAX9 NA ORDEM PRIMATA	29
I-8. PROBLEMAS IGUAIS, SOLUÇÕES SEMELHANTES. MESMAS FERRAMENTAS?	33
CAPÍTULO II - OBJETIVOS	38
CAPÍTULO III – RESULTADOS	39
III-1- MANUSCRITO EM PREPARAÇÃO	39
PAX9 EXON 3 VARIABILITY IN NATIVE AMERICANS AND IN MAJOR HUMAN GROUPS: INTER-POPULATION AND INTRA-GENE EVOLUTIONARY DYNAMICS.	40
ABSTRACT	41
INTRODUCTION	42
MATERIALS AND METHODS	43
POPULATIONS EXAMINED	43
DNA EXTRACTION, PCR AMPLIFICATION AND SEQUENCING	44
DATA ANALYSIS	45
RESULTS	46
MUTATIONS AND POLYMORPHIC SITES	46
GENETIC VARIABILITY	49
THIRD MOLAR AGENESIS AND THE ALA240PRO POLYMORPHISMS	51
DEMOGRAPHIC SIMULATIONS AND NEUTRAL/NATURAL SELECTION TESTS	52
DISCUSSION	53
LITERATURE CITED	57
III-2. RESULTADOS ADICIONAIS	76
III-2.1. A VARIAÇÃO DO EXON 3 DO GENE PAX9 DENTRO DO GÊNERO CEBUS E ENTRE ESPÉCIES DE MACACOS DO NOVO MUNDO	79
III-2.2. A VARIAÇÃO DA PORÇÃO CODIFICADORA DO GENE PAX9 EM MAMÍFEROS	91
III-2.2.1. ONIPRESENTE SELEÇÃO PURIFICADORA NO EXON 2	96

<i>III-2.2.2. O EXON 3 NA LINHAGEM DOS MAMÍFEROS</i>	97
<i>III-2.2.3. A VARIAÇÃO DO EXON 4 DO GENE PAX9 EM MAMÍFEROS</i>	99
<i>III-2.3. CONVERGÊNCIA E PARALELISMO MOLECULAR</i>	103
<u>CAPÍTULO IV – DISCUSSÃO</u>	109
<u>CAPÍTULO V - REFERÊNCIAS ONLINE</u>	113
<u>CAPÍTULO VI - REFERÊNCIAS BIBLIOGRÁFICAS</u>	115
<u>APÊNDICE I</u>	130
<i>AI-1 AMOSTRAS</i>	131
<i>AI-2 REAÇÃO DE AMPLIFICAÇÃO</i>	132
<i>AI-3 SEQÜENCIAMENTO</i>	134
<i>AI-4 ANÁLISE DE DADOS</i>	134

RESUMO

Um dos maiores objetivos na ciência atual é poder relacionar alterações nos genes a mudanças nas morfologias de organismos multicelulares. Além disso, é importante poder determinar se estas modificações foram obras do acaso ou se a seleção natural teve um papel central moldando a forma e o tamanho das criaturas vivas. A busca por nutrientes e sua utilização é uma das questões mais básicas para a sobrevivência de uma espécie. Em mamíferos a diversificação e especialização dos dentes para triturar, rasgar e mastigar diferentes tipos de comida possibilitou um incremento na geração de energia e representou uma inovação chave para a sobrevivência deste grupo. No presente estudo foi investigado o gene *PAX9*, que codifica um fator de transcrição chave no desenvolvimento da dentição de mamíferos. Para 125 indivíduos de origem ameríndia foi seqüenciado o exon 3 (138 pb), conjuntamente com as regiões intrônicas flangeadoras 5' e 3' (232 pb e 220 pb, respectivamente) e os dados integrados com a informação disponível para a mesma região de 115 indivíduos de diferentes origens geográficas. Além disso, os exons 2, 3 e 4 (627 pb, 138 pb e 240 pb respectivamente) foram seqüenciados do DNA de 25 espécies de mamíferos e agrupados com os dados disponíveis de mais 29 espécies.

Em humanos, o polimorfismo Ala240Pro possui uma distribuição variada entre os Ameríndios da América do Sul e está presente em Esquimós, Asiáticos e Europeus, mas não está presente entre os Afro-Americanos. O panorama que surge é que provavelmente a partir da saída da África a frequência da mutação Ala240Pro aumenta, sendo possivelmente selecionada positivamente, o que pode estar relacionado à agenesia do terceiro molar, ou ainda, estar associado a genes conectados a adaptações ao frio. Com a chegada dos Ameríndios a regiões equatoriais, outro cenário parece dominar a evolução do exon 3, seja pela perda inicial da possível vantagem adaptativa, ou por causa da ação da deriva genética amplamente documentada nestas populações. Provavelmente em algum momento na história destas populações, fatores casuais poderiam ter sido preponderantes às restrições funcionais da proteína. Em geral, há uma maior variabilidade em nível de aminoácidos no exon 3 considerando todos os mamíferos estudados quando comparada com a do exon 2, ficando o exon 4 numa posição intermediária. O exon 3 seria um exemplo notável de uma situação que conferiria "evolubilidade" ao *PAX 9*.

ABSTRACT

One of the main goals in science today is to establish the relation between gene changes and variations in the morphologies of multicellular organisms. In addition, it is important to determine if these changes were randomic or whether natural selection had a central role shaping the form and size of living beings. Searching for nutrients and its use is one of the most basic issues for a species survival. In mammals teeth diversification and specialization to grind, tear and chew different types of food were a key innovation that allowed an increase in the generation of energy and in the survival of this group. In this study an investigation was performed in the Paired box gene 9 (*PAX9*), which codes for a transcription factor that was a key factor in the development of mammal dentition. A total of 125 persons from Amerindian populations were studied by sequencing the *PAX9* gene exon 3 (138 base pairs) as well as its 5' and 3' flanking intronic segments (232 bp and 220 bp, respectively) and the data integrated with the information available for the same genetic region from 115 individuals of different geographical origins. Moreover, exons 2, 3 and 4 (627 bp, 138 bp and 240 bp, respectively) were sequenced from DNA of 25 mammalian species and the results combined with the data available from other 29 additional species.

In humans, the Ala240Pro polymorphism has a varied distribution among the South American Amerindians and is present in Eskimos, Asians and Europeans, but it is not present among Afro-Americans. The pattern that emerges is that probably starting with the out-Africa migration, and possibly as a result of positive selection, the Ala240Pro mutation frequency increased, that could be related to third molar agenesis or an associated gene connected to cold adaptation. With the Amerindian arrival again in equatorial regions, another pattern seems prevalent in exon 3 evolution, either due to the loss of a possible initial adaptive advantage, or to the action of genetic drift, widely documented in these populations. Probably at some point in the history of these populations, random factors could have been dominant despite the protein functional restrictions. In a general way, there is a greater variability at the amino acid level in exon 3 of all mammals considered when compared to exon 2, exon 4 occupying an intermediary position. Exon 3 would be a remarkable example of a condition that would confer *PAX9* evolvability.

CAPÍTULO I – INTRODUÇÃO

I-1. Considerações Gerais

Vivemos em um universo que se expande, em um planeta onde a vida surgiu há aproximadamente quatro bilhões de anos em oceanos primitivos. E desde então os organismos estão em constante mudança. Há cerca de 100.000 anos atrás, na África, surgiram os primeiros humanos modernos. Somos fruto de uma intensa evolução biológica e cultural. Desde o início olhamos para os céus e para nós mesmos e perguntamos: De onde viemos? O que somos? Para onde vamos?. Na busca por respostas, foram alcançados feitos científicos extraordinários. Inclusive técnicas que permitiram a investigação em nível do DNA. Através desta molécula recontamos a história, revelamos processos evolutivos, estabelecemos novos paradigmas e vivenciamos uma era de progressos, que abre um enorme leque de possibilidades para novas pesquisas e debates. A ciência, assim como a vida, também evolui.

I-2. Caixa de Ferramentas

Aproximadamente a 530 milhões de anos, surgiu a “caixa de ferramentas genéticas básicas” dos metazoários (Kirschner e Gerhart, 1998). Protozoários, fungos, vegetais e animais compartilham genes mestres que determinaram as características iniciais responsáveis pela diversificação deste grupo. Por tanto, uma das mais surpreendentes descobertas da atualidade é que a maioria dos animais, não importando como são diferentes na aparência, compartilham inúmeras famílias de genes que regulam os maiores aspectos dos padrões corporais (Carroll, 2000).

Algumas teorias foram propostas para explicar como este conjunto inicial de genes possibilitou a estes organismos possuírem processos moleculares conservados, mas passarem por rápidas e diversas mudanças na morfologia, na organização de tecidos, na fisiologia e no desenvolvimento (Kirschner e Gerhart, 1998).

Um consenso é que houve vários *rounds* iniciais de duplicação e diversificação destes genes controladores, muitos estudos reportam o surgimento de novas redes regulatórias, por duplicação, divergência e posterior ganho de função. Um exemplo clássico é a família de genes *HOX*, responsáveis por determinar a identidade no eixo antero-posterior de diversas partes do embrião dos animais, embora variem

consideravelmente em número, estes genes possuem seqüências ortólogas tanto em artrópodes quanto em tetrápodes.

Há trinta anos atrás, King e Wilson (1975) descreveram a grande similaridade entre proteínas de humanos e chimpanzés. Estes autores propuseram que as grandes diferenças evolucionárias na anatomia e no comportamento entre estas duas espécies de primatas, seriam conseqüências de mudanças em mecanismos que controlavam a expressão dos genes e não em seqüências que codificavam proteínas estruturais. A publicação dos resultados dos projetos do Genoma Humano e do Genoma do Chimpanzé, que reportam um nível de similaridade genômica entre as duas espécies em torno de 99%, corroborou a idéia inicial de que as diferenças estariam relacionadas com mecanismos regulatórios (*International Human Genome Sequencing Consortium, 2001; The Chimpanzee Sequencing and Analysis Consortium, 2005*). Por tanto, diferenças morfológicas importantes observadas entre algumas espécies podem ser explicadas por pequenas variações em seqüências reguladoras.

Uma questão controversa é se as alterações morfológicas encontradas entre e dentro das linhagens evolutivas são produto de alterações em regiões controladoras transcritas, mas não traduzidas (LCR – *locus control region*) (Abzhanov e cols., 2006; Saphiro e cols., 2004) ou resultado de alterações diretas em genes estruturais e/ou

reguladores, Nei (2007) focando sua observação na espécie humana, sustenta que características morfológicas, por exemplo, podem ser produto de pequenas alterações de grande efeito em genes importantes.

Generalizando, genes que codificam fatores de transcrição envolvidos nos estágios iniciais do desenvolvimento são altamente conservados (Vogel e cols., 2006; Nei, 2007). Entretanto, mutações que levam a alguma alteração na expressão ou ainda na estrutura de uma proteína regulatória, podem afetar de uma maneira indireta, mas determinante, uma bateria de outros genes (sujeitos a regulação pela proteína alterada) que não apresentam nenhuma modificação em sua estrutura molecular original (Carroll, 2003). Sintetizando, a maioria dos eventos evolutivos aconteceria quando velhos genes aprendem novos truques (Brakefield, 2006).

I-3. Evolução dos Mamíferos

A origem da Ordem *Mammalia* remonta do final ao período Triássico há cerca de 150 milhões de anos. Entretanto, sua primeira grande diversificação ocorreu na era Cenozóica, há 65 milhões de anos. A radiação dos mamíferos vem sendo relacionada com a conquista de vários novos nichos depois da extinção dos dinossauros. A primeira

fauna de mamíferos era representada por animais noturnos e generalistas de pequeno e médio porte, com hábitos na maioria arborícolas, sendo alguns terrestres. Predadores maiores surgiram no final do Paleoceno, aproximadamente há 60 milhões de anos atrás, e herbívoros com especializações nos dentes somente no final do Eoceno, 45 milhões de anos no passado (Pough e *cols.*, 2003).

Estudos da evolução de mamíferos frequentemente usam dados obtidos pela análise de fósseis de dentes, pelo fato destes serem o registro mais abundante, e por serem estruturas que muito provavelmente sofrem grande pressão seletiva. O desenvolvimento de estruturas especializadas, para triturar, rasgar e mastigar diferentes tipos de comida possibilitou um incremento na geração de energia e representou uma inovação chave para evolução dos mamíferos. A geração de diferentes padrões dentários, bem como diferentes tipos de dentes são determinados por uma rede de genes durante o desenvolvimento. Assim, os dentes estão se tornando um grande modelo não só nos estudos sobre desenvolvimento, mas também nos estudos de biologia evolutiva, visto que uma questão importante a ser entendida é como uma mutação afeta o desenvolvimento e como e quando ela é traduzida em mudanças na morfologia (Jernvall e *cols.*, 1999, 2000, 2003; Line, 2003; Salazar-Ciudad e *cols.*, 2003).

I-4. Ordem Primata

A Ordem Primata possui uma ampla variedade de espécies, compreende desde os pequenos lêmures voadores de Madagascar, até os grandes macacos e o homem. Embora vários estudos moleculares indiquem que a linhagem dos primatas divergiu de outros mamíferos placentários há ~90 milhões de anos, as características morfológicas que os caracterizam provavelmente emergiram mais tarde. Isto explicaria porque o registro fóssil indica um tempo de divergência menor, entre 54 e 81 milhões de anos atrás (Tavaré e cols., 2002). Já os primatas do Novo Mundo ou neotropicais (infra-ordem *Platyrrhini*) teriam divergido da linhagem que acabou dando origem aos grandes macacos e humanos (infra-ordem *Catarrhini*) por volta de 35 – 48 milhões de anos (Chiu e cols., 1996; Kumar e Hedges, 1998).

Bem mais recente, entre cinco e sete milhões de anos é a separação do gênero *Homo* (homem e homínídeos extintos) do gênero *Pan* (chimpanzé e bonobos). Os representantes fósseis mais antigos pertencentes ao gênero *Homo* foram encontrado na África e estão datados em 1,8 -1,9 milhões de anos. Já os humanos anatomicamente modernos surgiram nas savanas africanas há ~130 mil anos antes do presente (Ebersberger e cols., 2002; Jobling e cols., 2004). A partir de

então, uma bem sucedida dispersão pelos outros continentes aconteceu (Fagundes e *cols.*, 2007), tendo como resultado a colonização efetiva de ambientes tão distintos quanto florestas tropicais e zonas árticas.

I-5. Biologia Evolutiva do Desenvolvimento

Segundo Brakefield (2006), os genes de desenvolvimento que delineiam as formas dos animais são freqüentemente expressos em diversos tipos de células e tecidos, e em diferentes tempos durante o desenvolvimento. Eles são altamente pleiotrópicos, e há crescentes evidências de que mudanças evolutivas tendem a ocorrer via ajustes no complexo aparato de regulação gênica. Muitas das diferenças em traços morfológicos complexos não ocorrem devido à presença ou ausência de genes particulares, mas sim devido a mudanças nos mecanismos que os controlam, afetando quando e onde um gene é expresso.

Famílias dos genes reguladores do desenvolvimento são amplamente distribuídas pelo reino animal, e uma fração substancial destes é expressa de maneira similar em embriões de espécies pertencentes a filos distantes filogeneticamente (Carroll, 2003; Wray, 2003). Porém, existem diferenças marcantes entre todos os organismos vivos, sendo que parte delas poderiam ser atribuídas a mudanças em genes reguladores. Dessa forma, o desenvolvimento embrionário é

considerado um momento crucial para o acontecimento de mudanças evolucionárias, pois a cascata de eventos que é acionada quando o óvulo é fecundado apresenta passos chaves (multiplicação, migração e diferenciação celular, apoptose) para a determinação do plano corporal e das estruturas do embrião. Sendo cada um desses passos controlado por fatores genéticos, é esperado que mutações nos genes envolvidos possam levar ao surgimento de grandes novidades evolutivas.

A maioria dos genes reguladores ativos no desenvolvimento embrionário são inativados ao seu término. Porém, é importante ressaltar, que alguns destes genes voltam a ter papel destacado após o estágio embrionário. Este é o caso do gene *PAX9*.

I-6. Os Genes PAX

Os genes *PAX* constituem uma grande família de genes para fatores de transcrição de importância fundamental no desenvolvimento de órgãos e tecidos. Inicialmente foram descobertos em *Drosophila* (*Pox meso*). Subseqüentemente foram encontrados homólogos em diversos animais, sejam eles vertebrados ou invertebrados. Caracterizam-se por compartilhar um domínio conservado, chamado de domínio *paired* que apresenta normalmente 128 aminoácidos, com propriedades de se ligar a um elemento promotor do DNA alvo (Dahl e cols., 1997; Wilm e cols.,

1998; Peters e *cols.*, 1999). Motivos de ligação para as proteínas PAX foram encontrados em diversas regiões promotoras de genes que codificam uma variedade de proteínas, incluindo outros fatores de transcrição, receptores, hormônios peptídeos e proteínas de adesão. Por este motivo suas funções estão relacionadas à proliferação, especificação, migração e sobrevivência celular particularmente durante o período embrionário (Robson e *cols.*, 2006).

Como aconteceu com as outras famílias gênicas encontradas no genoma eucarioto, é provável, na medida em que emergiram os animais mais complexos, que os genes *PAX* se diversificaram a partir de um gene ancestral comum, através de duplicação e subsequente divergência, resultando, inclusive, no surgimento de novos domínios expressos (Balczarek e *cols.*, 1997; Peters e *cols.*, 1998; Ogasawara e *cols.*, 1999, 2000). Em vertebrados foram encontrados nove genes *PAX* (*PAX 1... 9*), sendo que estes podem ser agrupados em quatro grupos parálogos (Tabela 1). Tem sido considerado ainda que os quatro grupos de genes provavelmente originaram-se antes da radiação Cambriana, que se iniciou há 542 milhões de anos atrás (Breitling e *cols.*, 2000; Hetzer-Egger e *cols.*, 2000; Miller e *cols.*, 2000). A classificação nestes quatro grupos é baseada na presença ou ausência de duas outras seqüências conservadas, um homeodomínio e um octapeptídeo, e outras características estruturais (Figura 1), ou seja, no nível da organização

TABELA 1 . Estrutura e localização cromossômica dos genes das quatro famílias *Pax* de mamíferos

GRUPO	GENE	ESTRUTURA			LOCALIZAÇÃO CROMOSSÔMICA	
		Domínio- <i>paired</i>	Octapeptídeo	Homeodomínio	Rato	Homem
I	<i>Pax1</i>	Sim	Sim	Não	2	20p11
	<i>Pax9</i>				12	14q12-q13
II	<i>Pax2</i>	Sim	Sim	Sim	19	10q25
	<i>Pax5</i>				4	9p13
	<i>Pax8</i>				2	2q12-q14
III	<i>Pax3</i>	Sim	Sim	Sim	1	2q35
	<i>Pax7</i>				4	1p36
IV	<i>Pax4</i>	Sim	Não	Sim	6	7p32
	<i>Pax6</i>				2	11p13

Fonte: Peters e cols. (1998) com adaptações.

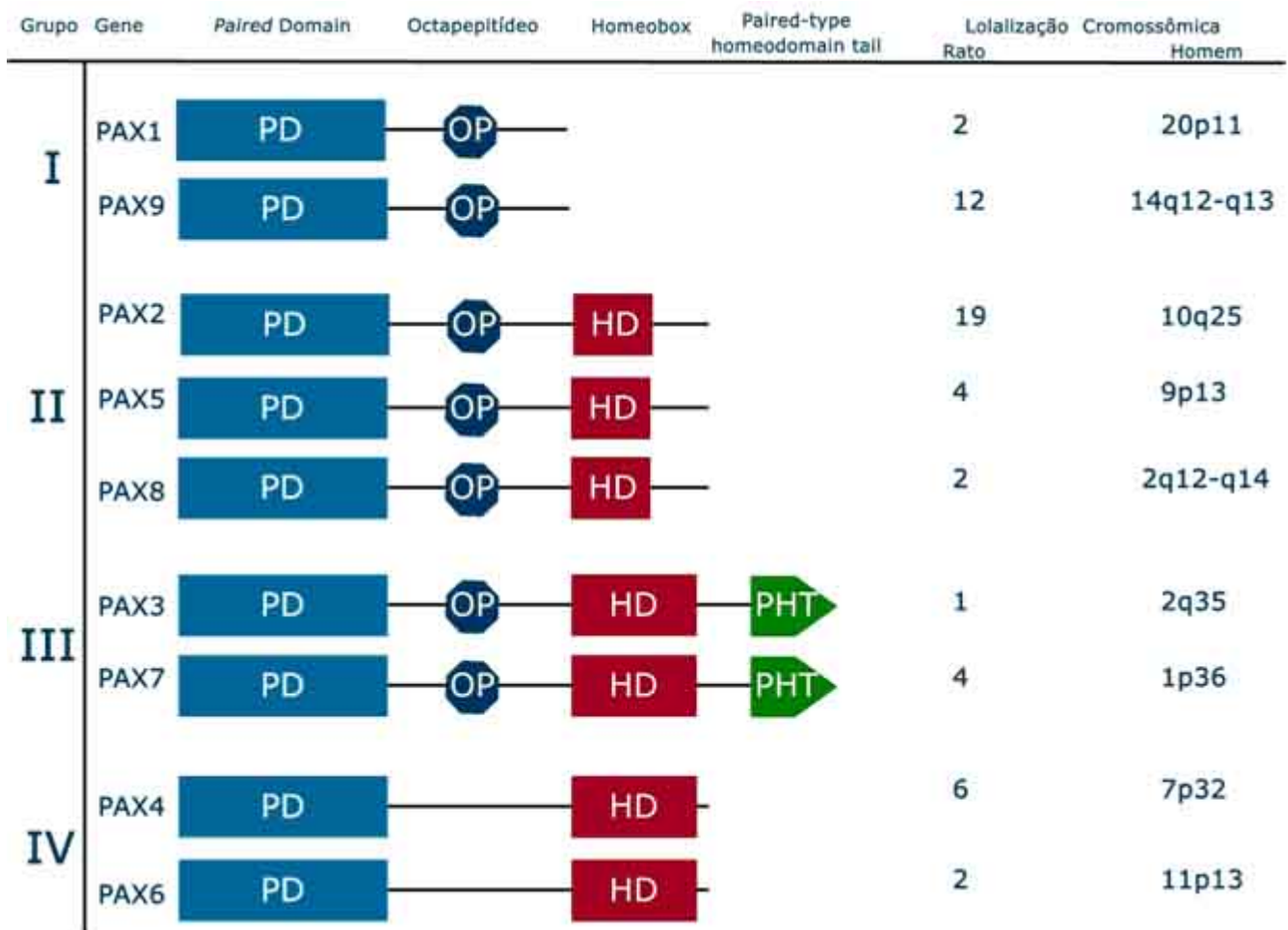


FIGURA 1. Esquema de classificação das subfamílias PAX. (Modificado de Balczarek, Lai e Kumar (1997), Breitling e Gerber (2000))

genômica conservada (Peters e cols., 1998; Ogasawara e cols., 1999). *PAX1* e *PAX9* constituem o Grupo I, com estrutura homóloga ao *Pox meso* (*Drosophila*) e ao *AmphiPAX* (*Amphioxus*) que é caracterizado pela presença do domínio *paired*, além de um octapeptídeo (Balczarek e cols., 1997; Breitling e cols., 2000). Curiosamente não apresentam homeodomínio. Os 128 aminoácidos do domínio *paired* dos genes *PAX1* e *PAX9* apresentam alta similaridade (98%; Peters e cols., 1999). Também pouca variação foi observada quando as regiões codificadoras dos dois genes foram seqüenciadas (Hetzer-Egger e cols., 2000). Alguns estudos têm mostrado ainda que *PAX1* e *PAX9* são co-expressos e podem ter funções similares e complementares (Wilm e cols., 1998; Peters e cols., 1999). Já os Grupos II (*PAX2*, *PAX5* e *PAX8*), III (*PAX3* e *PAX7*) e IV (*PAX4* e *PAX6*) apresentam tanto o domínio pareado quanto o homeodomínio. Assim como acontece com o *PAX1* e *PAX9*, a atividade de redundância funcional dos genes parálogos dentro das quatro subfamílias *PAX* parece ser um mecanismo geral que atua de maneira compensatória. Desta forma, a expressão de um contrabalançaria a não expressão de outro, levando a um fenótipo normal ou próximo do normal, pelo menos em algumas circunstâncias (Peters e cols., 1999).

Um conjunto grande de estudos demonstra que estes genes têm um papel importante em inúmeras redes regulatórias (*PAX6* - desenvolvimento ocular; *PAX3* - migração da crista neural; *PAX2* -

sobrevivência e morfogênese de tecidos como do fígado, orelhas e glândulas mamárias; *PAX8* – desenvolvimento da tiróide). Entretanto, a exata extensão das funções dos genes *PAX* em fenômenos biológicos relacionados ao desenvolvimento, sejam eles normais ou patológicos, bem como seus papéis em indivíduos adultos, ainda não é totalmente conhecida (Tabela 2).

TABELA 2. Funções e fenótipos mutantes dos componentes da família *PAX*, em camundongos e humanos.

Grupo	Gene	Domínio de expressão no Embrião	Camundongo		Homem	
			Nulo	Fenótipo Heterozigotos	Homozigotos	Fenótipo Heterozigotos
I	PAX1	esqueleto, 3rd/4th fendas faríngeas, timo	<i>Undulated</i> ,	Normal	Hipoplasia do corpo de dos discos vertebrais	Síndrome de Klippel-Feil Síndrome de Jarcho-Levin
	PAX9	esqueleto, Crânio-facial, timo, glândulas paratiróides, dentes	*	Normal	Palato secundário fendido, duplicação dos dígitos anteriores, perda de dentes timo e Gld paratiróides	Oligodontia, hipodontia
II	PAX2	SNC, Rins	<i>Krd</i> ,	Redução dos rins, coloboma óptico	Sem rins ou trato genital, coloboma óptico	Síndrome Renal-Coloboma
	PAX5	SNC, Células B	*	Normal	Sem células B, Defeitos no Cerebelo e Colliculous inferior	Linfomas
	PAX8	SNC, Rins, Tiróide	*	Normal	Sem tiróide	Displasia da Tiróide, Carcinomas
III	PAX3	SNC, Crista Neural, músculos esqueléticos	<i>Splotch</i>	Falta de pigmentação	Sem músculo nos membros, falta ou hipoplasia dos derivados da crista neural, truncus arteriosus	Síndrome Waardenburg, rhabdomyosarcoma
	PAX7	SNC, músculos esqueléticos, Crânio-Facial	*	Normal	Falta de células musculares satélite, processo maxilar e cápsula nasal;	rhabdomyosarcoma
IV	PAX4	Pâncreas	-	Normal	-	Diabetes
	PAX6	SNC, pâncreas, Olhos	<i>Sey</i> , anirídia, *	Microfítmia corneal e defeito na íris	Sem olhos, bulbo olfatório, perda de estruturas cerebrais;	Anirídia, Cataratas

* mutações alvo – transgênicos;

I-7. PAX9 -Uma pequena proteína com grandes funções

O gene *PAX9* foi mapeado no cromossomo 14 (14q12-q13) em humanos (Stapleton e cols., 1993), mas em camundongos ele se localiza no cromossomo 12. Santagati e cols. (2003) encontraram evidências de que o *PAX9* é herdado em um bloco com outros genes (*Mbip—Nkx2-1—Nkx2-9—Pax9—Slc25a21—Mipol1—Foxa1*), formando um grupo sintênico, parálogo ao do *PAX1* (*Nkx2-4—Nkx2-2—Pax1—Slc25a5l—Foxa2*; Figura 2). Esta sintenia seria conservada, sendo encontrada em mamíferos e peixes. Assim a associação física refletiria um domínio transcricional multigênico, pois estes genes não podem ser separados sem comprometerem suas funções. Existem fortes indícios de que elementos regulatórios do *PAX9* estariam intercalados pelo bloco multigênico, e estes seriam o fator decisivo da conservação da ligação física entre os genes. A presença de seqüências regulatórias intercaladas em regiões multigênicas conservadas representaria um ponto chave na evolução dos genomas.

O *PAX9* possui aproximadamente 15.847 pares de bases (pb) e codifica uma proteína de 341 aminoácidos. Apresenta três introns, que correspondem a maior parte de sua seqüência nucleotídica (13.136 pb), duas regiões não traduzidas 5'UTR e 3'UTR e quatro exons (Figura 3). O primeiro exon codifica unicamente a metionina inicial, os exons 3 e 4 são responsáveis pela região C-terminal da proteína que se supõe ser

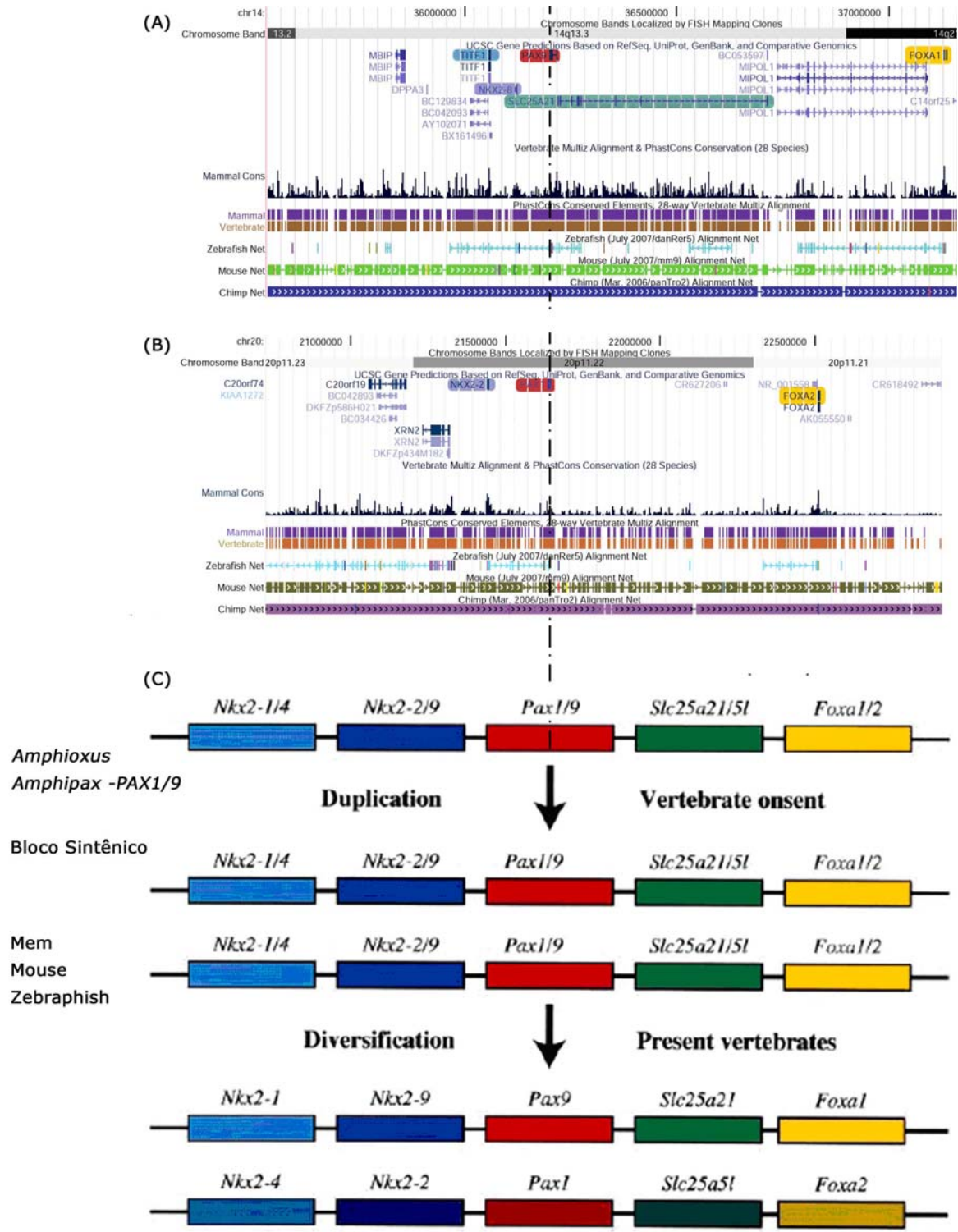


FIGURA 2. Imagem gerada pelo UCSC Genome Browser on Human Mar. 2006 Assembly (<http://genome.ucsc.edu/index.html?org=Human>). (A) Bloco do PAX9; (B) Bloco do PAX1 (C) Bloco sintênico PAX1/9 Santagati e cols. (2003).

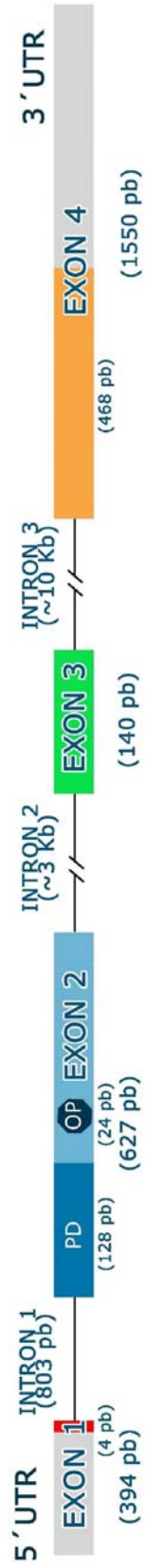


FIGURA 3. Esquema geral da estrutura do gene PAX9. PD : domínio *paired*. OP: octapeptídeo. 5' UTR: região 5' não traduzida. 3' UTR: região 3' não traduzida. Esquema gerado a partir dos dados de Santagati e cols. (2001), modificado de Pereira e cols. (2006).

um domínio de transativação. O exon 2 codifica o octapeptídeo e o domínio *paired*. O domínio *paired* consiste de subdomínios C- e N-terminais que são estruturalmente independentes. Na metade da região N-terminal se formam os motivos hélice-alça-hélice e um motivo alça- β que conferem à proteína a capacidade de se ligar ao DNA (Hetzer-Egger e cols., 2000; Stockton e cols., 2000).

O fator de transcrição codificado pelo *PAX9*, como a maioria dos fatores de transcrição de eucariotos, provavelmente regula a expressão de dezenas de outros genes. Sua função tem sido considerada fundamental durante o desenvolvimento embrionário, pois atua na formação dos arcos branquiais, timo, glândulas paratireóides e ossos (vértebras, membros, face) (Peters e cols., 1997; Hetzer-Egger e cols., 2000; Santagati e cols., 2003). As investigações nesta área têm mostrado ainda o papel essencial deste gene no desenvolvimento de uma estrutura específica dos vertebrados, que na maioria das espécies só estará presente após o nascimento: os dentes (Peters e cols., 1998; 1999; Frazier-Bowers e cols., 2002; McKeown e cols., 2002).

I-7.1 O gene PAX9 e o desenvolvimento dentário

O desenvolvimento dos dentes em vertebrados é um processo regulado temporal e espacialmente, devido a interações recíprocas entre células ectodermais e mesenquimais. Este processo envolve mais de duas dezenas de genes, que codificam uma variedade de fatores de transcrição, bem como de outras proteínas envolvidas na cascata molecular necessária para o desenvolvimento dentário normal (Peters e Balling, 1999; Chen e cols., 2000; Stockton e cols., 2000; Thesleff, 2000; Line, 2001; Frazier-Bowers e cols., 2002; McKeown e cols., 2002; Kist e cols., 2005; Mostowska e cols., 2006).

Vários estudos têm mostrado ainda que os efeitos combinatórios de um grupo de genes mestres formam gradientes e geram sub-regiões específicas. Esta rede regulatória é responsável por uma segmentação molecular, que resulta na formação, um tanto quanto independente, dos diferentes tipos de dentes encontrados, por exemplo, nos mamíferos (molares, pré-molares, caninos e incisivos). A perda total de um conjunto de dentes, mas não de outros, pode desta forma, ser explicada por esta segmentação genética ou modularidade (Peters e Balling, 1999; Chen e cols., 2000; Line, 2003; Kist e cols., 2005; Peres, 2005).

Os estudos de mutações que provocam agenesia de dentes em humanos e em camundongos (Peters e Balling, 1999; Chen e cols.,

2000; Kist e *cols.*, 2005) conjuntamente fornecem sólidas evidências de que o gene *PAX9* teria papel chave e destacado na rota genética da odontogênese. Mais que isso, demonstra através de abordagens que utilizam modelos animais, que alterações no *PAX9* estariam definitivamente relacionadas ao estabelecimento de diferentes fórmulas dentais, o qual estaria relacionado com as forças direcionadoras que permearam a diversificação de inúmeras espécies, incluindo os mamíferos (Kist e *cols.*, 2005).

I-7.2 Evolução molecular do PAX9 na ordem Primata

O modelo primitivo de dentição dos primatas inclui, em um determinado quadrante dois incisivos, um canino, três pré-molares e três molares, num total de 36 dentes. Este padrão ainda é conservado nos lêmures e nos macacos do Novo Mundo, enquanto nos grandes macacos e homem são encontrados apenas dois pré-molares, totalizando 32 dentes (Lewin, 1999). Já os terceiro-molares em humanos são freqüentemente impactados, o que já não ocorre nos grandes macacos (Varki e Altheide, 2005). Além disso, observa-se grande variação na agenesia de terceiros molares dentro de nossa espécie (Rozkovcová e *cols.*, 1999).

Considerando que mutações em genes importantes podem ser responsáveis por grandes mudanças morfológicas (Nei, 2007), Pereira e cols. (2006) avaliaram a possibilidade do gene *PAX9* estar envolvido em alterações dentárias encontradas na nossa e em outras espécies de primatas. Para isso foram seqüenciados todos os 4 exons do gene *PAX9* (1026 pares de bases) em 86 indivíduos (europeus, asiáticos e ameríndios). Neste trabalho foi possível demonstrar que as variações na nossa espécie estavam praticamente limitadas ao exon 3. Análises envolvendo seqüências ortólogas do *PAX9* disponível para outros organismos também foram consideradas. Os resultados mostraram que seleção purificadora estaria atuando no *PAX9*. Merece destaque, no entanto, que nos humanos foi detectada uma mutação G → C no exon 3 que resulta numa mudança não sinônima (alanina → prolina) na posição 240 da cadeia de aminoácidos, com distribuição polimórfica em todos os grupos humanos estudados. Foi sugerido pelos autores que esta mutação representaria um polimorfismo transiente em nossa espécie e que a possível ação de seleção positiva poderia estar sendo mascarada. A presença do aminoácido prolina, numa região de alfa-hélice, poderia ser detrimental para a proteína em si, mas vantajosa em nível do organismo já que poderia levar a agenesia dos terceiros – molares. Assim, a presença do alelo derivado C/Pro resultaria numa redução no número de dentes e isso poderia ter tido alguma vantagem

ao longo da trajetória evolutiva dos hominídeos, particularmente após mudanças importantes da dieta, como o hábito de comer alimentos cozidos (Pereira e cols., 2006). Estes resultados demonstraram que um gene conservado pode deixar de sê-lo em algum (ns) ramo(s) de uma dada filogenia. Além disso, o exon 3 poderia ser uma janela evolutiva num gene extremamente conservado. O conceito da janela evolutiva implica numa situação onde as mutações em um fator de transcrição incidem somente sobre um exon específico, pois esta região não estaria sob pressão de restrições funcionais. O relaxamento da seleção natural sob este exon em especial e a preservação de outras regiões do gene promoveria, desta forma, a manutenção da função original, promover a transcrição de outros genes, mas poderia alterar o seu leque de atuação. Este último aspecto poderia representar uma forma alternativa de evolução molecular dentro de uma família gênica. Normalmente, a proliferação que resulta em famílias gênicas fornece material para inovações funcionais, desde que após a duplicação de um dado gene, uma de suas cópias preserve sua função original enquanto outra acumule mutações que sob determinadas circunstâncias poderiam ser selecionadas e fixadas (Wang e Gu, 2001). Foi demonstrado também que determinados exons de um gene específico também poderiam estar “livres” para acumular mutações, desde que os outros exons do gene preservassem algumas das funções primordiais do gene como um todo

(Pereira e cols., 2006). Em fatores de transcrição, como é o caso do *PAX9*, tais alterações manteriam a estrutura tridimensional e não acarretariam modificações nos sítios ativos, preservando a função geral da proteína. Contudo, sutis alterações seriam suficientes para interromper a interação entre o fator de transcrição e alguns genes ou proteínas de sua via de atuação, gerando epistasia restrita. Como resultado, fenótipos específicos seriam gerados.

Outro aspecto que merece destaque no trabalho de Pereira e cols. (2006) comentado é a possibilidade de existirem mutações fixadas no exon 3 em macacos do Novo Mundo, particularmente uma, a substituição de uma alanina por uma treonina no resíduo 253 da cadeia de aminoácidos. O aminoácido treonina mostrou estar fixado em todas as espécies de macacos estudadas por Pereira e cols. (2006). Recentemente, a possibilidade deste "alelo" estar fixado nos macacos do Novo Mundo ganhou força com o trabalho de Perry e cols. (2006). Os últimos autores investigaram inúmeras espécies de primatas com relação ao gene *PAX9* e somente nos macacos do Novo Mundo aparece esta alteração.

Esse conjunto de estudos envolvendo o *PAX 9* mostra a amplitude de possibilidades envolvendo esse gene e o quanto ele pode estar envolvido em questões instigantes relativas a modificações morfológicas na nossa e em outras espécies.

I-8. Problemas iguais, soluções semelhantes. Mesmas ferramentas?

Similaridades morfológicas entre organismos vivos na maior parte das vezes reportam uma trajetória evolutiva em comum. Entretanto, muitas destas similaridades são observadas em espécies não relacionadas, mas que compartilham os mesmos hábitos de vida. Por exemplo, pangolins e tatus/tamanduás pertencentes a distintas ordens (*Pholidota* e *Xenarthra*, respectivamente) estão entre as mais peculiares espécies de mamíferos, possuindo em comum, únicas e especializadas características anatômicas adaptadas para comer insetos (Yang e cols., 2006).

Do ponto de vista taxonômico, convergência seria o desenvolvimento de características similares separadamente em duas ou mais linhagens sem um ancestral comum que justificasse a similaridade. Em contraste, o paralelismo é a ocorrência independente de mudanças similares em grupos com um ancestral comum. Segundo Wood e cols. (2005), em ambos os casos, a seleção natural seria a responsável pela recorrência da(s) característica(s) em comum, porém não podemos descartar totalmente o papel das forças do acaso agindo sob o desenvolvimento. De forma similar, sob a perspectiva da biologia molecular mudanças nucleotídicas ou aminoacídicas a partir de um mesmo estado ancestral que resulta no mesmo estado derivado é

referido como paralelismo, enquanto mudança de diferentes estados ancestrais para um estado derivado comum é referida como convergência (Zhang e Kumar, 1997).

Uma visão evolutiva busca integrar a análise das características compartilhadas com o estabelecimento de como, quando e onde surgiu a variante genética. Nosso interesse é abordar as possíveis conexões entre genótipo e fenótipo, particularmente quando estas se enquadram no contexto de paralelismo/convergência, ambos os fenômenos mediados principalmente pela ação da seleção natural, já que não é esperado que eventos estocásticos levem ao surgimento de fenótipos complexos repetidamente (Wiens e cols., 2003; Wood e cols., 2005). A pressão de seleção pode, inclusive, ser a mesma ao longo de uma escala de tempo, já que extrema convergência de planos corporais tem sido descrita em diferentes *taxa* que habitaram a mesma zona adaptativa, porém em épocas distintas (Nesbitt e Norell, 2006). A similaridade da forma do corpo e dos dentes dos ictiossauros e golfinhos ilustram bem este fenômeno.

Melville e cols. (2006) recentemente discutiram um aspecto extremamente interessante envolvendo convergência adaptativa. Quando diferentes comunidades habitam ambientes similares, a convergência evolutiva é esperada. Entretanto, se a contingência histórica (eventualidade, acaso; Pena, 2006 e citações lá contidas)

tivesse um papel preponderante na trajetória evolutiva de um grupo em particular, a convergência evolutiva seria improvável e não freqüente (Melville e *cols.*, 2006). Estudando lagartos de desertos na Austrália e nos Estados Unidos, os autores mostraram como forças evolutivas determinísticas moldaram os padrões morfológicos nestas populações, sobrepondo-se de maneira avassaladora às eventuais contingências históricas únicas relacionadas a cada linhagem em particular.

Sendo assim, a recorrência de similaridades em distintas linhagens evolucionárias tem sido descrita em vários níveis de organização biológica desde seqüências de nucleotídeos, passando por alterações cromossômicas até chegar a características fenotípicas complexas (Wood e *cols.*, 2005; Yang e *cols.*, 2006). No entanto, a conexão entre os mecanismos moleculares e/ou cromossômicos envolvidos e responsáveis pelos inúmeros casos de mudanças fenotípicas, que evoluíram em paralelo na natureza, permanece desconhecida (Colosimo e *cols.*, 2005).

Diversos estudos tentam estabelecer uma conexão entre a convergência morfológica ou ainda a evolução em paralelo com alterações moleculares. Trabalhos recentes como o de Protas e *cols.* (2006) que estudaram duas espécies de peixes do gênero *Astyanax* (habitantes de águas escuras e gélidas em cavernas subterrâneas) concluíram que uma de suas características morfológicas mais

marcantes, o albinismo, presente em ambas as espécies, era produto de mutações independentes no mesmo gene. Mais que isso, os autores concluíram que a perda da função da proteína *OCA2* era decorrente de deleções em exons/intron diferentes no gene *OCA2* em cada uma das espécies. O gene *MCR1* responsável pela coloração em mamíferos (Majerus e Mundi, 2003) e aves (Hoeskstra e Price, 2004) é também descrito como um exemplo de evolução paralela.

Além disso, foi visto ainda que a ocorrência de mutações em sítios regulatórios de genes do desenvolvimento estaria associada a mudanças fenotípicas recorrentes encontradas em diferentes espécies de *Drosophila* (Sucena e cols., 2003) e borboletas (Brakefield, 2006).

Estes trabalhos apontam para a idéia de que as mesmas soluções genéticas estariam sendo “privilegiadas” em espécies diferentes sujeitas às mesmas pressões seletivas, já que mutações no mesmo gene seria o modo mais otimizado, ou ainda, o único modo, de um fenótipo em particular evoluir de forma independente (Majerus e Mundi, 2003; Sucena e cols., 2003; Hoeskstra e Price, 2004; Wood e cols., 2005; Brakefield, 2006; Protas e cols. 2006).

É o caso já citado das espécies popularmente conhecidas como pangolins (Ordem *Pholidota*), tatus e tamanduás (Ordem *Xenarthra*). Estes animais compartilham especializações adequadas ao seu hábito de capturar e ingerir formigas e cupins, que incluem mudanças na

mandíbula, na forma e no comprimento do crânio, bem como na redução do número e no tamanho de dentes.

As especializações no aparelho mastigatório, que facilitaram a captura e o processamento dos alimentos, representaram uma inovação chave na evolução dos mamíferos. Deste modo, e considerando o que sabemos sobre o papel do *PAX9* no desenvolvimento de fórmulas dentais, é possível especular que mutações independentes neste gene possam estar envolvidas na cascata de modificações que levaram a morfologia similar como àquelas encontradas em espécies distantes filogeneticamente, mas que apresentam em comum, adaptações no crânio e, particularmente, nos dentes como aquelas citadas acima.

CAPÍTULO II - Objetivos

Os objetivos do presente projeto são apresentados a seguir:

- 1) Testar com uma amostra maior a hipótese apresentada em Pereira e cols. (2006) de que na nossa espécie poderia estar havendo seleção positiva para o alelo derivado C/Pro, e adicionalmente delinear um cenário mais específico que explique a possível ação dessa força evolutiva;
- 2) Avaliar a variabilidade das regiões codificadoras do *PAX9* em diversas espécies de mamíferos;
- 3) Investigar a possibilidade da convergência morfológica encontrada em algumas das espécies investigadas no presente estudo ser reflexo de uma convergência molecular envolvendo o gene *PAX 9*;
- 4) Contribuir para o delineamento do cenário evolutivo em nível molecular do gene *PAX9*.

CAPÍTULO III – Resultados

III-1- Manuscrito em preparação

Os resultados obtidos referentes às populações humanas investigadas e que buscam responder ao primeiro objetivo da presente dissertação podem ser vistos no manuscrito (em preparação) que segue.

PAX9 exon 3 variability in Native Americans and in major human groups: inter-population and intra-gene evolutionary dynamics.

V. Paixão-Côrtes, S.L. Bonatto, I. Bandeira, S. Mazières, F.M. Salzano, D. Meyer, and Bortolini M. C.

Abstract

A total of 125 persons from 10 South Amerindian and one Eskimo populations were studied in relation to the Paired box gene 9 (*PAX9*) exon 3 (138 base pairs) as well as its 5' and 3' flanking intronic segments (232 pb and 220 pb, respectively) and integrated with the information available for the same genetic region from 115 individuals of different geographical origin. Nine mutations were scored in exon 3 and four in its flanking regions; four of them are new South American tribe-specific singletons. The Ala240Pro mutation shows a wide distribution in South Amerindians and is present in Eskimo, Asian, European or Euro-derived persons; however it was not present among Afro-Americans. Exon3 nucleotide diversity is several orders of magnitude higher than their intronic borders, and F_{ST} values are different than those obtained with neutral markers. The pattern that emerges is the increase in frequency of the Ala240Pro derived allele from Africa to other continents probably as a result of positive selection, which could be related to third molar agenesis or an associated gene connected to environmental adaptation.

Introduction

The central goal of genetics is to understand how heritable information encoded in the genome determines the phenotype of an organism (Lee *et al.*, 2008). Some results support the idea that morphological evolution can occur basically by a small proportion of major effect mutations in structural or regulatory genes (major gene effect hypothesis; Nei 1987, 2007). Considering this possibility, we recently studied the entire coding region and exon-intron boundaries of the Paired box homeotic gene 9 (*PAX9*), and highlighted the importance of small changes in major protein-coding genes that might have been recently selected positively in the human lineage only (Pereira *et al.*, 2006).

The *PAX9* gene codes for a transcription factor that has demonstrated to be essential in the embryogenesis of vital organs and in the development of the mammalian dentition (Peters *et al.*, 1998). In this respect, although a number of mutations have been associated with both nonsyndromic molar hypodontia and oligodontia (Stokton *et al.*, 2000; Nieminen *et al.*, 2001; Das *et al.*, 2002; Pereira *et al.*, 2006; Tállon-Walton *et al.*, 2007), evidence regarding a nearly complete lack of variation in exons 1, 2 and 4 of the human *PAX9* are unequivocal. Conversely, we showed that exon 3 presents high variability in our species, which can be regarded as an “evolutionary window” in an extremely

conserved gene. Additionally we postulated the hypothesis that this exon might be a target for a recent positive selection (Pereira *et al.*, 2006).

Here we present sequence information on *PAX9* exon 3 and its bordering intronic regions in nine Native American, and one Eskimo populations, and integrated our data together with those earlier published involving other major human geographic groups. The objective was to extend our knowledge about the genetic variability of the human *PAX9* exon 3 in an attempt to better understand the contribution of non-random and random microevolutionary processes in its existence, pattern and maintenance.

Materials and Methods

Populations Examined

PAX9 exon 3 and its flanking intronic sequences were obtained from 125 individuals from ten South Amerindian (N = 81) and one Siberian Eskimo (N = 44) populations. Additional nucleotide sequences of the same DNA region from 15 Europeans, 14 Japanese, and 57 South Amerindians were compiled from our previous study (Pereira, 2006), whereas data from 14 Euro-Americans and 15 African Americans were obtained through the HAPMAP (<http://www.hapmap.org/cgiperl/>)

[gbrowse/hapmap_B35/](#)) and dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP/>) databases. Relevant information about the Amerindian and Eskimo populations can be seen in Table 1.

The investigations with these Native Amerindian samples were approved by the Brazilian National Ethics Commission (CONEP Resolution nº 123/1998).

DNA Extraction, PCR Amplification and Sequencing

Genomic DNA was extracted from plasma and glycerolized red blood cells stored in our laboratory as a result of previous studies, listed in Table 1, using the QIAamp DNA MiniKit (Qiagen) and following the manufacturer's instructions.

PAX9 exon 3 (138pb) and its 5' and 3' flanking intronic segments (232pb and 220pb, respectively) were amplified using primers and conditions previously described in Pereira *et al* (2006). The PCR products were purified following the Exonuclease I and Alkaline Phosphatase (Amersham Biosciences) protocols.. Both DNA strands were sequenced using MegaBace500 and MegaBace1000 (Amersham Biosciences) sequencing machines. When a possible mutation was found, the sample was re-amplified and re-sequenced at least once again to confirm it.

Data Analysis

Sequences were aligned and their quality, as well as the assessment of the accuracy of the resulting data, were obtained using the Phred, Phrap and Consed programs (<http://www.genome.washington.edu>). Additionally, all chromatograms were visualized and checked manually.

Population parameters such as nucleotide diversity (π) (Nei, 1987) and the degree of population structure (F_{ST} ; Weir and Cockeran, 1984; Excoffier *et al.*, 1992) were obtained employing the Arlequin 3.11 software (<http://lgb.unige.ch/arlequin>). To detect departures from a standard neutral model, Tajima's D and Fu and Li's D* and F* tests (Mega 4.0 and DnaSP 4.10.9) were performed. Both positive and negative significant values for these statistics cause the rejection of the null hypothesis of mutation-drift equilibrium. However, it is well-known that these parameters are also affected by the demographic histories of the population considered (Bamshad and Wooding, 2003). Therefore, we incorporated in the analysis coalescence simulations using Rogers' algorithm (Rogers, 1995) as implemented by Wooding *et al.* (2004) in the DFSC 1.1 software (www.xmission.com/~wooding/DFSC/). Distinct demographic scenarios compatible with the available data about times of expansion and initial population sizes were simulated as follows: initial population of 1,000 or 10,000 individuals (Liu *et al.*, 2006; Fagundes *et*

al., 2007) for all samples, 1- to 100-fold expansions, starting 60,000 years ago (Alonso and Armour, 2001; Santos-Lopes *et al.*, 2007; Fagundes *et al.*, 2007). For the Native American and Eskimo dataset, initial populations of 80 or 500 individuals (Hey, 2005; Fagundes *et al.*, 2008), 1 – 100-fold expansion, starting 18,000 years ago (Fagundes *et al.*, 2008); whereas for Europeans and Japanese we considered an initial population size of 10,000 and 3,000, 1 – 100 fold expansions, starting 40,000 and 37,000 years, respectively (Cavalli-Sforza *et al.*, 1994). The mutation rate (μ) used to obtain the θ values under the different scenarios was 1×10^{-9} (Wooding *et al.*, 2004).

A synthetic map of the Old World considering the agenesi s of third molar frequencies was obtained using an interpolation or smoothing algorithm furnished by the UNIRAS A/S SOBERG (Denmark) program. Frequencies were added to geographical coordinates and a color-graded distribution (Piazza *et al.*, 1981) obtained.

Results

Mutations and polymorphic sites

A total of 590 base pairs, corresponding to *PAX9* exon 3 and its 5' and 3' flanking intronic regions, of 250 human chromosomes were

sequenced. The data were analyzed together with 172 other sequences published earlier (HAPMAP and dbSNP web sites; Pereira *et al.*, 2006). The results shown in Table 2 reveal three substitutions in the 5' Flanking and one in the 3' Flanking intronic regions, in the Apalaí, Mekranoti and Galibi populations. Only one of these, the variant allele A at chromosome 14 in position 36,205,241 was present in homozygous state in one Galibi individual. Overall, the sequence alterations observed in the human exon 3 were in heterozygous state with two exceptions: variants located at codons 239 and 240 (C→T, His239His; and G → C, Ala240Pro; see below).

The G → A (Ser226Asn) transition, G →A (Ala236Thr) transition, T → A (Asn242Asp) transversion, and A → T (Glu250Val) transversion are new South American tribe-specific singletons (Apalaí, Tenharim and Arara and tribes; Table 2). The G → A (Ser226Asn) mutation is located in the same position as an earlier described among the Bari from Venezuela, which also caused an amino acid replacement (G →T, Ser226Ile). Another C → A (Gln251Lys) mutation, previously seen exclusively in the Guarani (Pereira *et al.*, 2006) was not observed in the present investigation. On the other hand, the derived T allele of the His239His SNP earlier observed only in the Kaingang from southern Brazil (Pereira *et al.*, 2006) was now also found in the Eskimo sample. Since this variant was also observed with high frequencies in European

(27%) and Euro-derived (21%) populations, its presence among the Kaingang (10%) could be due to admixture with non-Indians, which has been reported in earlier studies with this population (Marrero et al., 2007).

The substitution of a G for A at chromosome 14 position 36,205,504, which results in an amino acid change at residue 240, was detected in the Munducuru, Tenharim and Xikrin. Curiously, in this same chromosome location variant C of the Ala240Pro SNP can be found. Our results indicate that the *C/Pro* allele is widely distributed in South Amerindians living in distant geographic regions and speaking different linguistic stocks, as well as in Eskimos, suggesting that it came from Asia to the New World with the first migrants, whereas *A/Thr* seems to be an American autochthonous mutation, but with relatively ancient origin, since it is present in individuals who speak languages classified in the two major South Amerindian linguistic families (Tupian and Jêan; split at least ~5 thousand years before present; Urban, 1998). Finally, although allele *T* at the His239His SNP is only present in populations where the Ala240Pro allele *C* is also found (Table 2), they are not associated.

To better explore the results obtained with the Ala240Pro SNP, detailed information on its genotype and allele frequencies is given in Table 3. Although sample sizes are small, (*C/Pro*) frequencies vary

considerably over the 15 South American tribes analyzed: values range from zero (in seven populations) to 40% (Jamamadi). The highest *C/Pro* frequencies can be found in Eskimo, European and Euro-derived populations ($\sim 30\%$), while it was not detected in Afro-descendants from the USA. No *PAX9* data for Africans are available yet, but the absence of *C/Pro* in Afro-Americans can indicate that it is virtually absent, or at least rare, in sub-Saharan populations.

Genetic variability

Nucleotide diversity (π) values for *PAX9* exon 3 and its intronic bordering regions are summarized in Table 4. Considering all the 590 base pairs and exon 3 only, π ($\times 10^4$) for the Amerindian sample as a whole (4.7 and 17.4, respectively) are lower than those obtained for the Eskimo (8.7 and 37.2) and European/Euro-descendants (14.3/56.1 and 13.1/56.0), contrasting with the complete absence of diversity in Afro-Americans. Furthermore, π values for exon 3 are generally several orders of magnitude higher than those calculated for the non-coding segment in all populations.

The level of interpopulation differentiation obtained among the Native Americans, considering the coding portion only (F_{ST} : 7.8; data not shown), although high since it is similar to that calculated for neutral

markers (678 autosome microsatellite loci: F_{ST} : 8.1; Wang et al., 2007), is lower than that obtained when all other human populations were compared (F_{ST} = 18.6; data not shown). This last value, on the other hand, contrasts with the degree of population differentiation observed at the worldwide level by Wang et al. (2007), F_{ST} : 7.1, with the same set of microsatellites indicated above.

Studies with neutral genetic loci have generally shown that Native Americans present lower genetic diversity (as measured by π and other heterozygosity indices) and higher levels of population structure (as determined using F_{ST} or other similar statistics) than those seen in populations/groups from other continents (Cavalli-Sforza et al., 1994; Bortolini and Salzano, 1996; Wang et al., 2007). The opposite is generally found when Sub-Saharan populations are investigated (Rosenberg et al., 2002; Wang et al., 2007). These diversity/divergence patterns are mainly due to demographic processes (successive founder effects with later expansions) related to the dispersal of modern *Homo sapiens* from Africa to other continents (Alonso and Armour, 2001; Ramachandran et al., 2005; Fagundes et al., 2007; Liu et al., 2007; Santos-Lopes et al., 2007; Wang et al., 2007). Since demographic effects affect the whole genome, deviations from this classical diversity/divergence model, as that observed here, is expected when portions of the genome are under natural selection pressure

(Bamshad and Wooding, 2003; Romeo *et al.*, 2007; Barreiro *et al.*, 2008).

Third molar agenesis and the Ala240Pro polymorphisms

Physical anthropological studies, based on large samples, determined that agenesis of the third molar vary within and between human groups, although part of these findings can be attributable to the absence of panoramic X-ray examinations (Bredy *et al.*, 1991; Irish, 1993, 1997, 2003; Turner, 1986, 1987, 1990; Rozkovicova *et al.*, 1999; 2004; Delgado-Bueno, 2007). The main result of these anthropological investigations, however, has little possibility to be due to methodological problems: a tendency of the ancestral dental trait, third molar presence, generally decreases, whereas its absence increases in frequency from Sub-Sahara to Europe and Asia (Figure 1). Interestingly, if we assume that the *C/Pro* allele in the Ala240Pro SNP is absent or rare in sub-Saharan populations as it is in Afro-Americans, our results are suggesting that *C/Pro* increases in frequency from Africa to other continents, in opposition with the presence of third molar cline.

The anthropological studies conducted with Native Americans reveal that missing third molars are relatively frequent among Eskimos (~ 20%), but it is difficult to establish a general pattern in relation to

Native America as a whole due to the large amount of variation (range from zero to $\sim 40\%$; Gómez-Valdés *et al.*, 2005; Rozkovcova *et al.*, 1999; Turner, 1981, 1986; Sutter, 2005; Hydenblit, 1996). In a way this resembles the heterogeneous Ala240Pro allele distribution presented in Table 3. But of course, other known and presently unknown genes, as well as epigenetic factors may influence tooth development processes (Tállon-Walton *et al.*, 2007; Swinnen *et al.*, 2008)

Demographic simulations and neutral/natural selection tests

To better investigate the possibility of departures from a standard neutral model with the *PAX9* human data, we used Tajima's D , and Fu and Li's D^* and F^* statistics, and afterwards considered them under the assumption of several demographic scenarios. The strong negative values found for South Amerindians (Table 5 and Figure 2) are probably inflated by demographic processes, since their significance is lost when scenarios of bottlenecks with later expansion are simulated (Table 6). Note, however, that for the other groups practically all values are positive and significant under the different population growth scenarios, considering the exon 3 datasets only (Tables 5 and 6).

Discussion

Even considering that many polymorphisms are eliminated or fixed as a consequence of genetic drift, it is expected that part of the human molecular variation is result of varied selective pressures in different geographical regions, each with its own different climate, pathogens and sources of food (Bamshad and Wooding, 2003; Harris and Meyer, 2006; Hughes *et al.*, 2008). Additionally, natural selection act differently in diverse genome segments; while portions of a given gene, for example, are largely invariable due to strong negative selection, others vary due to adaptative selection, relaxation of purifying selection, or balancing selection, among others (Yang *et al.*, 2000; Hughes, 2007). Our findings show that the human *PAX9* is a good example of this inter-population and intra-gene evolutionary dynamics.

Evolutionary conservation is a potentially powerful method to identify functional regions of the genome (Drake *et al.*, 2006). Although introns are subjected to less evolutionary constraint than coding sequences, they may not evolve neutrally for many reasons, including maintenance of the splicing efficiency or alternative splicing, as well as conservation of transcription factor binding sites (Majewski and Ott, 2002; Waterston *et al.*, 2002; Hare and Palumbi, 2003; Sorek and Ast, 2003; Bejerano *et al.*, 2004; Voelker and Berglund, 2007; Hughes *et al.*, 2008). The high intron conservation in the *PAX9* gene had been

demonstrated earlier (Pereira *et al.*, 2006). Here we detected Tajima's D , and Fu and Li's D^* and F^* negative values considering the 5' and 3' flanking intronic regions in all populations where variation was observed (Table 5). Although these numbers are not significant under population increase demographic scenarios, the tendency for negative values can be interpreted as a signature of functional constraint maintained by purifying selection.

On the other hand, the level of *PAX9* exon 3 nucleotide diversity and divergence to in our sample as a whole ($\pi = 27.4 \times 10^{-4}$) is about four times higher in comparison with that reported for the coding region of other human genes ($\pi = \sim 7.0 \times 10^{-4}$; Li and Sadler, 1991). The differentiation among all populations, including Native Americans (F_{ST} : 11%; data not shown) is also higher than that estimated for neutral markers ($F_{ST} = 7.1\%$; Wang *et al.*, 2007). The lack of intra and inter population conservation of a specific exon in a development gene is surprising. Our analyzes the human *PAX9* exon 3 together with those orthologous from 13 other vertebrate species demonstrated that the driving force behind the evolution of this coding region is apparently under intense purifying selection (Pereira *et al.*, 2006). But, positive selection effects are often masked by negative selection (Fay *et al.*, 2002) and the former could be occurring in human lineages (Pereira *et al.*, 2006). The Tajima's D and Fu and Li's D^* and F^* positive estimates

considering exon 3 only, plus the F_{ST} results (positive selection tends to increase F_{ST} ; Barreiro et al., 2008) obtained here vigorously reinforce this suggestion. The central role of the Ala240Pro polymorphism in this context is also demonstrated since when we excluded it the significance of the neutrality tests is lost in practically all simulated scenarios (data not shown). The sliding window method (Figure 2) was also used to illustrate this situation.

Our earlier results also suggested that due to the dramatic changes in the way of living and diets that occurred in the human lineage, reduction in the number of teeth could be advantageous, especially due to third molar implantation problems (Pereira *et al.*, 2006). The data presented in present study seems to indicate that this view is correct and additionally permit us to hypothesize a possible sequence of events to explain them as a whole. We have to assume that initially there was selection relaxation on exon 3, allowing new mutations to be maintained there. It is possible that the G → A Ala240Pro mutation occurred in Africa, but only after the human migration from there it acquired adaptative value. The Asian and European high frequencies of the derived allele are in accordance with this suggestion. In relation to South Amerindians, genetic drift and cultural changes could have disturbed or reversed the postulated adaptative advantage of the *C/Pro* allele. It is also possible that the

reduced number of teeth is not the main factor that is subject to positive selection. The *C/Pro* allele could be associated with variants connected with cold adaptation or others, necessary for the Out-of-Africa migrants. Several alleles selected for conditions different from those present in equatorial Africa were recently identified (Harris and Meyer, 2006; Hancock *et al.*, 2008; Hughes *et al.*, 2008).

Acknowledgments

Thanks are due to the people who, after appropriate information, donated blood for this investigation. The logistic support of the National Indian Foundation (Fundação Nacional do Índio-FUNAI) in the Brazilian field work is also gratefully acknowledged. Financial support was provided by the Institutos do Milênio and Apoio a Núcleos de excelência Programs, Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul. We also thank T.V. Pereira for previous discussions of the manuscript, and V. A. Alonso, M.E. Delgado-Burbano, J. Gómez-Valdez for sources of information about third molar agenesis.

Literature Cited

- Alonso S, Armour JA (2001) A highly variable segment of human subterminal 16p reveals a history of population growth for modern humans outside Africa. *Proc Natl Acad Sci U S A* 98(3):864-9.
- Bamshad M, Wooding SP (2003) Signatures of natural selection in the human genome. *Nat Rev Genet* 4(2):99-111.
- Bamshad MJ, Wooding S, Watkins WS, Ostler CT, Batzer MA, Jorde LB (2003) Human population genetic structure and inference of group membership. *Am J Hum Genet* 72(3):578-89.
- Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L (2008) Natural selection has driven population differentiation in modern humans. *Nat Genet*. Epub ahead of print.
- Bejerano G, Haussler D, Blanchette M (2004) Into the heart of darkness: large-scale clustering of human non-coding DNA. *Bioinformatics* 20(1):40-48.
- Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D (2004) Ultraconserved elements in the human genome. *Science* 304(5675):1321-1325.
- Belich MP, Madrigal JA, Hildebrand WH, Zemmour J, Williams RC, Luz R, Petzl-Erler ML, Parham P (1992) Unusual HLA-B alleles in two tribes of Brazilian Indians. *Nature*. 357(6376):326-329.
- Bredy E, Erbring C, Hübenthal B (1991) The incidence of hypodontia with the presence and absence of wisdom teeth. *Dtsch Zahn Mund Kieferheilkd Zentralbl* 79(5):357-363.
- Bortolini MC, Salzano FM (1996) mtDNA diversity analysis in Amerindians and other human populations – how different are they? *Brazilian Journal of Genetics*. 19(3):527-534.
- Bortolini MC, Salzano FM, Zago MA, Da Silva Júnior WA, Weimer Tde A (1997) Genetic variability in two Brazilian ethnic groups: a comparison of mitochondrial and protein data. *Am J Phys Anthropol*. 103(2):147-156.
- Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SP, Bau CH, Hutz MH, Layrisse Z, Petzl-Erler ML, Tsuneto LT, Hill K,

- Hurtado AM, Castro-de-Guerra D, Torres MM, Groot H, Michalski R, Nymadawa P, Bedoya G, Bradman N, Labuda D, Ruiz-Linares A (2003) Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet.*73(3):524-539.
- Brown ID (1974) The unpredictable lower third molar. A case report. *Br Dent J.* 136(4):155-156.
- Casillas S, Barbadilla A, Bergman CM (2007) Purifying selection maintains highly conserved noncoding sequences in *Drosophila*. *Mol Biol Evol.* 24(10):2222-2234.
- Cavalli-Sforza LL, Menozzi P, Piazza A. (1994) The history and geography of human genes. Princeton: Princeton University Press.
- Cunha MC. 1992. História dos Índios do Rio Grande do Sul. São Paulo, Brasil: Companhia das Letras. p 608.
- Das P, Stockton DW, Bauer C, Shaffer LG, D'Souza RN, Wright T, Patel PI (2002) Haploinsufficiency of PAX9 is associated with autosomal dominant hypodontia. *Hum Genet* 110(4):371-376.
- Drake JA, Bird C, Nemesh J, Thomas DJ, Newton-Cheh C, Reymond A, Excoffier L, Attar H, Antonarakis SE, Dermitzakis ET, Hirschhorn JN (2006) Conserved noncoding sequences are selectively constrained and not mutation cold spots. *Nat Genet* 38(2):223-227.
- Delgado-Burbano ME (2007) Population affinities of African Colombians to Sub-Saharan Africans based on dental morphology. *Homo.* 58(4):329-356.
- Erdesz S, Shubin SV, Shoch BP, Krylov My, Mylov NM, Chekalina NA, Benevolenskaya LI (1994) Spondyloarthropathies in circumpolar populations of Chukotka (Eskimos and Chukchi): epidemiology and clinical characteristics. *J Rheumatol.* 21(6):1101-1104.
- Excoffier L, Smouse PE, Quattro JM (1992) Analysis of molecular variance inferred from metric distances among DNA haplotypes: application to human mitochondrial DNA restriction data. *Genetics* 131(2):479-491.

- Fagundes NJ, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L (2007) Statistical evaluation of alternative models of human evolution. *Proc Natl Acad Sci U S A* 104 (45):17614-17619.
- Fagundes NJ, Kanitz R, Eckert R, Valls AC, Bogó MR, Salzano FM, Smith DG, Silva-Jr WA, Zago MA, Ribeiro-dos-Santos AK, Santos SE, Petzl-Erler ML, Bonatto SL (2008) Mitochondrial population genomics supports a single pré-Clovis origin with a coastal route for the peopling of the Américas. *Am J Hum Genet*.
- Fay JC, Wyckoff GJ, Wu CI. (2002) Testing the neutral theory of molecular evolution with genomic data from *Drosophila*. *Nature*. 28;415(6875):1024-1026.
- Gómez-Valdés J (2005) Antropología dental en poblaciones del Occidente de Mesoamérica. Tesis para optar al grado de Licenciado en Antropología Física. Escuela Nacional de Antropología e Historia. México.
- Hancock AM, Witonsky DB, Gordon AS, Eshel G, Pritchard JK, Coop G, Di Rienzo A (2008) Adaptations to Climate in Candidate Genes for Common Metabolic Disorders. *PLoS Genet*. 4(2):e32.
- Hare MP, Palumbi SR (2003) High intron sequence conservation across three mammalian orders suggests functional constraints. *Mol Biol Evol* 20(6):969-978.
- Harris EE, Meyer D (2006) The molecular signature of selection underlying human adaptations. *Am J Phys Anthropol. Suppl* 43:89-130.
- Hey J. (2005) On the number of New World founders: a population genetic portrait of the peopling of the Americas. *PLoS Biol.* 3(6):e193.
- Hetzer-Egger C, Schorpp M, Boehm T (2000) Evolutionary conservation of gene structures of the Pax1/9 gene family. *Biochim Biophys Acta*;1492(2-3):517-521.
- Hydenblit R (1996) Dental variation among four prehispanic Mexican populations. *Am J of Phys Anthropol*. 100: 225-246.

- Hughes AL (2007) Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity*. 99(4):364-373.
- Hughes I, Binkley J, Hurle B, Green ED, Comparative Sequencing Program N, Sidow A, Ornitz DM (2008) Identification of the Otopetrin Domain, a conserved domain in vertebrate otopetrins and invertebrate otopetrin-like family members. *BMC Evol Biol*. 8(1):41
- Irish, J. D. (1993). Biological Affinities of Late Pleistocene through Modern African Aboriginal Populations: The Dental Evidence. Ph.D. Dissertation, Arizona State University, Tempe.
- Irish JD (1997) Characteristic high- and low-frequency dental traits in sub-Saharan African populations. *Am J Phys Anthropol* 102(4):455-467.
- Irish JD, Guatelli-Steinberg D (2003) Ancient teeth and modern human origins: an expanded comparison of African Plio-Pleistocene and recent world dental samples. *J Hum Evol* 45(2):113-144.
- Jakobsson M, Scholz SW, Scheet P, Gibbs JR, VanLiere JM, Fung HC, Szpiech ZA, Degnan JH, Wang K, Guerreiro R, Bras JM, Schymick JC, Hernandez DG, Traynor BJ, Simon-Sanchez J, Matarin M, Britton A, van de Leemput J, Rafferty I, Bucan M, Cann HM, Hardy JA, Rosenberg NA, Singleton AB (2008) Genotype, haplotype and copy-number variation in worldwide human populations. *Nature* 451(7181):998-1003.
- Krylov M, Erdesz S, Alexeeva L, Benevolenskaya L, Arnett FC, Reveille JD (1995) HLA class II and HLA-B27 oligotyping in two Siberian native population groups. *Tissue Antigens*. 46(5):382-386.
- Krylov MIu, Erdes SH, Alekseeva LI, Benevolenskaia LI (1995) DNA typing of HLA class II genes among the aboriginal inhabitants of Chukotka. *Genetika*. 31(6):852-858.
- Lee I, Lehner B, Crombie C, Wong W, Fraser AG, Marcotte EM.(2008)A single gene network accurately predicts phenotypic effects of gene perturbation in *Caenorhabditis elegans*. *Nat Genet*. (2):181-188.

- Liu CC, Lin CC, Li KC, Chen WS, Chen JC, Yang MT, Yang PC, Chang PC, Chen JJ (2007) Genome-wide identification of specific oligonucleotides using artificial neural network and computational genomic analysis. *BMC Bioinformatics*. 8:164.
- Liu H, Prugnolle F, Manica A, Balloux F. (2006) A geographically explicit genetic model of worldwide human-settlement history. *Am J Hum Genet*.79(2):230-237.
- Majewski J, Ott J (2002) Distribution and characterization of regulatory elements in the human genome. 12(12):1827-1836.
- Marrero AR, Bravi C, Stuart S, Long JC, Pereira das Neves Leite F, Kommers T, Carvalho CM, Pena SD, Ruiz-Linares A, Salzano FM, Bortolini MC (2007) Pre- and post-Columbian gene and cultural continuity: the case of the Gaucho from southern Brazil. *Hum Hered* 64(3):160-171.
- Marrero AR, Silva-Junior WA, Bravi CM, Hutz MH, Petzl-Erler ML, Ruiz-Linares A, Salzano FM, Bortolini MC (2007) Demographic and evolutionary trajectories of the Guarani and Kaingang natives of Brazil. *Am J Phys Anthropol* 132(2):301-310.
- Mazières S, Guitard E, Crubézy E, Dugoujon JM, Bortolini MC, Bonatto SL, Hutz MH, Bois E, Tiouka F, Larrouy G, Salzano FM (2008) Uniparental (mtDNA, Y-chromosome) polymorphisms in French Guiana and two related populations--implications for the region's colonization. *Ann Hum Genet*. 72(Pt 1):145-156.
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol*. 3(5):418-426.
- Nei M (2007) The new mutation theory of phenotypic evolution. *Proc Natl Acad Sci U S A* 104(30):12235-12242.
- Nieminen P, Arte S, Tanner D, Paulin L, Alaluusua S, Thesleff I, Pirinen S (2001) Identification of a nonsense mutation in the PAX9 gene in molar oligodontia. *Eur J Hum Genet* 9(10):743-746.
- Pereira TV, Salzano FM, Mostowska A, Trzeciak WH, Ruiz-Linares A, Chies JA, Saavedra C, Nagamachi C, Hurtado AM, Hill K, Castro-

- de-Guerra D, Silva-Júnior WA, Bortolini MC (2006) Natural selection and molecular evolution in primate PAX9 gene, a major determinant of tooth development. *Proc Natl Acad Sci U S A* 103(15):5676-5681.
- Peters H, Neubüser A, Balling R (1998) Pax genes and organogenesis: Pax9 meets tooth development. *Eur J Oral Sci* 106(1):38-43.
- Piazza A, Menozzi P, Cavalli-Sforza L. (1981) The making and testing of geography gene frequency maps. *Biometrics* 37: 635–659.
- Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW, Cavalli-Sforza LL (2005) Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc Natl Acad Sci U S A* 102(44):15942-15947.
- Reich DE, Schaffner SF, Daly MJ, McVean G, Mullikin JC, Higgins JM, Richter DJ, Lander ES, Altshuler D (2002) Human genome sequence variation and the influence of gene history, mutation and recombination. *Nat Genet* 32 (1):135-142.
- Rogers J, Comuzzie AG (1995) When is ancient polymorphism a potential problem for molecular phylogenetics? *Am J Phys Anthropol* 98(2):216-218.
- Romeo S, Pennacchio LA, Fu Y, Boerwinkle E, Tybjaerg-Hansen A, Hobbs HH, Cohen JC (2007) Population-based resequencing of ANGPTL4 uncovers variations that reduce triglycerides and increase HDL. *Nat Genet.* 39(4):513-516.
- Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW (2002) Genetic structure of human populations. *Science* 298(5602):2381-2385.
- Rozkocová E, Marková M, Dolejší J (1999) Studies on agenesis of third molars amongst populations of different origin. *Sb Lek* 100(2):71-84.

- Rozkovcová E, Marková M, Láník J, Zvárová J (2004a) Development of third molar in the Czech population. *Prague Med Rep* 105(4):391-422.
- Rozkovcová E, Marková M, Láník J, Zvárová J (2004b) Agenesis of third molars in young Czech population. *Prague Med Rep* 105(1):35-52.
- Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S, Mullikin JC, Mortimore BJ, Willey DL, Hunt SE, Cole CG, Coggill PC, Rice CM, Ning Z, Rogers J, Bentley DR, Kwok PY, Mardis ER, Yeh RT, Schultz B, Cook L, Davenport R, Dante M, Fulton L, Hillier L, Waterston RH, McPherson JD, Gilman B, Schaffner S, Van Etten WJ, Reich D, Higgins J, Daly MJ, Blumenstiel B, Baldwin J, Stange-Thomann N, Zody MC, Linton L, Lander ES, Altshuler D; International SNP Map Working Group (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*. 409(6822):928-933.
- Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* 4(4):406-425.
- Salzano FM, Neel JV, Weitkamp LR, Woodall JP (1972) Serum proteins, hemoglobins and erythrocyte enzymes of Brazilian Cayapo Indians. *Hum Biol.* 44(3):443-458.
- Salzano FM, Callegari Jacques SM, Franco MH, Hutz MH, Weimer TA, Silva RS, Da Rocha FJ (1980) The Caingang revisited: blood genetics and anthropometry. *Am J Phys Anthropol.* 53(4):513-524.
- Salzano FM Black FL Callegari-Jacques SM Santos SE Weimer TA Mestriner MA Pandey JP Hutz MH Rieger TT (1988) Genetic variation within a linguistic group: Apalai-Wayana and other Carib tribes. *Am J Phys Anthropol.* 75(3):347-56.
- Salzano FM, Callegari-Jacques SM. 1988. South American Indians: a case study in evolution. Oxford, England: Oxford University Press. 260p.
- Salzano FM (1991) Interdisciplinary approaches to the human biology of South Amerindians. *Hum Biol.* 63(6):875-882.

- Salzano FM, Black FL, Callegari-Jacques SM, Santos SE, Weimer TA, Mestriner MA, Kubo RR, Pandey JP, Hutz MH (1991) Blood genetic systems in four Amazonian tribes. *Am J Phys Anthropol.* 85(1):51-60.
- Salzano FM, Franco MH, Weimer TA, Callegari-Jacques SM, Mestriner MA, Hutz MH, Flowers NM, Santos RV, Coimbra CE Jr (1997) The Brazilian Xavante Indians revisited: new protein genetic studies. *Am J Phys Anthropol.* 104(1):23-34.
- Santos SE, Ribeiro-Dos-Santos AK, Guerreiro JF, Santos EJ, Weimer TA, Callegari-Jacques SM, Mestriner MA, Franco MH, Hutz MH, Salzano FM (1998) New protein genetic studies in six Amazonian Indian populations. *Ann Hum Biol.* 1998 25(6):505-522.
- Santos-Lopes SS, Pereira RW, Wilson IJ, Pena SD (2007) A worldwide phylogeography for the human X chromosome. *PLoS ONE* 2(6):e557.
- Silvestri AR, Jr, Singh I (2003) The unresolved problem of the third molar: would people be better off without it? *J Am Dent Assoc.* 134(4):450-455.
- Sorek R, Ast G (2003) Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genomics.* 13(7):1631-1637.
- Stockton DW, Das P, Goldenberg M, D'Souza RN, Patel PI (2000) Mutation of PAX9 is associated with oligodontia. *Nat Genet* 24(1):18-19.
- Sutter R (2005) The prehistoric peopling of the South America as inferred from epigenetic dental traits. *Andean Past*, Vol. 7.
- Swinnen S, Bailleul-Forestier I, Arte S, Nieminen P, Devriendt K, Carels C (2008) Investigating the etiology of multiple tooth agenesis in three sisters with severe oligodontia. *Orthod Craniofac Res.* 11(1):24-31.
- Tallón-Walton V, Manzanares-Céspedes MC, Arte S, Carvalho-Lobato P, Valdivia-Gandur I, Garcia-Susperregui A, Ventura F, Nieminen P (2007) Identification of a novel mutation in the PAX9 gene in a family affected by oligodontia and other dental anomalies. *Eur J Oral Sci* 115(6):427-432.

- Trimmell JB (2004) M.Sc. thesis. (University of Missouri, Kansas City).
- Turner, C.G.II., 1986. The first Americans: the dental evidence. *Natl. Geogr. Res.* 2, 37-46.
- Turner CG 2nd (1987) Late Pleistocene and Holocene population history of East Asia based on dental variation. *Am J Phys Anthropol.* 73(3):305-321.
- Turner, C. G. II & Markowitz, M. (1990). Dental discontinuity between late Pleistocene and recent Nubians. I. Peopling of the Eurafrikan-South Asian Triangle. *Homo* 41, 32-41.
- Urban G. (1998) A história da cultura brasileira segundo as linguas natives. In: Carneiro da Cunha MC, editor. *História dos índios no Brasil*. São Paulo: Companhia das Letras. P 87-102.
- Voelker RB, Berglund JA (2007) A comprehensive computational characterization of conserved mammalian intronic sequences reveals conserved motifs associated with constitutive and alternative splicing. *Genome Res* 17(7):1023-1033.
- Wall JD, Przeworski M (2000) When did the human population size start increasing? *Genetics.* 155(4):1865-1874.
- Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, Bedoya G, Rojas W, Parra MV, Molina JA, Gallo C, Mazzotti G, Poletti G, Hill K, Hurtado AM, Labuda D, Klitz W, Barrantes R, Bortolini MC, Salzano FM, Petzl-Erler ML, Tsuneto LT, Llop E, Rothhammer F, Excoffier L, Feldman MW, Rosenberg NA, Ruiz-Linares A (2007) Genetic Variation and Population Structure in Native Americans. *PloS Genet* 3(11):e185.
- Waterston RH, Lander ES, Sulston JE (2002) On the sequencing of the human genome. *Proc Natl Acad Sci U S A* 99(6):3712-3716.
- Weir, B. S. and Cockerman, C. C. (1984) Estimating F-statistics for the analysis of population structure. *Evolution* 38: 1358-1370.
- Wooding S, Ostler C, Prasad BV, Watkins WS, Sung S, Bamshad M, Jorde LB (2004) Directional migration in the Hindu castes:

inferences from mitochondrial, autosomal and Y-chromosomal data. *Hum Genet* 115(3):221-229.

Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*. 155(1):431-449.

Yang J, Wang ZL, Zhao XQ, Wang de P, Qi de L, Xu BH, Ren YH, Tian HF (2008) Natural selection and adaptive evolution of leptin in the ochotona family driven by the cold environmental stress. *PLoS ONE*. 3(1):e1472.

TABLE 1.- Sample sizes, geographic location and linguistic information regarding fifteen South Amerindian and one Eskimo populations analyzed in this study

Population	No. of individuals	Geographic location	Country/Region	Language classification ^a	References for additional information
South Amerindians					
Aché	20	23°30' - 24°10' S 55°50' - 56°30' W	Paraguay	Tupian	Brown <i>et al.</i> (1974); Hill and Hurtado (1996); Bortolini <i>et al.</i> , 2003
Apalaí	9	1° 20' N - 54° 40' W	Brazil/North	Cariban	Salzano <i>et al.</i> (1988); Mazières <i>et al.</i> (2008)
Arara	21	3° 30' S - 54° 10' W	Brazil/North	Arara	Salzano <i>et al.</i> (1991)
Bari	4	12° 50' S - 73° 5' W	Venezuela	Chibchan	Bortolini <i>et al.</i> (2003)
Cayapo	1	10° 20' S - 53° 5' W	Brazil/North	Jêan	Salzano <i>et al.</i> (1972)
Guarani	5	23° 6' S - 55° 12' W	Brazil/South	Tupian	Salzano <i>et al.</i> (1997); Bortolini <i>et al.</i> (2003); Marrero <i>et al.</i> (2007)
Galibi	7	3° 45' N - 51° 15' W	Brazil/North	Cariban	Santos <i>et al.</i> (1998)
Jamamadi	5	7° 15' S - 66° 41' W	Brazil/North	Arahuan	Salzano <i>et al.</i> (1991)
Kaingang	20	27° 20' S; 52° 45' W	Brazil/South	Jêan	Salzano <i>et al.</i> (1980); Belich <i>et al.</i> (1992); Bortolini <i>et al.</i> , (2003); Marrero <i>et al.</i> (2007)
Mekranoti	5	8° 40' S - 54° W	Brazil/North	Jêan	Salzano <i>et al.</i> (1972)
Mundurucu	8	6° 23' S - 59° 9' W	Brazil/North	Tupian	Salzano <i>et al.</i> (1991)
Tenharim	8	8° 20' S - 54° W	Brazil/North	Tupian	Santos <i>et al.</i> (1998)
Warao	4	11° 50' S - 74° 5' W	Venezuela	Warao	Bortolini <i>et al.</i> (2003)
Xikrin	17	5° 55' S 51° W	Brazil/North	Jêan	Salzano <i>et al.</i> (1972)
Yucpa	4	12° 50' S - 71° 5' W	Venezuela	Chibchan	Bortolini <i>et al.</i> (2003)
Northern Arctic					
Eskimo	44	64° N 175° W	Russia	Eskimo-Aleut	Erdesz <i>et al.</i> (1994); Krylov <i>et al.</i> (1995)

^aAccording to Campbell (1997)

TABLE 3 .Genotype and allele distributions of the Ala240Pro polymorphism in 15 South Amerindian and other human populations

Population	Number of ind.	Genotype frequency			Allele frequency		Reference
		Ala/ Ala C/C (%)	Ala/Pro C/G (%)	Pro/Pro G/G (%)	C (Ala)	G (Pro)	
<i>South Amerindian</i>							
Aché	20	20 (100)	0	0	1	0	Pereira et al. (2006)
Apalaí	9	7 (78)	2 (22)	0	0.89	0.11	Present study
Arara	21	18 (86)	3 (14)	0	0.86	0.14	Present study
Bari	4	4 (100)	0	0	1	0	Pereira et al. (2006)
Cayapo	1	1 (100)	0	0	1	0	Present study
Galibi	7	6 (86)	1 (14)	0	0.86	0.14	Present study
Guarani	5	5 (100)	0	0	1	0	Pereira et al. (2006)
Jamamadi	5	2 (40)	2(40)	1(20)	0.6	0.40	Present study
Kaigang	20	15 (75)	5 (25)	0	0.88	0.12	Pereira et al. (2006)
Mekranoti	5	5(100)	0	0	1	0	Present study
Munducuru	8	7 (88)	1(12)	0	0.94	0.06	Present study
Tenharim	8	6 (75)	2 (25)	0	0.88	0.12	Present study
Warao	4	4 (100)	0	0	1	0	Pereira et al. (2006)
Xikrin	17	15 (88)	2 (12)	0	0.94	0.06	Present study
Yucpa	4	4 (100)	0	0	1	0	Pereira et al. (2006)
Total	138	119 (86)	18 (13)	1 (<1)	0.93	0.07	
<i>Eskimos</i>	44	21(48)	20(45)	3 (7)	0.70	0.30	Present study
<i>Asians</i>	14	11(79)	2 (14)	1 (7)	0.86	0.14	Pereira et al. (2006)
<i>Europeans</i>	365	169 (46)	172 (47)	24 (7)	0.70	0.30	Pereira et al. (2006)
<i>Afro-Americans</i>	15	15 (100)	0	0	1	0	HAPMAP/dbSNP
<i>Euro-Americans</i>	14	10 (71)	0	4 (29)	0.71	0.29	HAPMAP/dbSNP

TABLE 4. Additional informations about mutations and nucleotide sequence diversity estimates of *PAX9* exon 3 and non-exon 3 bordering regions in several human populations

Population	REGION	Number of sequences	Number of sites	Number of mutations	Non-synonymous mutations	$\pi \pm SE$
Aché	Exon 3	40	138	0	0	0
	Non-coding		452	0		0
	TOTAL		590	0		0
Apalaí	Exon 3	18	138	3	3	0.00313 \pm 0.00330
	Non-coding		452	1		0.00025 \pm 0.00047
	TOTAL		590	4		0.00092 \pm 0.00089
Arara	Exon 3	42	138	2	2	0.00162 \pm 0.00217
	Non-coding		452	0		0
	TOTAL		590	2		0.00038 \pm 0.00051
Bari	Exon 3	8	138	1	1	0.00311 \pm 0.00354
	Non-coding		452	0		0
	TOTAL		590	1		0.00073 \pm 0.00083
Cayapo	Exon 3	2	138	0	0	0
	Non-coding		452	0		0
	TOTAL		590	0		0
Galibi	Exon 3	14	138	1	1	0.00191 \pm 0.00250
	Non-coding		452	1		0.00058 \pm 0.00076
	TOTAL		590	2		0.00089 \pm 0.00089
Guarani	Exon 3	10	138	1	1	0.00258 \pm 0.00307
	Non-coding		452	0		0
	TOTAL		590	1		0.00060 \pm 0.00072
Jamamadi	Exon 3	10	138	1	1	0.00386 \pm 0.00398
	Non-coding		452	0		0
	TOTAL		590	1		0.00090 \pm 0.00092
Kaingang	Exon 3	40	138	2	1	0.00199 \pm 0.00244
	Non-coding		452	0		0
	TOTAL		590	2		0.00046 \pm 0.00057
Mekranoti	Exon 3	10	138	0	0	0
	Non-coding		452	2		0.00088 \pm 0.00101
	TOTAL		590	2		0.00068 \pm 0.00077
Munducuru	Exon 3	16	138	2	2	0.00175 \pm 0.00236
	Non-coding		452	0		0
	TOTAL		590	2		0.00041 \pm 0.00055
Tenharim	Exon 3	16	138	2	2	0.00326 \pm 0.00341
	Non-coding		452	0		0
	TOTAL		590	2		0.00076 \pm 0.00080
Warao	Exon 3	8	138	0	0	0
	Non-coding		452	0		0
	TOTAL		590	0		0
Xikrin	Exon 3	34	138	2	2	0.00084 \pm 0.00152
	Non-coding		452	0		0
	TOTAL		590	2		0.00020 \pm 0.00036
Yupca	Exon 3	8	138	0	0	0
	Non-coding		452	0		0
	TOTAL		590	0		0

TABLE 4. Cont.

Population	REGION	Number of sequences	Number of sites	Number of mutations	Non-synonymous mutations	$\pi \pm SE$
South Amerindians (Average)	Exon 3	276	138	9	8	0.00164 \pm 0.00215
	Non-coding		452	4		0.00008 \pm 0.00025
	TOTAL		590	13		0.00045 \pm 0.00055
Eskimos	Exon 3	88	138	2	1	0.00372 \pm 0.00351
	Non-coding		220	0		0
	TOTAL		590	2		0.00087 \pm 0.00082
Asians (Japanese)	Exon 3	28	138	1	1	0.00184 \pm 0.00236
	Non-coding		452	0		0
	TOTAL		590	1		0.00043 \pm 0.00055
Europeans (Polish/Spaniard)	Exon 3	30	138	1	1	0.00561 \pm 0.00469
	Non-coding		452	1		0.00014 \pm 0.00034
	TOTAL		590	2		0.00143 \pm 0.00116
Euro-Americans	Exon 3	28	138	2	1	0.00560 \pm 0.00469
	Non-coding		452	0		0
	TOTAL		590	2		0.00131 \pm 0.00110
Afro-Americans	Exon 3	30	138	0		0
	Non-coding		452	0		0
	TOTAL		590	0		0
All	Exon 3	480	138	15	12	0.00264 \pm 0.00282
	Non-coding		452	5		0.00005 \pm 0.00020
	TOTAL		590	20		0.00066 \pm 0.00069

TABLE 5. Statistical tests to detect departures from a standard neutral model

Population	Tajima's D (Non-Coding)	Tajima's D (Exon 3)	Fu and Li's Test (Non-Coding)		Fu and Li's Test (Exon 3)	
			D*	F*	D*	F*
South Amerindians	-2.13561	-1.84693	-3.84723	-2.13561	-3.03182	-3.08405
Eskimos	0	0.46730	0	ND	0.69567	0.73072
Japanese	0	-0.01865	0	ND	0.60275	0.49703
Europeans	-1.14700	1.09500	-1.68214	-1.76554	0.80615	1.02510
Euro-Americans	0	1.05245	0	ND	0.81535	1.01741
Afro-Americans	0	0	0	ND	0	ND
Total (less South Native Americans)	-0.95003	0.7049	-2.20545	-2.13561	0.63545	0.77585

Note: Statistical significance (p values) of these numbers under different demographic scenarios are presented in Table 6.
 ND. Not determined.

TABLE 6. p values of Tajima's D , and Fu and Li's D^* and F^* statistics calculated under several demographic scenario;

Neutrality tests	Populations													
	All less South Amerindians	Afro-American	European	Euro-American	Japanese	Eskimo	South Amerindian							
	1-fold	100-fold	1-fold	100-fold	1-fold	100-fold	1-fold	100-fold	1-fold	100-fold				
Non-coding Tajima's D	0.809	0.215	0.440	0.229	0.791	0.588	0.459	0.286	0.436	0.245	0.443	0.121	0.999	0.606
D^*	0.927	0.145	0.625	0.225	0.844	0.668	0.647	0.307	0.629	0.293	0.670	0.081	0.996	0.673
F^*	0.930	0.141	n/c	n/c	0.826	0.595	n/c	n/c	n/c	n/c	n/c	n/c	0.999	0.748
Exon 3 Tajima's D	0.237	0.039	0.452	0.289	0.225	0.112	0.189	0.135	0.437	0.283	0.314	0.123	0.999	0.867
D^*	0.042	0.005	0.734	0.403	0.013	0.006	0.011	0.008	0.120	0.062	0.043	0.012	0.991	0.577
F^*	0.218	0.031	n/c	n/c	0.203	0.011	0.039	0.020	0.437	0.267	0.263	0.091	0.991	0.574
First scenario - 6 000 years ago No=10 000														
Non-coding Tajima's D	0.824	0.098	0.424	0.047	0.733	0.190	0.412	0.300	0.424	0.080	0.418	0.023	0.999	0.599
D^*	0.840	0.415	0.732	0.170	0.746	0.238	0.594	0.339	0.737	0.156	0.790	0.114	0.993	0.657
F^*	0.840	0.388	n/c	n/c	0.746	0.226	n/c	n/c	n/c	n/c	n/c	n/c	0.999	0.734
Exon 3 Tajima's D	0.999	0.727	0.455	0.061	0.234	0.035	0.201	0.041	0.436	0.078	0.322	0.021	0.999	0.969
D^*	0.862	0.568	0.763	0.185	< 0.001	< 0.001	< 0.001	< 0.001	0.016	0.004	< 0.001	< 0.001	0.999	0.851
F^*	0.862	0.547	n/c	n/c	0.003	0.001	0.006	< 0.001	0.440	0.085	0.231	0.017	0.999	0.851
Second scenario - 6 000 years ago No=1000														
Non-coding Tajima's D					0.765	0.489	0.450	0.230	0.444	0.242	0.438	0.124	0.999	0.562
D^*					0.798	0.553	0.707	0.332	0.662	0.323	0.740	0.128	0.989	0.328
F^*					0.794	0.534	n/c	n/c	n/c	n/c	n/c	n/c	0.999	0.451
Exon 3 Tajima's D					0.182	0.072	0.185	0.071	0.445	0.232	0.299	0.049	0.999	0.459
D^*					0.055	0.025	0.055	0.016	0.208	0.079	0.092	0.011	0.986	0.200
F^*					0.069	0.031	0.071	0.026	0.451	0.166	0.248	0.024	0.986	0.234
Third scenario - 40,000 years ago No=10 000														

TABLE 6.. Cont.

Neutrality tests	Population										
	All less South Amerindians	Afro-American	European	Euro-American	Japanese	Eskimo	South Amerindian				
1-fold	100-fold	1-fold	100-fold	1-fold	100-fold	1-fold	100-fold	1-fold	100-fold	1-fold	100-fold
Non-coding Tajima's D					0.411				0.193		
D*					0.692				0.284		
F*					n/c				n/c		
Exon3 Tajima's D					0.439				0.223		
D*					0.036				0.024		
F*					0.446				0.209		
Fourth scenario - 37,000 years ago No=3,000											
Non-coding Tajima's D						0.394			0.048		0.998
D*						0.792			0.184		0.919
F*						n/c			n/c		0.989
Exon3 Tajima's D						0.340			0.049		0.999
D*						< 0.001			< 0.001		0.909
F*						0.308			0.036		0.907
Fifth scenario - 18,000 years ago No=500											
Non-coding Tajima's D						0.414			0.013		0.926
D*						0.787			0.386		0.681
F*						n/c			n/c		0.865
Exon3 Tajima's D						0.343			0.016		0.999
D*						< 0.001			0.001		0.982
F*						0.295			0.015		0.982
Sixth scenario - 18,000 years ago No=80											

Note: Statistically significant values are shown in bold faces; n/c: statistical not computed since F* statistical is dependent on the number of segregating sites. No.: size of the founding population. The intermediate values between 1 and 100-fold expansions are not shown.

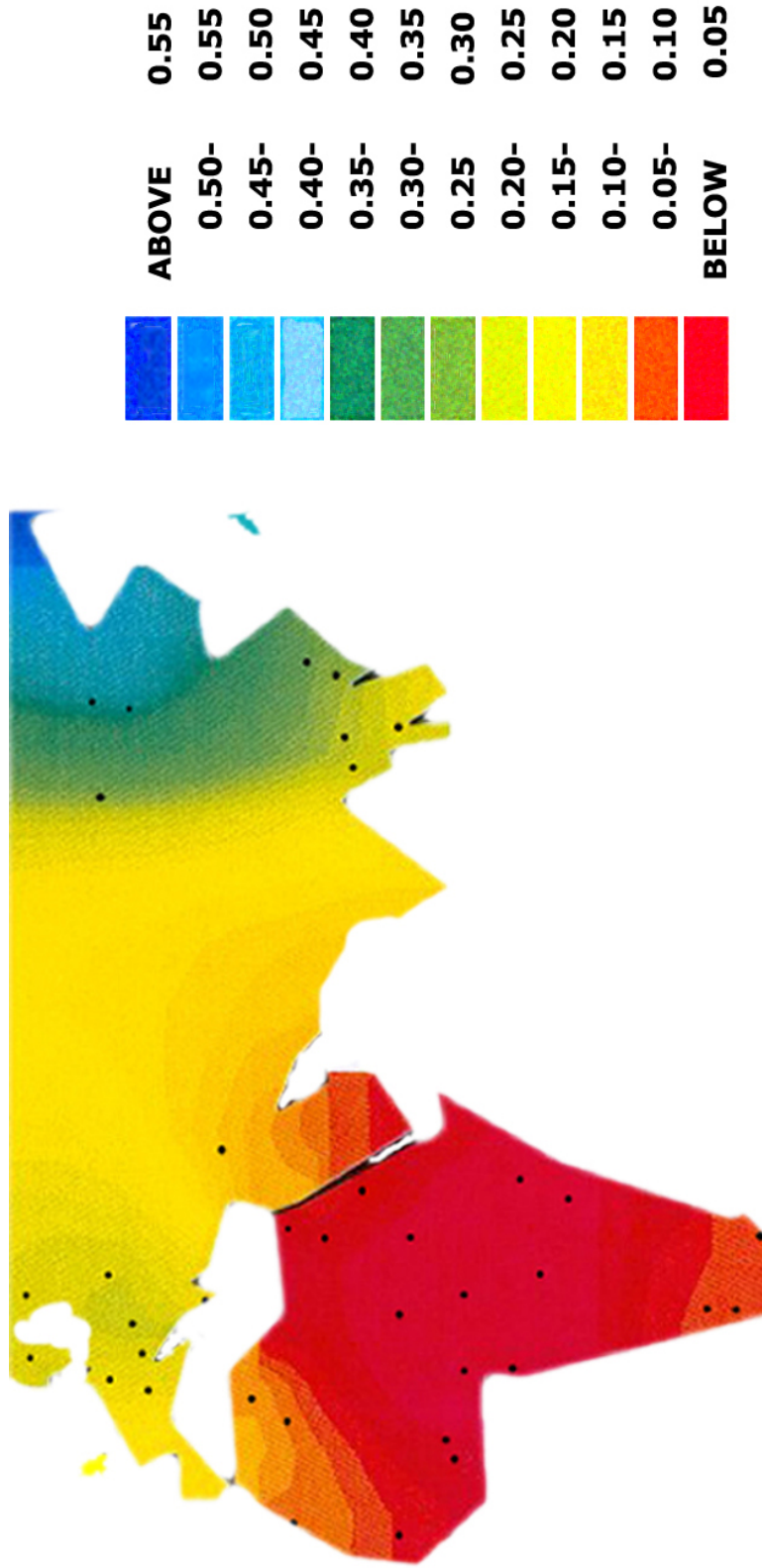


FIGURE 1. Synthetic map illustrating the distribution of ageneses of the third molar in the old world; Variation in the frequency of ageneses compiled: Turner ,1987; Irish,, 1993; Rozcová, 2004a e 2004b; Sutter, 2005.

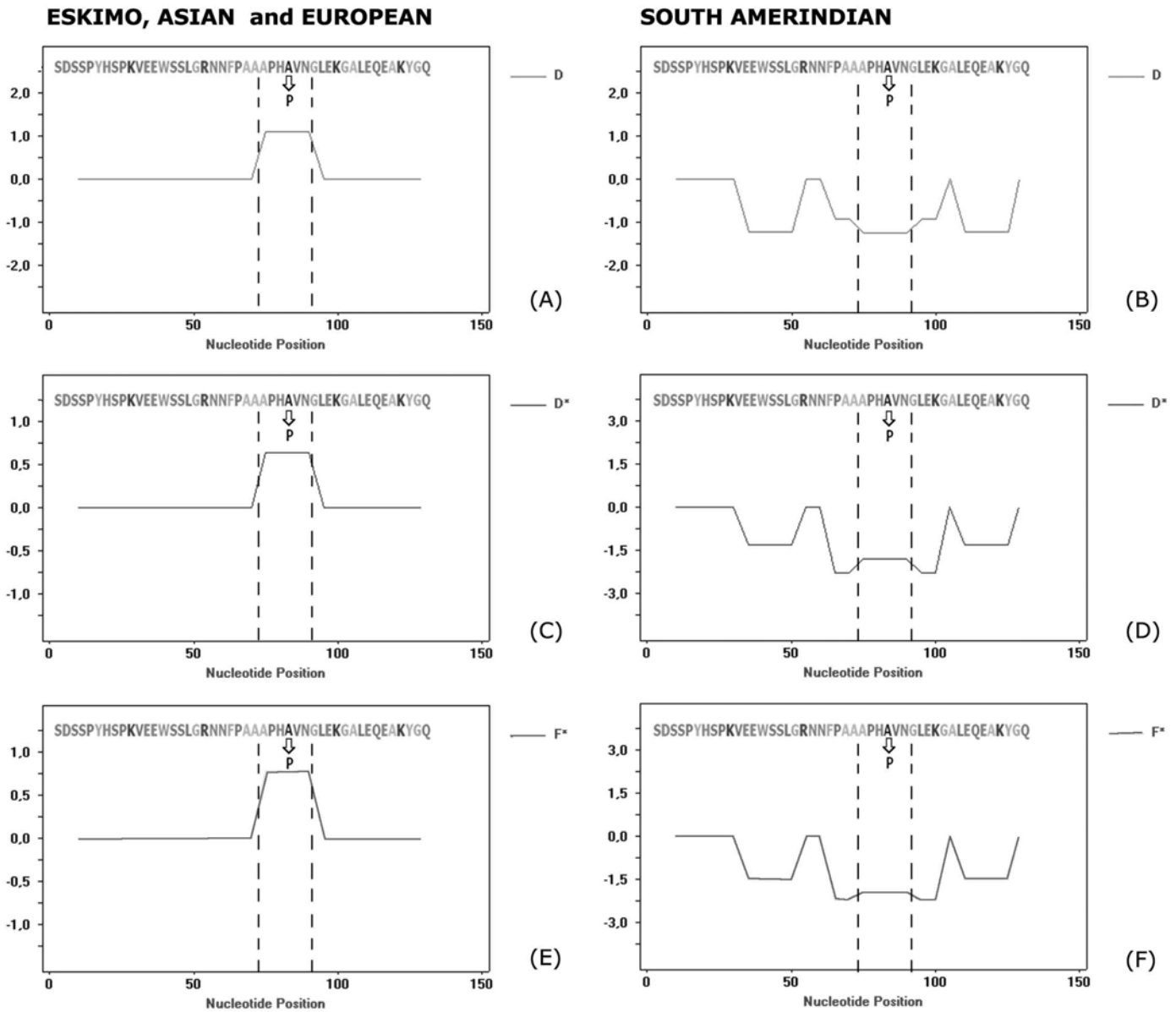


FIGURE 2. Graphical of the Sliding Window method, perform for DNAsp 4.0, (window length = 20 ntd; step size = 5 ntd). In the graph Tajima's D and Fu and Li's D* and F* values (Y axis) can be plotted against the nucleotide position (X axis). Note that the graphics are not with the same scale. The sequence of the protein was placed to illustrate the target region. Tajima's D neutrality test (A) Asian, Eskimo and European (B) South Amerindian; Fu and Li's D* test (C) Asian, Eskimo and European (D) South Amerindian; Fu and Li's F* test (E) Asian, Eskimo and European (F) South Amerindian.

III-2. Resultados adicionais

Além das amostras humanas (ver manuscrito) foram também investigadas 53 outras espécies de mamíferos (Tabelas 3 e 4). Destas, 25 tiveram os exons 2, 3 e 4 do gene *PAX9* seqüenciados para o presente estudo (Tabela 3), enquanto dados das demais foram compilados de *sites* especializados em bancos com seqüências de DNA (Tabela 4). Com isso pretende-se atingir os objetivos 2 e 3 dessa dissertação.

No apêndice I pode ser visto os detalhes sobre as amostras, bem como sobre os métodos utilizados para a obtenção dos resultados que serão apresentados a seguir. Vale destacar que algumas análises envolveram somente primatas enquanto outras envolveram o banco de dados completo. Salienta-se ainda que diante do volume de dados, as análises que seguem são de caráter preliminar, sendo que deverão ser expandidas num futuro próximo.

TABELA 3. Relação das amostras dos mamíferos.

Espécie	ID	N	Ordem	Fórmula Dental: (I - C- PM -M)	Alimentação	E2	E3	E4
<i>Alouatta</i>	Bugiu	2	Primata (NWM)	(2/2 - 1/1 - 3/3 - 3/3) X 2 = 36	Herbívoros	Ok	Ok	Ok
<i>Brachyteles</i>	Muriqui	1	Primata (NWM)	(2/2 - 1/1 - 3/3 - 3/3) X 2 = 36	Herbívoros	Ok	Ok	Ok
<i>Bradypus</i>	Sloth	1	Xenarthra	5/4-5 = 18-20	Herbívoros	Ok	Ok	Ok
<i>Cacajao melanocephalus</i>	Uakari	2	Primata (NWM)	(2/2 - 1/1 - 3/3 - 3/3) X 2 = 36	Herbívoros	NO	Ok	NO
<i>Callicebus coimbrai</i>	Titi	1	Primata (NWM)	(2/2 - 1/1 - 3/3 - 3/3) X 2 = 36	Herbívoros	NO	Ok	NO
<i>Canis domesticus</i>	Dog	7	Carnívora	(3/3 - 1/1- 4/4 - 2/2) + (3/3 - 1/1- 4/4 - 2/2) = 38	Onívoros	Ok	Ok	Ok
<i>Cebus robustus</i>	CpBlack	4	Primata (NWM)	(2/2 - 1/1 - 3/3 - 3/3) X 2 = 36	Onívoros	NO	Ok	NO
<i>Cebus xanthosternos</i>	Capuchin	6	Primata (NWM)	(2/2 - 1/1 - 3/3 - 3/3) X 2 = 36	Onívoros	NO	Ok	NO
<i>Chiropotes uthaickae</i>	Saki	1	Primata (NWM)	(2/2 - 1/1 - 3/3 - 3/3) X 2 = 36	Herbívoros	NO	Ok	NO
<i>Ctenomys Lami</i>	TucoL	3	Rodentia	(1/1 - 0/0 - 1/1 - 3/3) = 20	Herbívoros	Ok	NO	Ok
<i>Ctenomys torquatus</i>	TucoT	2	Rodentia	(1/1 - 0/0 - 1/1 - 3/3) = 20	Herbívoros	Ok	Ok	Ok
<i>Cyclopes didactylus</i>	AnteaterS	1	Xenarthra	0	Insetívoros	Ok	NO	Ok
<i>Dasybus</i>	Tatu	1	Xenarthra	7-8/7-8 até 18/19	Insetívoros	Ok	Ok	Ok
<i>Dugong dugon</i>	Dugong	1	Sirênia	2/3, 0/1, 3/3, 3/3 = 36	Herbívoros	Ok	Ok	NO
<i>Eubalaena australis</i>	Whale	4	Cetacea (Mist)	0	Planctívoras	Ok	Ok	Ok
<i>Gorilla gorilla</i>	Gorilla	1	Primata (HOM)	(2/2 - 1/1- 2/2 - 3/3) X 2= 32	Herbívoros	Ok	Ok	Ok
<i>Grampus griseus</i>	RDolphin	1	Cetacea (Odo)	0-7 homodontes	Carnívoros	Ok	NO	Ok
<i>Leontopithecus chrysomelas</i>	LionT	3	Primata (NWM)	(2/2 - 1/1 - 3/3 - 3/3) X 2 = 36	Onívoros	NO	Ok	NO
<i>Leontopithecus rosalia</i>	Tamarin	5	Primata (NWM)	(2/2 - 1/1 - 3/3 - 3/3) X 2 = 36	Onívoros	Ok	Ok	Ok
<i>Manis Javanica</i>	Pangolin	1	Pholidota	0	Insetívoros	NO	Ok	Ok
<i>Mesoplodon grayi</i>	GrayiWhale	1	Cetacea (Odo)	15 - 22/1 homodontes	Piscívoras	Ok	NO	Ok
<i>Myrmecophaga tridactyla</i>	Anteater	1	Xenarthra	0	Insetívoros	Ok	Ok	Ok
<i>Pontoporia blainvillei</i>	TDolphin	2	Cetacea (Odo)	53-58 / 51-56 = 114 homodontes	Piscívoras	Ok	Ok	Ok
<i>Tamandua tetradactyla</i>	AnteaterM	1	Xenarthra	0	Insetívoros	Ok	Ok	Ok
<i>Trichechus inunguis</i>	Manatee	1	Sirênia	6 (M - substituíveis)	Herbívoros	Ok	Ok	NO
25 Espécies	Total	54				18	23	17

N - Número de indivíduos; ID - identificação; I- incisivos; C- caninos; PM- pré-molares; M - molares; E2 - exon2; E3 - exon3; E4 - exon4; HOM - *Hominidae*; NWM - *new world monkey*; MYS - *Mysticeti*; ODO - *Odontoceti*; Dados Coletados: *NCBI taxonomy database* (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=taxonomy>), *Animal Diversity Web - University of Michigan Museum of Zoology* (<http://animaldiversity.ummz.umich.edu/site/index.html>), *Mammalian Crania Photographic Archive* (<http://taxa.soken.ac.jp/mcpa2/en/mammal.html>).

TABELA 4. Relação das amostras dos mamíferos *in silico*.

Espécie	ID	SEQ	Ordem	Fórmula Dental: (I - C- PM -M)	Alimentação	E2	E3	E4
<i>Aotus infulatus</i>	Aotus	NCBI	Primata (NWM)	(2/2 - 1/1 - 3/3 - 3/3) X 2 = 36	Onívoros	NO	Ok	NO
<i>Bos taurus</i>	Cow	NCBI	Artiodactyla	(0/3, 0/1, 2-3/3, 3/3) = 30-32	Herbívoros	Ok	Ok	Ok
<i>Callimico goeldii</i>	Goeldis	NCBI	Primata (NWM)	(2/2 - 1/1 - 3/3 - 3/3) X 2 = 36	Onívoros	Ok	Ok	Ok
<i>Callithrix jacchus</i>	Marmoset	NCBI	Primata (NWM)	(2/2 - 1/1 - 3/3 - 3/3) X 2 = 36	Herbívoros	Ok	Ok	Ok
<i>Dasyopus novemcinctus</i>	Armadillo	Ensemble	Xenarthra	7-8/7-8 to 18/19	Insetívoros	Ok	Ok	Ok
<i>Daubentonia madagascariensis</i>	Ayeaye	NCBI	Primata (LEM)	(1/1, 0/0, 1/0, 3/3) = 18	Onívoros	Ok	Ok	Ok
<i>Echinops telfairi</i>	Tenrec	Ensemble	Afrosoricida	2-3/2-3, 1/1, 2-3/2-3, 2-4/2-3 = 32-40	Onívoros	Ok	Ok	Ok
<i>Gallus gallus</i>	Chick	Ensemble	Aves	0	Onívoros	Ok	Ok	NO
<i>Lemur catta</i>	Lemur	NCBI	Primata (LEM)	2/2, 1/1, 3/3, 3/3 = 36	Herbívoros	Ok	Ok	Ok
<i>Lepilemur edwardsi</i>	Lepilemur	NCBI	Primata (LEM)	2/2, 1/1, 3/3, 3/3 = 36	Herbívoros	Ok	Ok	Ok
<i>Macaca mulatta</i>	Rhesus	NCBI	Primata (OWM)	(2/2 - 1/1 - 2/2 - 3/3) X 2 = 32	Onívoros	Ok	Ok	Ok
<i>Microcebus murinus</i>	Mlemur	Ensemble	Primate (LEM)	2/2, 1/1, 3/3, 3/3 = 36	Insetívoros	OK	OK	Ok
<i>Monodelphis domestica</i>	Opossum	NCBI	Didelphimorphia	5/4, 1/1, 3/3, 4/4 = 50	Onívoros	Ok	Ok	Ok
<i>Mus musculus</i>	Mouse	NCBI	Rodentia	(1/1 - 0/0 - 1/1 - 3/3) = 20	Onívoros	Ok	Ok	Ok
<i>Myotis lucifugus</i>	Bat	Ensemble	Chiroptera	1-2/2-3, 1/1, 1-3/2-3, 3/3 = 28 -38	Insetívoros	Ok	Ok	OK
<i>Pan troglodytes</i>	Chimp	NCBI	Primata (HOM)	(2/2 - 1/1 - 2/2 - 3/3) X 2 = 32	Onívoros	Ok	Ok	Ok
<i>Rattus norvegicus</i>	Rat	NCBI	Rodentia	(1/1 - 0/0 - 1/1 - 3/3) = 20	Onívoros	Ok	Ok	Ok
<i>Ornithorhynchus anatinus</i>	Platypus	NCBI	Monotrêmata	0	Piscívoras	Ok	Ok	Ok
<i>Otolemur garnettii</i>	Galago	Ensemble	Primate (LEM)	2/2, 1/1, 3/3, 3/3 = 36	Onívoros	OK	NO	Ok
<i>Perodicticus potto</i>	Potto	NCBI	Primata (LEM)	2/2, 1/1, 3/3, 3/3 = 36	Onívoros	Ok	Ok	Ok
<i>Propithecus verreauxi coquereli</i>	Sifaka	NCBI	Primata (LEM)	2/2 - 1/1 - 3/3 - 3/3) + (1/1 - 1/1 - 3/3 - 3/3) = 34	Herbívoros	Ok	Ok	Ok
<i>Saguinus oedipus</i>	Sagui	NCBI	Primata (NWM)	(2/2 - 1/1 - 3/3 - 3/3) X 2 = 36	Onívoros	Ok	Ok	Ok
<i>Saimiri boliviensis</i>	Saimiri	NCBI	Primata (NWM)	(2/2 - 1/1 - 3/3 - 3/3) X 2 = 36	Onívoros	Ok	Ok	Ok
<i>Saimiri sciureus</i>	Ssciureus	NCBI	Primata (NWM)	(2/2 - 1/1 - 3/3 - 3/3) X 2 = 36	Onívoros	Ok	Ok	Ok
<i>Sorex araneus</i>	Shrew	Ensemble	Soricomorpha	3/1-2, 1/0-1, 1-3/1, 3/3 = 26-32	Carnívoros	Ok	Ok	Ok
<i>Sus scrofa</i>	Pig	Ensemble	Artiodactyla	1-3/3, 1/1, 2-4/2 or 4, 3/3 = 34-44	Onívoros	Ok	Ok	Ok
<i>Tarsius syrichta</i>	Tarsier	NCBI	Primata (LEM)	2/2 - 1/1 - 3/3 - 3/3) + (1/1 - 1/1 - 3/3 - 3/3) = 34	Insetívoros	Ok	Ok	Ok
<i>Tupaia belangeri</i>	Tshrews	Ensemble	Scandentia	2/3, 1/1, 3/3, 3/3 = 38	Insetívoros	Ok	Ok	Ok
28 Espécies						27	27	26

N - Número de indivíduos; ID - identificação; I- incisivos; C- caninos; PM- pré-molares; M - molares; E2 -exon2; E3 -exon3; E4 -exon4;

HOM - *Hominidae*; NWM - *new world monkey*; OWM - *old world monkey*; MYS - *Mysticeti*; ODO - *Odontoceti*; LEM - *lemuriforme*; SEQ - origem da sequência, NCBI (<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>), Ensemble (<http://www.ensembl.org/Multi/blastview>)

Dados Coletados: NCBI taxonomy database (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=taxonomy>),

Animal Diversity Web - University of Michigan Museum of Zoology (<http://animaldiversity.ummz.umich.edu/site/index.html>),

Mammalian Crania Photographic Archive (<http://taxa.soken.ac.jp/mcpa2/en/mammal.htm> l).

III-2.1. A variação do exon 3 do gene PAX9 dentro do gênero *Cebus* e entre espécies de macacos do Novo Mundo

Como visto no manuscrito, a variabilidade no exon 3 na nossa espécie pode ser explicada por um conjunto de fenômenos, dentre os quais seleção Darwiniana. Qual seria a variabilidade do exon 3 considerando outras espécies de primatas?

Obviamente, o ideal para responder esta pergunta seria gerar dados, como foi feito no caso dos humanos, a partir de investigação de diversos indivíduos de várias espécies, incluindo grandes macacos, macacos do velho e do novo mundo, além dos primatas de Madagascar. Até o momento somente duas espécies das 24 de primatas não-humanos investigadas, tiveram mais de três indivíduos seqüenciados (10 indivíduos do gênero *Cebus*; 6 *Cebus xanthosternos*, 4 *Cebus robustus*; Tabela 3).

Considerando somente estes 10 indivíduos, não foi encontrada nenhuma substituição com troca de aminoácido dentro do gênero *Cebus* (Figuras 4 e 5; Tabela 5), e nenhum dos testes de neutralidade feitos, apresentou valores significativos (Tabela 6). Aparentemente, pelo menos, nessa linhagem dos *Platyrrhini* não há o relaxamento funcional no exon 3.

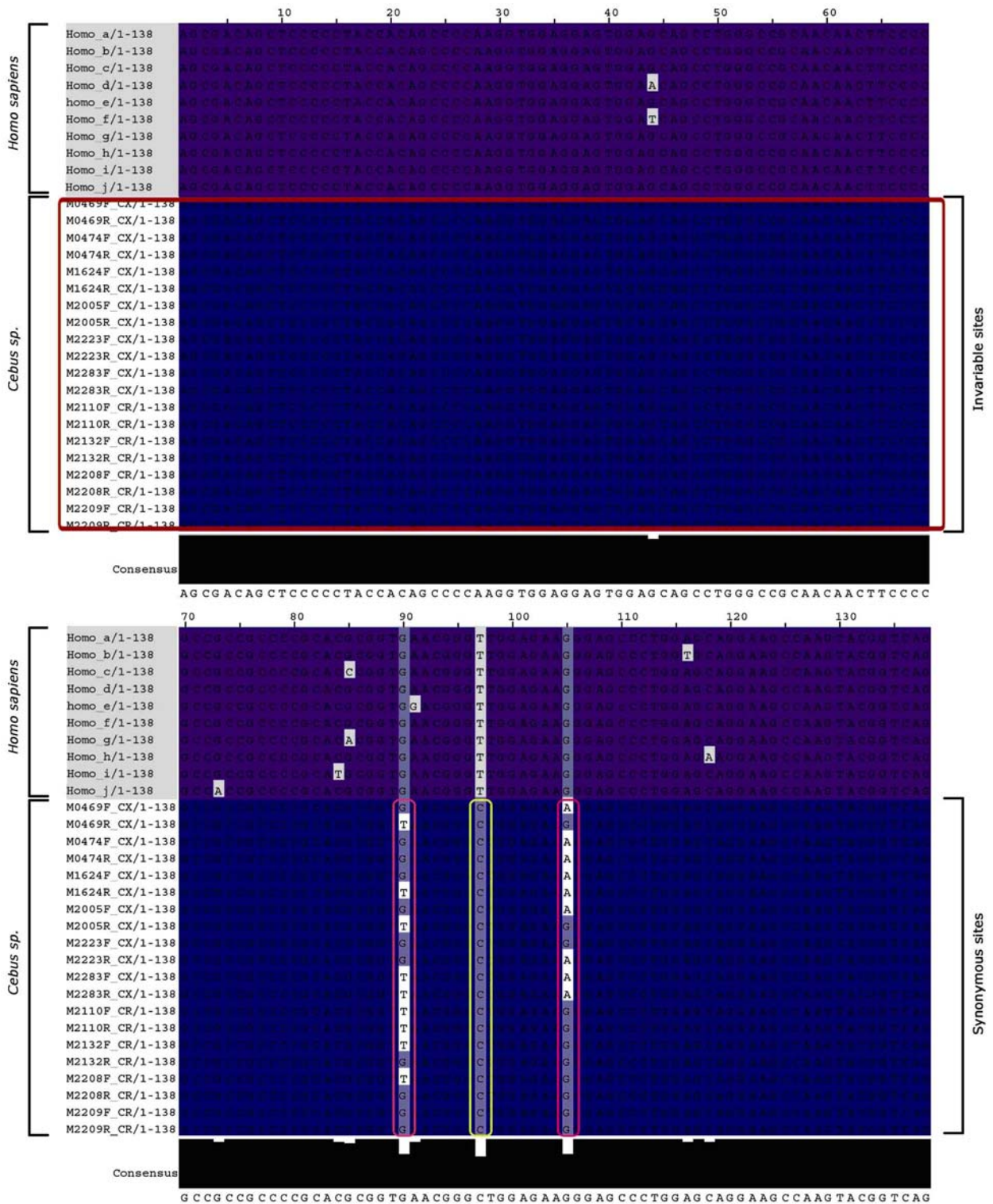


FIGURA 4. Seqüências de nucleotídeos dos haplótipos humanos e dos indivíduos do gênero *Cebus*. Gerada pelo *Jalview*.

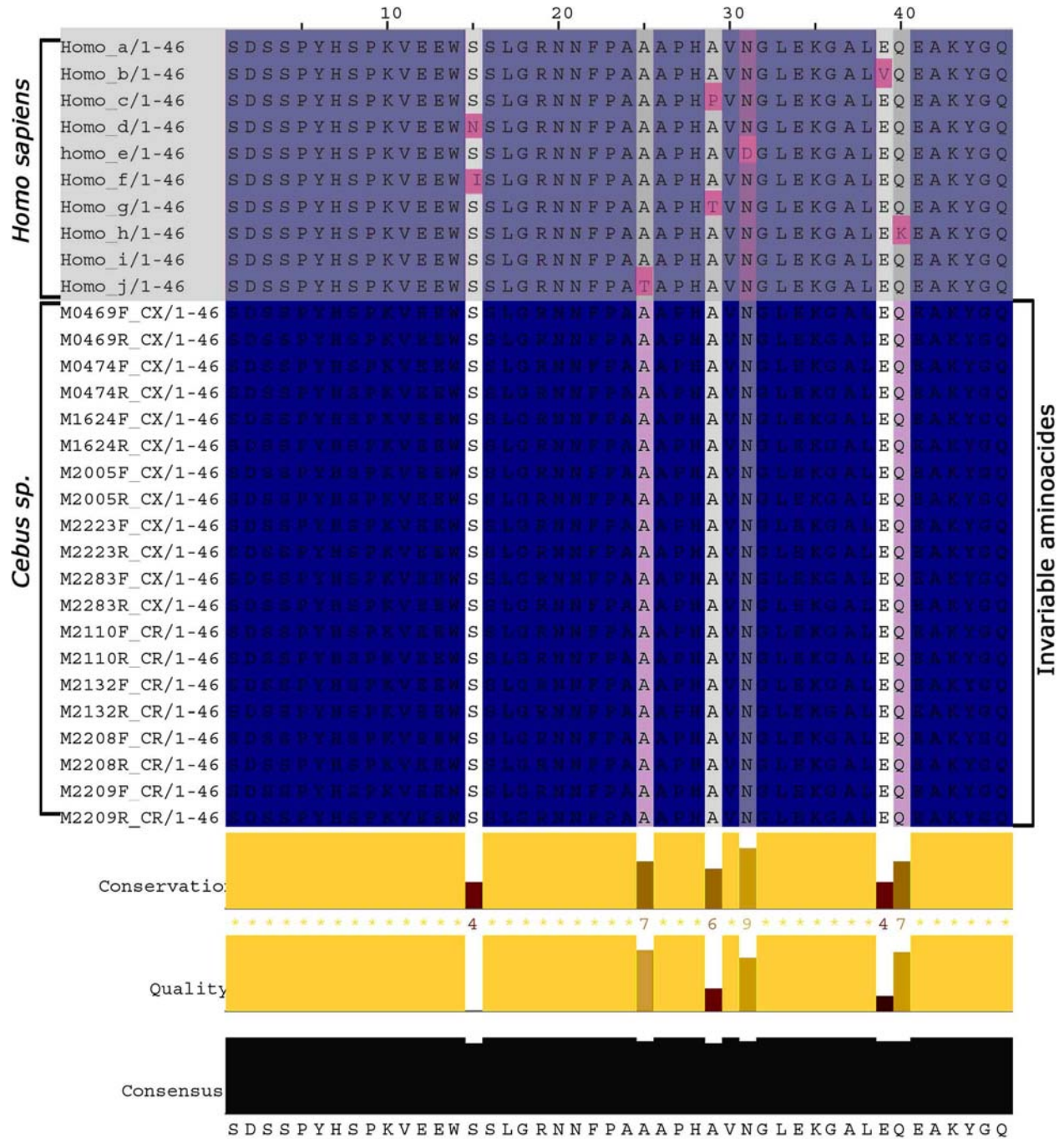


FIGURA 5. Seqüências de aminoácidos dos haplótipos humanos e das amostras do gênero *Cebus*. Gerada pelo *Jalview*.

TABELA 5. Polimorfismo no exon 3 do gene PAX9 de *Cebus*.

Região	Sites	SilSites	NSynSites	Mut(Tot)	SilMut	NSynMut
Região Codificadora	138	30,70	107,30	2	2	0
	π (Tot)	π (Sil)	Θ (Tot)	Θ (Sil)	K(JC-Tot)	K(JC-Sil)
	0,00755	0,03395	0,00409	0,01837	0,01539	0,06593

As estimativas do número de mutações, de θ (theta), e de π (Pi) foram obtidas a partir da comparação de dados intraespecíficos; Tot -total; Sil - silencioso; JC - correção de Jukes-Cantor.

TABELA 6. Testes de neutralidade considerando o exon 3 em amostras do gênero *Cebus*.

População	Tajima's D	Fu and Li's	
		D*	F*
<i>Cebus xanthosternos</i>	1,22270#	0,97295#	1,16768#
<i>Cebus robustus</i>	1,44416 #	0,88779#	1,10037#
Todos os <i>Cebus</i>	1,98958 #	0,86615#	1,34153#

$P > 0,05$ (não significante).

Numa outra análise foram consideradas todas as 24 amostras de primatas deste estudo, (N = 24, ou seja, um indivíduo por espécie; ver Tabelas 3 e 4, Figura 6 e 7).

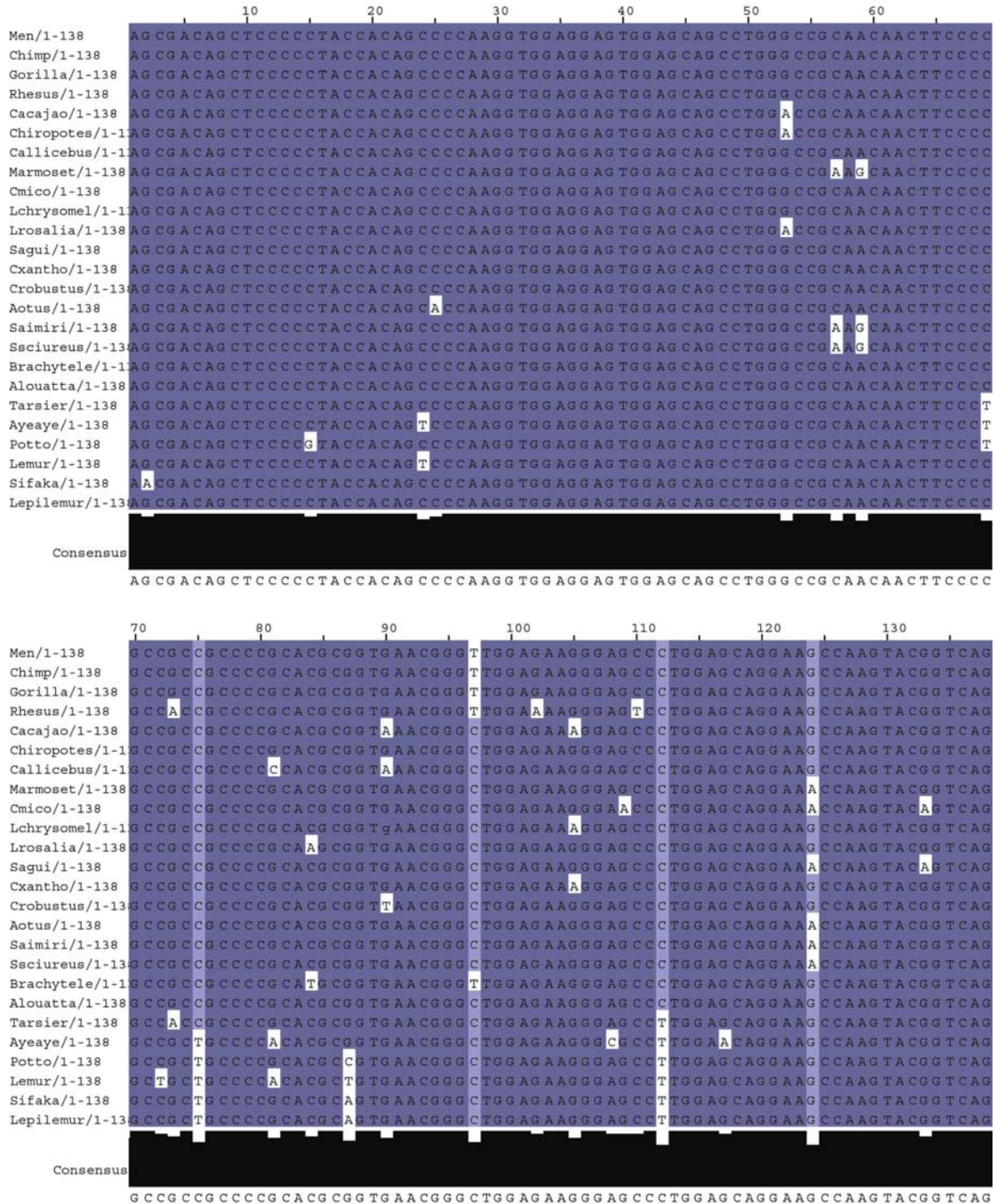


FIGURA 6. Alinhamento das seqüências dos nucleotídeos das 25 espécies de primatas, incluindo o homem. Gerado pelo *Jalview*.

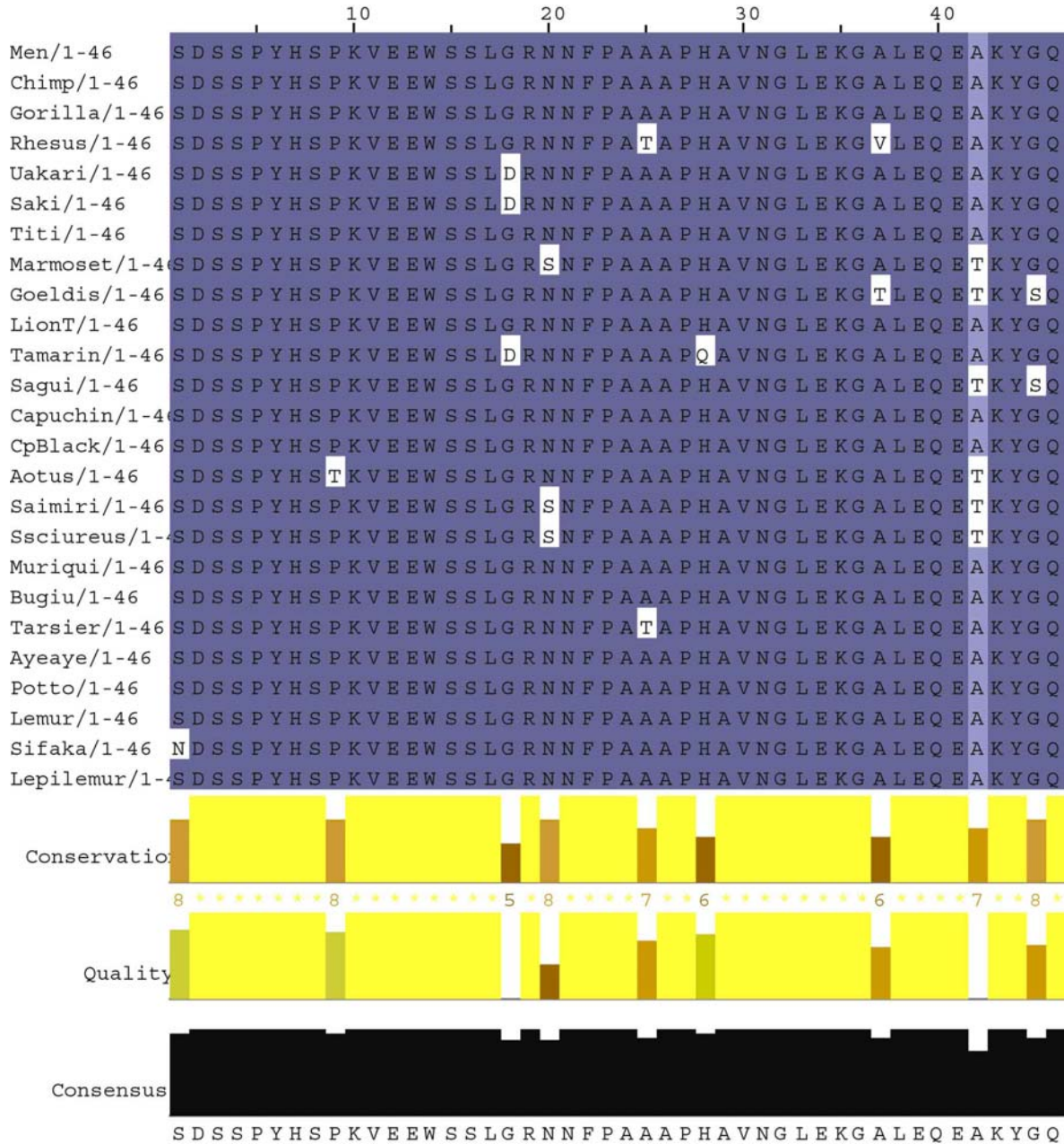


FIGURA 7. Alinhamento das seqüências dos aminoácidos das 25 espécies de primatas, incluindo o homem. Gerado pelo *Jalview*.

Estudos anteriores detectaram que o aminoácido treonina (Ala253Thr) estava presente em todas as espécies de macacos do Novo Mundo estudadas até então (Pereira *et cols.*, 2006; e Perry *et cols.*, 2006). Nestes trabalhos prévios foi levantada a hipótese de que este alelo estaria fixado na linhagem dos *Platyrrhini*. Nas espécies estudadas nesta dissertação esta hipótese não pode ser confirmada. O alelo treonina não foi encontrado em nenhum dos 25 indivíduos seqüenciados.

As taxas de substituições sinônimas e não-sinônimas (dN/dS), foram determinadas utilizando-se o programa *YN00*, do pacote do PAML, e podem ser vistas na tabela 8. Para o cálculo desta taxa foi utilizada uma árvore filogenética baseada em dados de marcadores neutros publicados por Schneider *et cols.* (2004) e Horvath *et cols.* (2007 e 2008) (Figura 8). Através da observação da ilustração apresentada na figura 9, pode-se perceber que embora os valores de dN/dS variem quando as espécies de primatas são comparadas, ele nunca excede 1, sendo que o maior valor da estatística envolve as espécies de lêmures *versus* os demais. Já a figura 10 destaca os valores considerando os humanos *versus* cada uma das outras espécies de primatas. Também se observa que os maiores valores são quando os lêmures são considerados. Isso poderia estar indicando algum sinal seletivo, ou ainda ser um sinal filogenético, já que os lêmures são distantes de todos os demais.

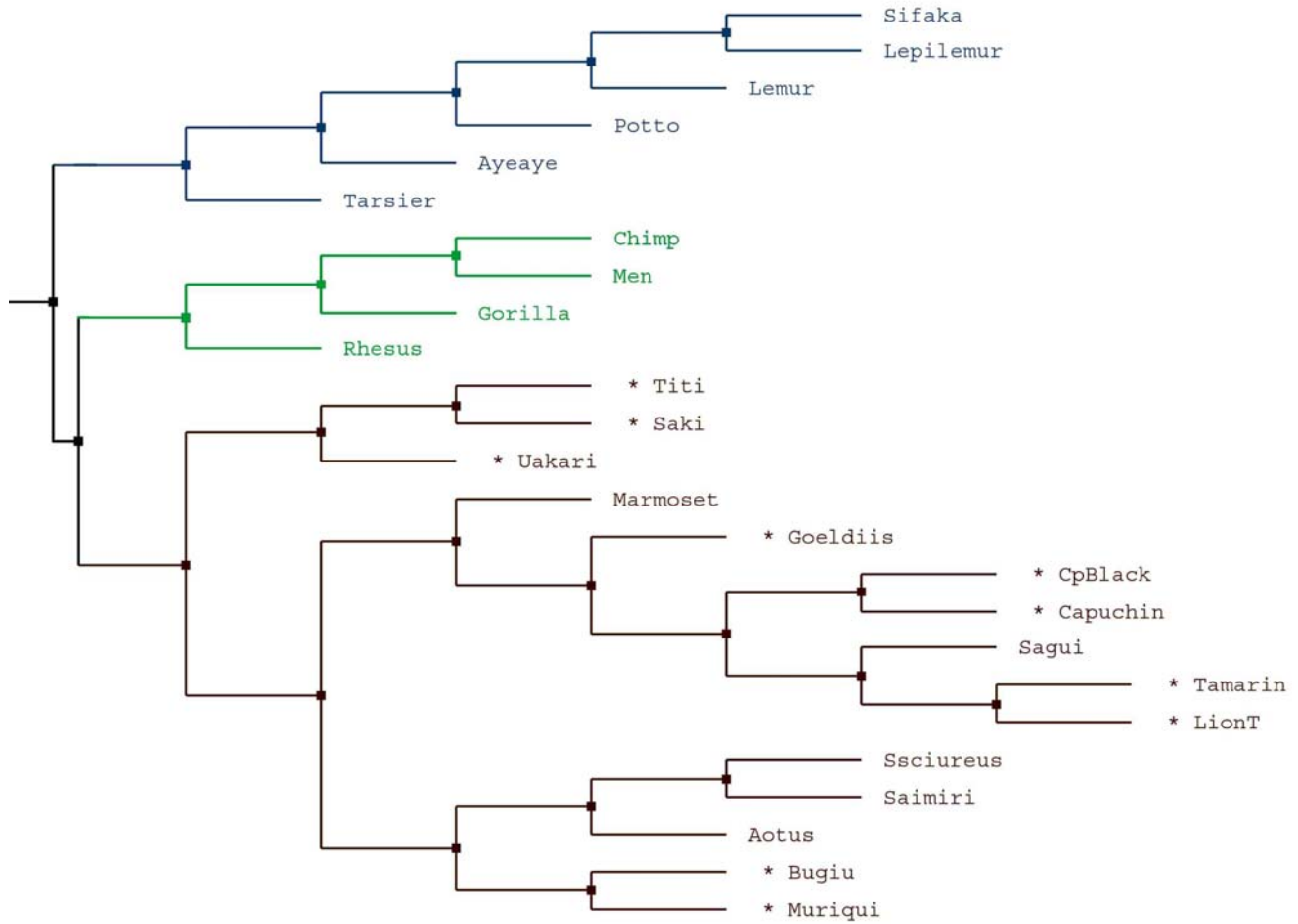


FIGURA 8. Árvore utilizada para a análise do exon 3 em primatas, construída baseada nas relações filogenéticas estabelecidas com marcadores neutros publicados por Schneider e tal (2004) e Horvath e cols. (2007 e 2008).

TABELA 7. Taxa de substituição sinônima por não sinônima (dN / dS) considerando o exon 3 do gene PAX9 de 25 espécies de primatas.

	Men	Chimp	Gorilla	Rhesus	Ualari	Saki	TI	Marmoset	Goeldi	Leont	Tamarin	Sagui	Capuchin	CpBbock	Aotus	Saimiri	Sacileurea	Muriqui	Bugiu	Tanarir	Ayseye	Polto	Lemur	Sibia	
Chimp	0																								
Gorilla	0	0																							
Rhesus	0,0688	0,0688	0,0688																						
Ualari	0,0983	0,0983	0,0983	0,1734																					
Saki	0,0448	0,0448	0,0448	0,1147	0,0482																				
TI	0,0806	0,0806	0,0806	0,1619	0,0702	0,0709																			
Marmoset	0,0927	0,0927	0,0927	0,1675	0,1444	0,0884	0,121																		
Goeldi	0,0884	0,0884	0,0884	0,1695	0,1374	0,0888	0,1159	0,0886																	
Leont	0,0484	0,0484	0,0484	0,1201	0,0446	0,0447	0,0734	0,0928	0,0882																
Tamarin	0,0688	0,0688	0,0688	0,1373	0,0686	0,0219	0,0928	0,1124	0,1125	0,0688															
Sagui	0,067	0,067	0,067	0,1376	0,1147	0,0683	0,0934	0,0672	0,0219	0,0689	0,0888														
Capuchin	0,0484	0,0484	0,0484	0,1201	0,0446	0,0447	0,0734	0,0928	0,0882	0	0,0688	0,0689													
CpBbock	0,0487	0,0487	0,0487	0,1225	0,0689	0,0449	0,0491	0,0929	0,0886	0,0483	0,067	0,0671	0,0483												
Aotus	0,067	0,067	0,067	0,1378	0,1147	0,0684	0,0931	0,0671	0,0684	0,0689	0,089	0,044	0,0689	0,0671											
Saimiri	0,0927	0,0927	0,0927	0,1675	0,1444	0,0884	0,121	0	0,0886	0,0928	0,1124	0,0672	0,0928	0,0929	0,0671										
Sacileurea	0,0927	0,0927	0,0927	0,1675	0,1444	0,0884	0,121	0	0,0886	0,0928	0,1124	0,0672	0,0928	0,0929	0,0671	0									
Muriqui	0,0228	0,0228	0,0228	0,0918	0,1278	0,0701	0,1155	0,1218	0,1148	0,0786	0,0687	0,0924	0,0786	0,0778	0,0928	0,1218	0,1218								
Bugiu	0,023	0,023	0,023	0,0928	0,07	0,0219	0,049	0,0671	0,0683	0,0229	0,044	0,0439	0,0229	0,0231	0,044	0,0671	0,0671	0,0482							
Tanarir	0,089	0,089	0,089	0,13	0,1508	0,0924	0,1386	0,1449	0,1376	0,096	0,1144	0,115	0,096	0,1004	0,115	0,1449	0,1449	0,1291	0,0702						
Ayseye	0,3889	0,3889	0,3889	0,6982	0,5178	0,3009	0,326	0,361	0,3686	0,3877	0,3031	0,3428	0,3877	0,3188	0,3287	0,361	0,361	0,4917	0,2768	0,1683					
Polto	0,2023	0,2023	0,2023	0,3073	0,2748	0,1715	0,246	0,226	0,2209	0,1941	0,1887	0,1978	0,1941	0,1841	0,1949	0,228	0,226	0,2648	0,149	0,0969	0,1682				
Lemur	0,2801	0,2801	0,2801	0,3818	0,3263	0,2188	0,272	0,2891	0,2884	0,2407	0,2286	0,2414	0,2407	0,2248	0,2376	0,2891	0,2891	0,3068	0,191	0,208	0,1304	0,1791			
Sibia	0,1291	0,1291	0,1291	0,2069	0,1827	0,1189	0,1809	0,1786	0,1686	0,1275	0,1414	0,1428	0,1275	0,154	0,1426	0,1786	0,1786	0,1687	0,0976	0,119	0,2084	0,0988	0,1286		
Lepilemur	0,107	0,107	0,107	0,1844	0,1604	0,0977	0,1582	0,1559	0,1429	0,1055	0,1188	0,1204	0,1055	0,1118	0,12	0,1559	0,1559	0,1387	0,0757	0,0988	0,1881	0,0739	0,1036	0,0219	

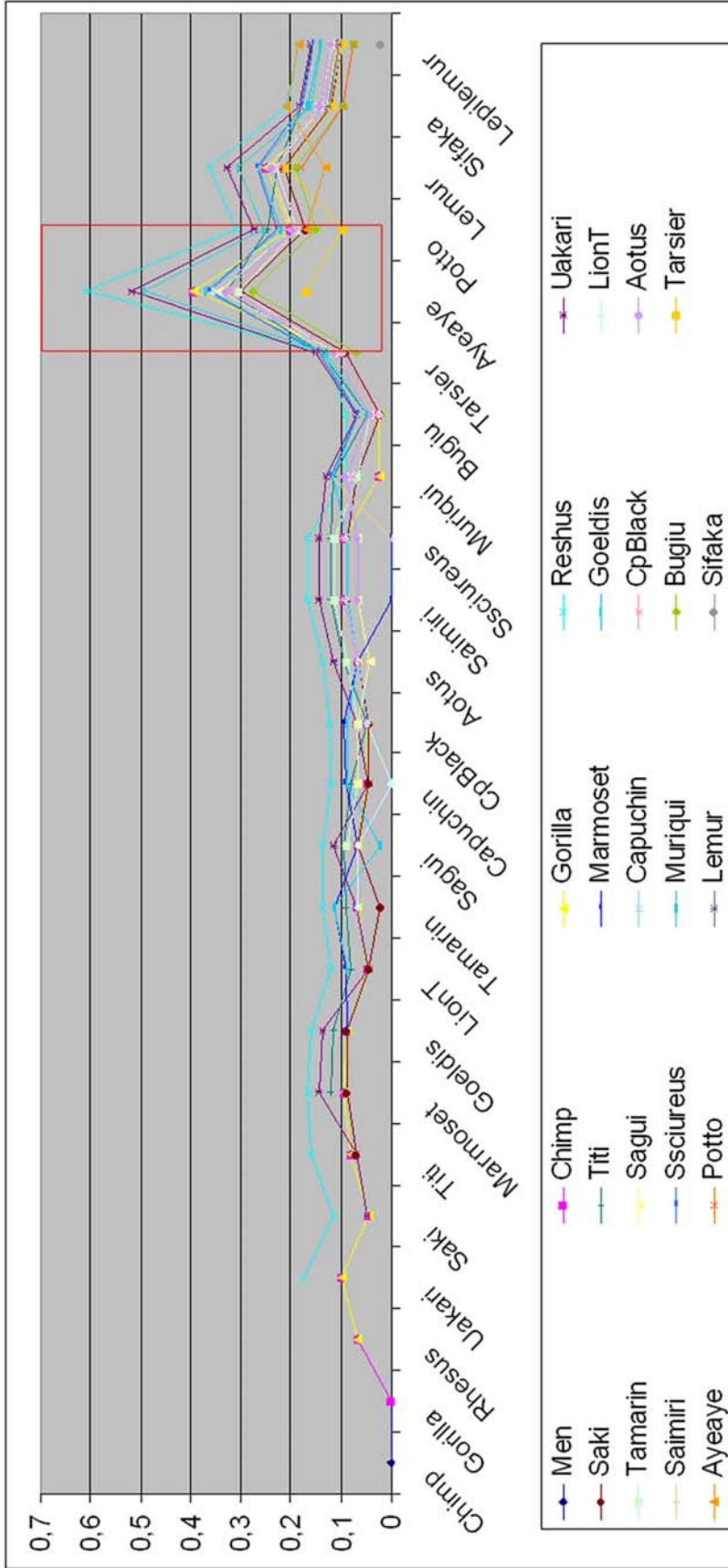


FIGURA 9. Gráfico com os valores de dN/dS dos 24 primatas presentes no estudo, observe, em destaque que os maiores valores são entre o grupo dos macacos e o dos lêmures. Valores de dN/dS obtidos no PAML..

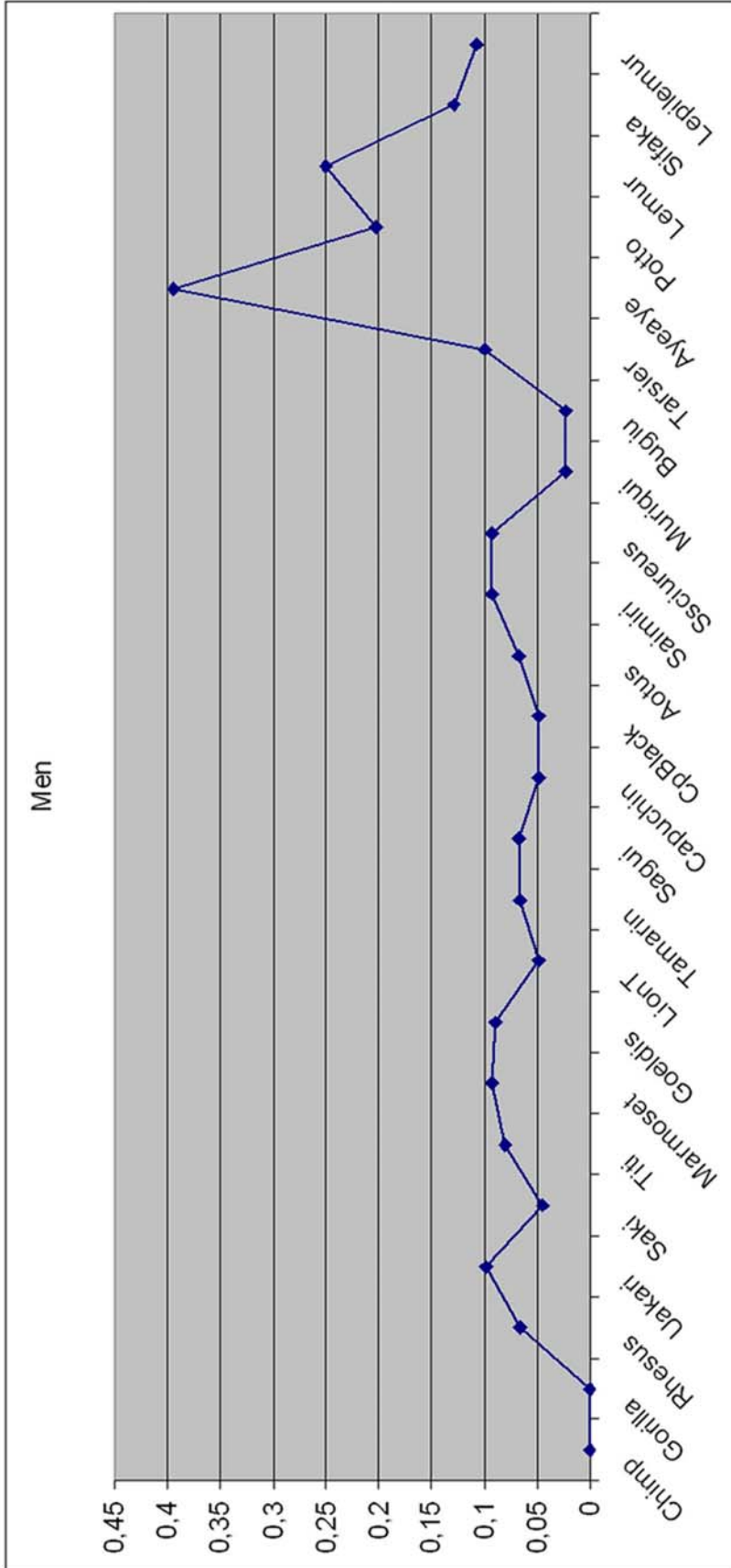


FIGURA 10. Gráfico com os valores de dN/dS dos 24 primatas presentes no estudo, em destaque a variação dos humanos. Valores de dN/dS obtidos no PAML.

Para tentar discriminar entre as duas possibilidades citadas acima, foi testado qual o melhor modelo de substituição de códons se aplica nesse caso, usando o programa *CODEML*, do pacote do PAML (ver Apêndice I).

Quando os modelos (M0 - M3, M1a - M2a e M7 - M8) são comparados por máxima verossimilhança, o melhor modelo é o M3 ($p < 0,001$), que admite variação nas classes de ω (Tabela 8). Ou seja, indica que a pressão seletiva varia através dos sítios de aminoácidos, o que já tinha sido observado através das figuras 9 e 10. Foi verificado que cerca de 27% dos códons estão sobre um regime de relaxamento funcional ($\omega_2 = 0,28387$). Desse modo, se algum sinal seletivo existe, ele não seria forte suficiente para gerar significância estatística. Concluindo, nesse momento só podemos sugerir que a variação nas taxas dN/dS entre os primatas são resultado de um afrouxamento da pressão seletiva.

TABELA 8. Parâmetros estimados sobre vários modelos de substituição de códons considerando o exon 3 em primatas.

Modelo ^a	dN/dS ^b	Parâmetros estimados ^c	ℓ	p -value
M0: <i>one-ratio</i> (1)	0,07316	$\omega = 0,07316$	-432,402917	<0,001
M3: <i>discrete</i> (5)	0,0759	$p_0 = 0,23127, p_1 = 0,502298, (p_2 = 0,26755)$ $\omega_0 = 0, \omega_1 = 0, \omega_2 = 0,28387$	-424,382447	
M1a: <i>neutral</i> (1)	0,1286	$p_0 = 0,90613, (p_1 = 0,09387)$ $(\omega_0 = 0,03831), (\omega_1 = 1)$	-428,159334	0,999
M2a: <i>selection</i> (3)	0,1286	$p_0 = 0,90613, p_1 = 0,09387, (p_2 = 0)$ ($\omega_0 = 0,03832$), ($\omega_1 = 1$), $\omega_2 = 16,86772$	-428,159334	
M7: <i>beta</i> (2)	0,0739	$P = 0,13682$ $q = 1,48168$	-425,274395	
M8: <i>beta&ω</i> (4)	0,1077	$p_0 = 0,98922$ ($p_1 = 0,01078$) $p = 0,14690$ $q = 1,93778$ $\omega = 3,21739$	-424,561039	0,489

a O número após o código modelo, entre parênteses, é o número de parâmetros livres na distribuição de ω . **b** O valor dN/dS ratio é um valor médio de todos os códons presentes no alinhamento do exon 3 do gene *PAX9*. **c** os parâmetros em parênteses não são parâmetros livres; ω dN/dS ratio; p , proporção de códons na classe de ω ; M0-M3: χ^2 df=4; M1a-M2a: χ^2 df=2; M7-M8: χ^2 df=2; LTR: $2\Delta \ell = 2(\ell - \ell_0)$.

III-2.2. A variação da porção codificadora do gene *PAX9* em mamíferos

Como pode ser visto nos dados apresentados até aqui, parece que pelo menos o exon 3 do *PAX9* está sujeito a diferentes pressões evolutivas considerando os primatas. A fim de se ampliar o conhecimento sobre o tema, as análises foram expandidas para os exons 2 e 4 considerando as outras espécies de mamíferos disponíveis para esse estudo. Isso permitiu atender o objetivo 2 dessa dissertação.

Primeiramente foram analisadas comparativamente as seqüências, dos exons 2, 3 e 4 considerando as 37 espécies, ou seja, todas aquelas que apresentavam dados completos para todos os 3 segmentos (Tabela 9 e Figura 10).

O resultado mais marcante é a grande diferença entre a porcentagem de mutações não-sinônimas encontradas no exon 3 quando este é comparado ao exon 2; enquanto somente 2% das mutações do exon 2 trocam o aminoácido, no exon 3 acontece à mudança em cerca de 32%. Isso indica que o fenômeno já observado nos humanos parece ser algo mais geral considerando outros mamíferos.

TABELA 9. Dados sobre três exons do gene *PAX9* em 37 espécies de mamíferos.

PAX9	Sítios (analisados)	Códons (analisados)	Sítios sinônimos (SIL)	Sítios não- sinônimos (NYS)	Mut (Tot)	SilMut (%)	NSynMut (%)
Exon 2 (627 pb)	555	185	147,57	407,43	237*/215	210(98)	5(2)
Exon 3 (138 pb)	135	45	30,32	104,68	67*/35	24(68)	11(32)
Exon 4 (255 pb)	240	85	62,73	177,27	104*/79	69(87)	10(13)

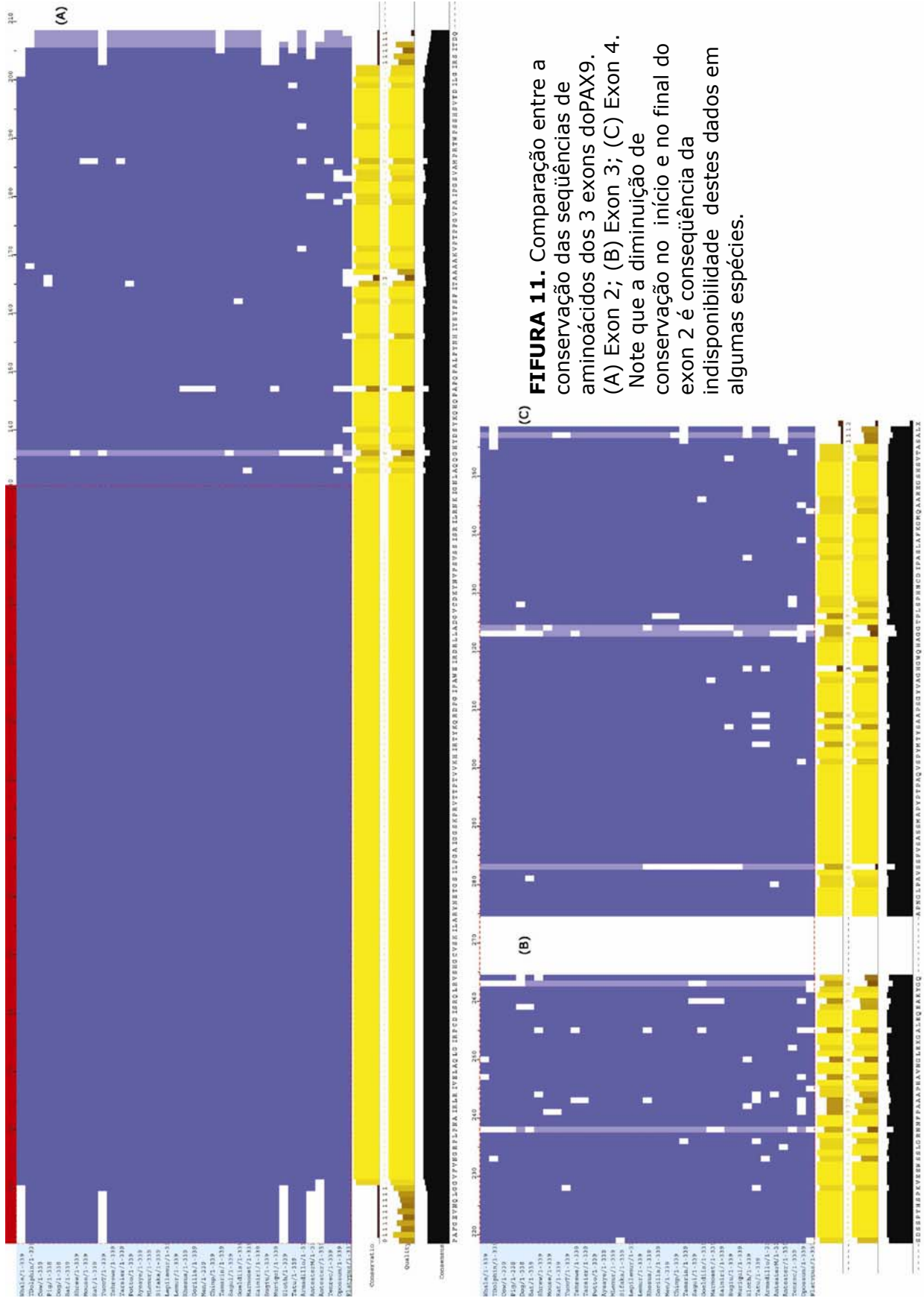
Sítios com *gaps* ou dados incompletos são excluídos da análise; Quando códons complexos, foram detectados as mutações sinônimas e não-sinônimas não foram contabilizadas *

É esperado que uma onipresente seleção purificadora domine completamente o panorama evolutivo do exon 2, onde está o domínio pareado (sinalizado com uma barra vermelha na figura 11) de um gene importante para o desenvolvimento. São 384 nucleotídeos envolvidos no domínio pareado e ali não há nenhuma alteração que leve a mudança de

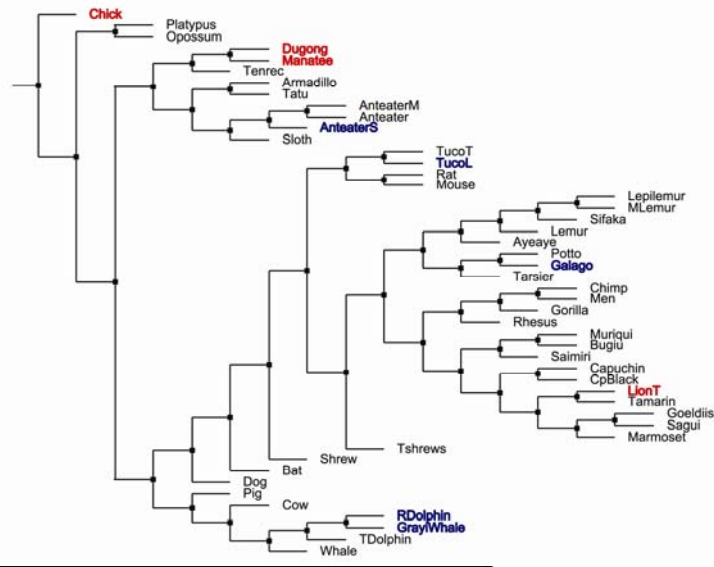
aminoácidos como pode ser visto no bloco azul compacto na figura 9(A). Ou seja, são exatamente os mesmos 128 aminoácidos para todos os 37 mamíferos analisados, entre eles o homem, a baleia e o ornitorrinco.

Já os exons 3 e 4 mostram importante variação em nível de aminoácidos (Figura 9B e C). Testes foram realizados considerando cada exon em separado para avaliar quais possíveis forças evolutivas estariam atuando na manutenção dessa diversidade.

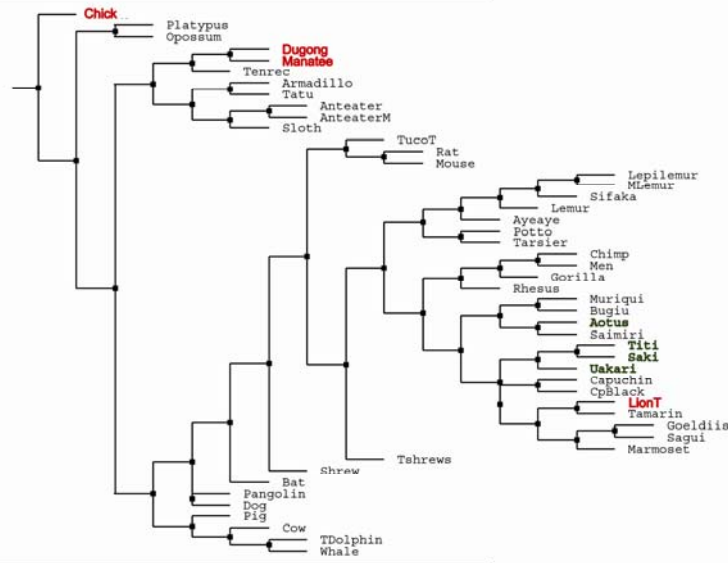
Para a análise específica de cada exon, utilizou-se uma árvore (Figura 12) baseada nas relações filogenéticas estabelecidas com marcadores neutros publicados por Murphy *et cols.* (2001), Eizirich *et cols.* (2001), Schneider *et cols.* (2004) e Horvath *et cols.* (2007 e 2008). Essa construção se faz necessária para rodar o programa PALM (ver material e métodos no Apêndice I).



Exon 2



Exon 3



Exon 4

■ Espécies presentes somente na análise dos exons 2 e 3;

■ Espécies presentes somente na análise dos exons 2 e 4;

■ Espécies presentes somente na análise do exon 3;

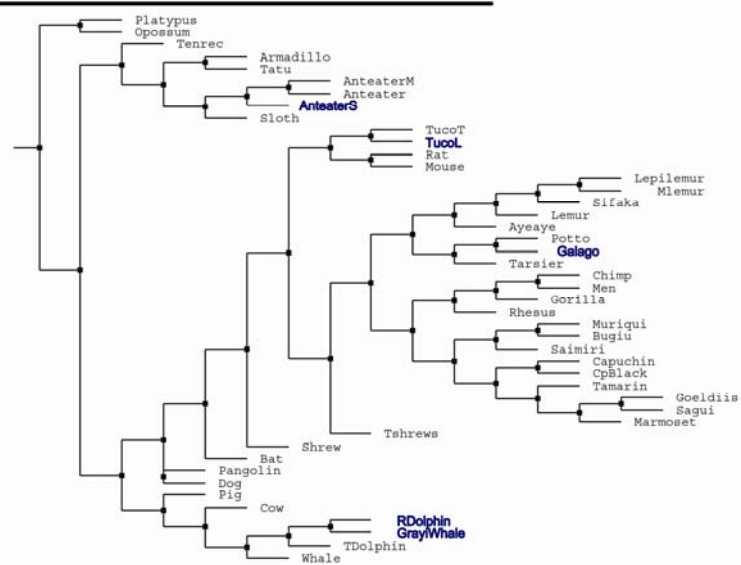


FIGURA 12. Árvores utilizadas para a análise dos exons 2, 3 e 4 nos mamíferos, foi construída baseada nas relações filogenéticas estabelecidas com marcadores neutros publicados por Murphy e cols. (2001), Eizirich e cols. (2001) Schneider e tal (2004) e Horvath e cols. (2007 e 2008);

III-2.2.1. Onipresente seleção purificadora no exon 2

Na análise específica do exon 2, quarenta e oito espécies puderam ser comparadas. Os resultados dos cálculos das taxas de substituições, bem como os modelos encontram-se na tabela 10. Quando os modelos de substituição são comparados (M0 – M3, M1a – M2a e M7 e M8), o que melhor se adapta aos dados é o M3 ($p < 0,001$), que admite variação nas classes de ω . A taxa de substituições sinônimas e não-sinônimas ($dN/dS - \omega = 0,00759$) é um valor significativamente menor que zero, sinalizando que a principal força agindo no exon 2 de mamíferos é, como esperado, seleção purificadora. Além disso, como o modelo M3 permite ramos com valores diferentes, encontramos que 92% dos sítios dentro do exon 2 estão sujeitos a uma taxa substituições sinônimas e não-sinônimas muito baixa (menor que 0,00125). Isso significa que uma esmagadora constrição funcional mantém a seqüência de aminoácidos inalterada por todas as linhagens dos mamíferos estudadas na presente dissertação.

TABELA 10. Parâmetros estimados sobre vários modelos de substituição de códons considerando o exon 2 em mamíferos.

Modelo ^a	dN/dS ^b	Parâmetros estimados ^c	ℓ	p-value
M0: <i>one-ratio</i> (1)	0,00732	$\omega = 0,0073$	-3678.522512	
M3: <i>discrete</i> (5)	0,00759	$p_0 = 0,44949, p_1 = 0,47187,$ $(p_2 = 0, 07863)$ $\omega_0 = 0,00124, \omega_1 = 0,00125 \omega_2 =$ $0,08564$	-3643.980262	<0,001
M1a: <i>neutral</i> (1)	0,1286	$p_0 = 1, (p_1 = 0)$ $(\omega_0 = 0,00732), (\omega_1 = 1)$	-3678.522661	0,999
M2a: <i>selection</i> (3)	0,1286	$p_0 = 1, p_1 = 0, (p_2 = 0) (\omega_0 = 0,00732),$ $(\omega_1 = 1), \omega_2 = 13,95102$	-3678.522512	
M7: <i>beta</i> (2)	0.0085	$P = 0.05618 \quad q = 4.42657$	-3644.412323	
M8: <i>beta&ω</i> (4)	0.0085	$p_0 = 1 (p_1 = 0) \quad p = 0.05618$ $q = 4.42656 \quad \omega = 2.20614$	-3644.412504	0,999

a O número após o código modelo, entre parênteses, é o número de parâmetros livres na distribuição de ω .

b O valor dN/dS ratio é um valor médio de todos os códons presentes no alinhamento do exon 2 do gene PAX9. **c** os parâmetros em parênteses não são parâmetros livres; ω dN/dS ratio; p , proporção de códons na classe de ω ; M0-M3: χ^2 df=4; M1a-M2a: χ^2 df=2; M7-M8: χ^2 df=2; LTR: $2\Delta \ell = 2(\ell_1 - \ell_0)$.

III-2.2.2. O exon 3 na linhagem dos mamíferos

Como já foi sinalizado pela imagem da figura 11 (B), o exon 3 é o mais variável também considerando as quarenta e sete espécies de mamíferos. Nesse caso, pode-se perguntar: que forças evolutivas estariam envolvidas na manutenção e no padrão dessa variação?

Os resultados dos cálculos das taxas de substituições, bem como os modelos encontram-se na tabela 11. Quando os modelos (M0 – M3, M1a – M2a e M7 - M8) são comparados, o melhor deles é o M3 ($p < 0,001$), que admite variação nas classes de ω .

TABELA 11. Parâmetros estimados sobre vários modelos de substituição de códons considerando o exon 3 em mamíferos.

Modelo ^a	dN/dS ^b	Parâmetros estimados ^c	ℓ	p-value
M0: <i>one-ratio</i> (1)	0,03925	$\omega = 0,03925$	-1096,684136	<0,001
M3: <i>discrete</i> (5)	0,0435	$p_0 = 0,27719, p_1 = 0,51180,$ $(p_2 = 0,21101)$ $\omega_0 = 0, \omega_1 = 0,03189, \omega_2 = 0,12860$	-1083,183286	
M1a: <i>neutral</i> (1)	0,0588	$p_0 = 0,97783, (p_1 = 0,02217)$ $(\omega_0 = 0,03744), (\omega_1 = 1)$	-1092,257013	0,960
M2a: <i>selection</i> (3)	0,0582	$p_0 = 0,97783, p_1 = 0,00794,$ $(p_2 = 0,01422) (\omega_0 = 0,03687), (\omega_1 = 1),$ $\omega_2 = 1$	-1092,209749	
M7: <i>beta</i> (2)	0,0437	$P = 0,61508 q = 12,86620$	-1083,283858	0,572
M8: <i>beta&ω</i> (4)	0,0573	$p_0 = 0,98156 (p_1 = 0,01844)$ $p = 0,71422 q = 16,64557 \omega = 1$	-1082,726387	

a O número após o código modelo, entre parênteses, é o número de parâmetros livres na distribuição de ω .

b O valor *dN/dS ratio* é um valor médio de todos os códons presentes no alinhamento do exon 2 do gene *PAX9*. **c** os parâmetros em parênteses não são parâmetros livres; ω *dN/dS ratio*; p , proporção de códons na classe de ω ; M0-M3: χ^2 df=4; M1a-M2a: χ^2 df=2; M7-M8: χ^2 df=2; LTR: $2\Delta \ell = 2(\ell_1 - \ell_0)$.

O valor médio da taxa de substituições sinônimas e não-sinônimas é menor do que 1 ($dN/dS = 0,0435$), indicando seleção purificadora em 73% dos códons. Entretanto, o modelo de substituição M3, que tem o melhor ajuste, permite taxas de substituições diferentes entre os códons, e encontramos que 21% dos sítios estão sujeitos a uma taxa de substituições sinônimas e não-sinônimas igual à $\omega_2 = 0,12860$, indicando que pelo menos um quarto dos códons do exon 3 encontrariam-se sob relaxamento funcional. Como já foi mencionada no manuscrito, a seleção purificadora pode mascarar seleção positiva em alguma linhagem ou ramo filogenético específico. Análises utilizando

parâmetros de maneira mais adstringente deverão ser utilizadas futuramente para avaliar mais detalhadamente se isso poderia estar ocorrendo nesse caso.

III-2.2.3. A variação do exon 4 do gene PAX9 em mamíferos

Para a análise envolvendo o exon 4, quarenta e cinco espécies de mamíferos foram consideradas. Os resultados dos cálculos das taxas de substituições, bem como os modelos encontram-se na tabela 12. Quando os modelos (M0 – M3, M1a – M2a e M7 - M8) são comparados, novamente é o M3 que melhor se ajusta ($p < 0,001$). A taxa de substituições sinônimas e não-sinônimas ($dN/dS - \omega^2 = 1,13501$) é um pouco maior que 1. Isso significa que deve haver sítios sujeitos a seleção Darwiniana. De fato, as análises indicam que pelo menos 2% dos sítios nesse exon estariam sob ação de seleção positiva. Adicionalmente, a tabela 13 mostra os valores de ω para cada códon. Pode-se observar que em dois códons 49 e 50, correspondentes aos aminoácidos 306 e 307 respectivamente, podem estar sujeitos à seleção positiva, com uma probabilidade posterior de mais de 99%.

TABELA 12. Parâmetros estimados sobre vários modelos de substituição de códons considerando o exon 4 em mamíferos.

Modelo ^a	dN/dS ^b	Parâmetros estimados ^c	ℓ	p-value
M0: <i>one-ratio</i> (1)	0,09330	$\omega = 0,09330$	-1633,037362	<0,001
M3: <i>discrete</i> (5)	0,0969	$p_0 = 0,50738, p_1 = 0,46305,$ $(p_2 = 0,02957)$ $\omega_0 = 0, \omega_1 = 0,13670, \omega_2 = 1,13501$	-1611,463377	
M1a: <i>neutral</i> (1)	0,1017	$p_0 = 0,95445, (p_1 = 0,04555)$ $(\omega_0 = 0,05884), (\omega_1 = 1)$	-1614,496761	0,999
M2a: <i>selection</i> (3)	0,1017	$p_0 = 0,95445, p_1 = 0,04555, (p_2 = 0)$ $(\omega_0 = 0,05885), (\omega_1 = 1),$ $\omega_2 = 33,30659$	-1614,496761	
M7: <i>beta</i> (2)	0,1037	$P = 0,23721 q = 1,95891$	-1614,665353	0,999
M8: <i>beta&ω</i> (4)	0,1038	$p_0 = 1 (p_1 = 0) p = 0,23721 q =$ $1,95891 \omega = 3,54727$	-1614,665432	

a O número após o código modelo, entre parênteses, é o número de parâmetros livres na distribuição de ω .

b O valor *dN/dS ratio* é um valor médio de todos os códons presentes no alinhamento do exon 4 do gene *PAX9*. **c** os parâmetros em parênteses não são parâmetros livres; ω *dN/dS ratio*; p , proporção de códons na classe de ω ; M0-M3: χ^2 df=4; M1a-M2a: χ^2 df=2; M7-M8: χ^2 df=2; LTR: $2\Delta \ell = 2(\ell_1 - \ell_0)$.

TABELA 13. Análise da probabilidade posterior *Naive Empirical Bayes* (NEB) – M3, de cada códon do exon 4.

AA	ω 0 %	ω 1%	ω 2%	Valor ω	P
1 A	0,69301	0,30685	0,00014	0,042	0,000
2 P	0,67606	0,32371	0,00023	0,045	0,000
3 N	0,74158	0,25840	0,00002	0,035	0,000
4 G	0,74564	0,25435	0,00002	0,035	0,000
5 L	0,74534	0,25464	0,00002	0,035	0,000
6 P	0,00002	0,99385	0,00614	0,143	0,006
7 A	0,00003	0,99980	0,00017	0,137	0,000
8 V	0,72753	0,27244	0,00003	0,037	0,000
9 G	0,00002	0,99917	0,00081	0,137	0,001
10 S	0,74108	0,25890	0,00002	0,035	0,000
11 F	0,74490	0,25509	0,00002	0,035	0,000
12 V	0,72677	0,27319	0,00003	0,037	0,000
13 S	0,67547	0,32430	0,00023	0,045	0,000
14 A	0,69790	0,30198	0,00012	0,041	0,000
15 S	0,76875	0,23124	0,00001	0,032	0,000
16 S	0,73226	0,26771	0,00003	0,037	0,000
17 M	0,73498	0,26500	0,00002	0,036	0,000
18 A	0,79268	0,20731	0,00000	0,028	0,000
19 P	0,76796	0,23203	0,00001	0,032	0,000
20 Y	0,69540	0,30450	0,00010	0,042	0,000
21 P	0,76326	0,23673	0,00001	0,032	0,000
22 T	0,79872	0,20128	0,00000	0,028	0,000
23 P	0,69329	0,30658	0,00013	0,042	0,000
24 A	0,78153	0,21846	0,00001	0,030	0,000
25 Q	0,62952	0,36949	0,00099	0,052	0,001
26 V	0,72627	0,27370	0,00003	0,037	0,000
27 S	0,00002	0,99650	0,00348	0,140	0,003
28 P	0,76729	0,23270	0,00001	0,032	0,000
29 Y	0,69549	0,30441	0,00010	0,042	0,000
30 M	0,00000	0,99425	0,00575	0,142	0,006
31 T	0,79105	0,20895	0,00000	0,029	0,000
32 Y	0,69442	0,30547	0,00011	0,042	0,000
33 S	0,00000	0,94944	0,05056	0,187	0,051
34 A	0,79067	0,20933	0,00000	0,029	0,000
35 A	0,00000	0,99814	0,00186	0,139	0,002
36 P	0,76751	0,23248	0,00001	0,032	0,000
37 S	0,77886	0,22113	0,00000	0,030	0,000
38 G	0,74614	0,25384	0,00002	0,035	0,000
39 Y	0,70134	0,29858	0,00008	0,041	0,000
40 V	0,79856	0,20144	0,00000	0,028	0,000
41 A	0,00003	0,99978	0,00019	0,137	0,000
42 G	0,65679	0,34278	0,00042	0,047	0,000
43 H	0,00000	0,89878	0,10122	0,238	0,101
44 G	0,68817	0,31168	0,00015	0,043	0,000
45 W	0,62867	0,37032	0,00101	0,052	0,001
46 Q	0,63036	0,36868	0,00096	0,051	0,001
47 H	0,72386	0,27610	0,00004	0,038	0,000
48 A	0,00003	0,99976	0,00022	0,137	0,000
49 G	0,00000	0,00044	0,99956	1.135	0,999
50 G	0,00000	0,00228	0,99772	1.133	0,998
51 T	0,80482	0,19518	0,00000	0,027	0,000
52 S	0,00002	0,99467	0,00532	0,142	0,005
53 L	0,71475	0,28519	0,00005	0,039	0,000
54 S	0,00000	0,99749	0,00251	0,139	0,003
55 P	0,00000	0,99404	0,00596	0,143	0,006
56 H	0,71546	0,28449	0,00005	0,039	0,000
57 N	0,73211	0,26787	0,00003	0,037	0,000
58 C	0,70899	0,29095	0,00006	0,040	0,000
59 D	0,75394	0,24605	0,00001	0,034	0,000
60 I	0,76624	0,23375	0,00001	0,032	0,000

TABELA 13. Continuação

AA	ω 0 %	ω 1%	ω 2%	Valor ω	P
61 P	0,75366	0,24633	0,00001	0,034	0,000
62 A	0,00000	0,85785	0,14215	0,279	0,142
63 S	0,69676	0,30313	0,00011	0,042	0,000
64 L	0,65592	0,34366	0,00042	0,047	0,000
65 A	0,00002	0,99840	0,00158	0,138	0,002
66 F	0,00002	0,99918	0,00080	0,137	0,001
67 K	0,67933	0,32049	0,00018	0,044	0,000
68 G	0,66968	0,33003	0,00028	0,045	0,000
69 M	0,73498	0,26500	0,00002	0,036	0,000
70 Q	0,00001	0,98302	0,01697	0,154	0,017
71 A	0,00002	0,99638	0,00360	0,140	0,004
72 A	0,00003	0,99973	0,00024	0,137	0,000
73 R	0,65251	0,34700	0,00048	0,048	0,000
74 E	0,68590	0,31395	0,00015	0,043	0,000
75 G	0,74472	0,25527	0,00002	0,035	0,000
76 S	0,74079	0,25919	0,00002	0,035	0,000
77 H	0,72082	0,27914	0,00004	0,038	0,000
78 S	0,00000	0,99790	0,00210	0,139	0,002
79 V	0,00002	0,99973	0,00025	0,137	0,000
80 T	0,00000	0,99085	0,00915	0,146	0,009

Probabilidade de que cada códon específico estar em cada uma das 3 classes de ω ; A seqüência do homem foi usada como referência. AA: aminoácidos P: probabilidade posterior;

III-2.3. Convergência e Paralelismo Molecular

O terceiro objetivo desta dissertação foi o de procurar alterações no gene *PAX9* que fossem compartilhadas por animais que tivessem as estruturas dentais e craniofaciais similares, mas que não fossem próximos filogeneticamente.

Interessantemente foi possível observar uma mutação no exon 3, que leva a troca de um aminoácido que é somente compartilhada entre o pangolin e os tamanduás (Figura 13), animais com aparelhos bucais altamente especializados para captura de formigas e cupins, e que não tem parentesco próximo (Figura 12). Vale a pena destacar que a mutação aparece nas duas espécies pertencentes a dois gêneros, mas em nenhuma outra da ordem Xenarthra (Tatu: *Dasybus sp.*; e Preguiça: *Bradypus sp.*). No entanto, os testes não detectaram sinal de seleção. Por outro lado, pangolins e tamanduás também compartilham uma mutação não-sinônima no exon 4 (Figura 14). Nesse caso, a mutação está presente no Tenrec (*Echinops telfairi*) e no Muriqui (*Brachyteles sp.*), e alguns outros mamíferos, não estando presente nos demais membros da ordem Xenarthra estudados. Mas o fato importante que deve ser frisado é que este sítio tem mais de 99% de provavelmente posterior de estar sob a ação de seleção positiva (Tabela 14).

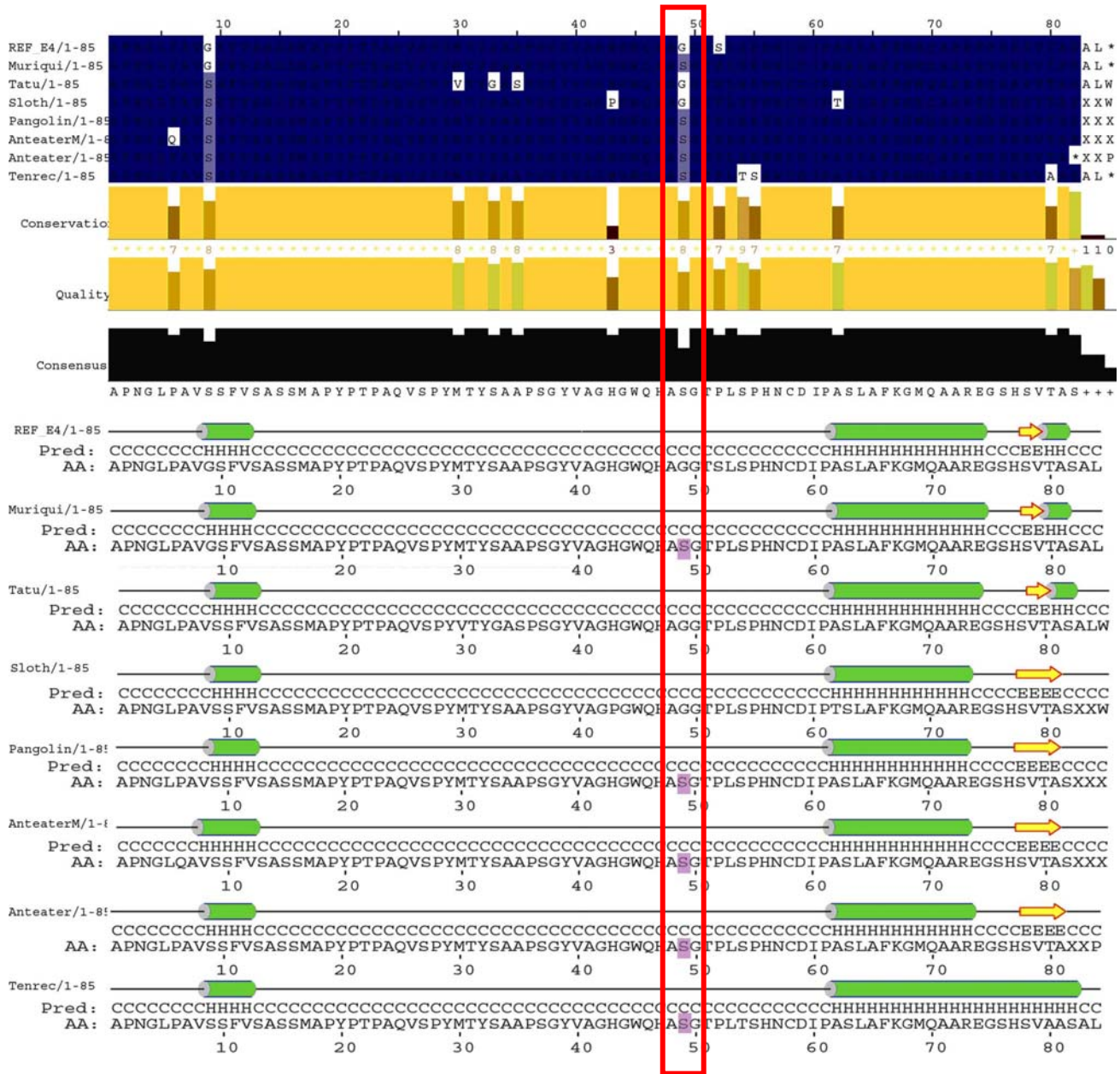


FIGURA 14. A mesma substituição nucleotídica no exon 4 em comum entre o pangolin e os tamanduás, o Tenrec e o Muriqui. A figura mostra que os demais componentes da ordem Xernathra não possuem esta substituição. A figura mostra a estrutura secundária do segmento da proteína correspondente ao exon 4, gerada pelo programa PSIPHRED.

TABELA 14. Análise da probabilidade posterior, *Naive Empirical Bayes* (NEB). Códon sobre a ação de seleção positiva (*: P>95%; **: P>99%)

Pr($\omega > 1$)	post mean +- SE for ω	
49 G	0.999**	1.135

Para investigar melhor a possibilidade de termos aqui uma convergência molecular, foi predito a estrutura secundária do segmento da proteína composto pelo exon 3 e exon 4, pois foi dentro desses dois exons mutações idênticas foram detectadas em tamanduás e pangolins.

Com relação ao exon 3 há algumas diferenças entre o fragmento da proteína dos tamanduás/pangolin com as demais mostradas na figura 11 (A) *versus* (B) e (C). Porém a estrutura secundária dos tamanduás/pangolins não difere de outras, tais como a encontrada em humanos (dado não mostrado). O mesmo também acontece quando o exon 4 foi modelado. Porém, não é possível descartar a possibilidade das diferenças estarem na estrutura terciária da proteína PAX9.

Infelizmente a falta de dados para o pangolin, relativo ao exon 2, impede que nesse momento seja feita análise semelhante com esse exon.

Mas é realmente instigante o fato destes animais compartilharem as mesmas mutações, pelo menos nos dois exons que puderam ser

comparados. Segundo Orr (2005), a probabilidade de evolução paralela é quase duas vezes maior sobre a ação da seleção natural do que sobre um regime neutro. Além disso, diversos trabalhos investigaram um gênero de peixes os *sticklebacks* que apresentam uma cintura pélvica com espinhos, (estudos prévios sugerem que esta estrutura protege os animais de predação) colonizaram ambientes de águas gélidas e em diversos momentos e em espécies diferentes ocorreu à redução e/ou a perda completa do esqueleto pélvico. Mutações no gene *Pitx1*, um gene de desenvolvimento, são a maior causa da redução pélvica neste sistema. A expressão alterada deste gene evoluiu múltiplas vezes em vários lagos separados no hemisfério norte (Shapiro e cols., 2004; Colosimo e cols., 2005). Outro caso instigante envolvendo convergência evolutiva é o papel do gene *PAX6* no desenvolvimento dos olhos altamente complexos dos cefalópodes e dos vertebrados (Gehring, 2001; Nilsson e cols., 2004). Deste modo, e considerando o que sabemos sobre o papel do *PAX9* no desenvolvimento de fórmulas dentais, é possível especular que estas mutações que ocorreram independentemente neste gene, possam estar envolvidas na cascata de modificações que levaram á morfologias similares encontradas nos tamanduás e pangolins.

Esses dados apontam para a idéia de que as mesmas soluções genéticas estariam sendo "privilegiadas" em espécies diferentes

(pangolin e tamanduás) sujeitas as mesmas pressões seletivas e *constraints* genéticos, já que mutações no mesmo gene seria o modo mais eficiente, ou ainda, a única maneira, de um fenótipo em particular evoluir de forma independente (Sucena *et cols.*, 2003; Wood *et cols.*, 2005; Brakefield, 2006; Carroll, 2006; Protas *et cols.*, 2006).

CAPÍTULO IV – Discussão

Pretende-se que o texto a seguir não seja repetitivo já que há uma discussão detalhada no manuscrito envolvendo os humanos, e os resultados envolvendo as outras espécies foram apresentados juntamente com comentários específicos. Dessa forma, aqui somente alguns aspectos serão explorados adicionalmente para ajudar no delineamento de um panorama geral sobre a evolução molecular do gene *PAX9* (objetivo 4 desta dissertação). Vale lembrar, contudo, que alguma sobreposição é inevitável visto a complexidade do tema abordado.

Inicialmente é importante destacar que modificações morfológicas são resultado da ação de um grande número de genes regulatórios e estruturais (Davidson, 2006), além de fatores epigenéticos e ambientais. Os resultados apresentados nessa dissertação apontam que o gene *PAX9* seria um dos elementos envolvidos nessa rede.

Foi possível verificar que as variações (ou ausência destas) contidas tanto em regiões codificadoras quanto não codificadoras do *PAX9* são mantidas por um conjunto de elementos que compõem um complexo cenário evolutivo intra-específico, bem como entre

espécies/gêneros/ordens. Embora a ação da seleção purificadora tenha um papel principal e determinístico na manutenção do gene como um todo em todas as espécies investigadas aqui, fugas deste padrão pontilham esta história.

Como foi visto para a espécie humana, logo após saída dos primeiros humanos modernos da África há um relaxamento da pressão seletiva sobre o exon 3, o que permitiu que novas mutações ali permanecessem. Provavelmente a mutação $G \rightarrow C$, Ala240Pro, que está ligada a agenesia do terceiro molar, teria sido selecionada positivamente fora da África. As altas frequências do alelo derivado na Europa e Ásia corroboram essa sugestão. Posteriormente, quando os primeiros migrantes entram na América, especificamente na América do Sul, outro cenário parece dominar a evolução do exon 3, seja pela perda inicial da possível vantagem adaptativa, ou por causa da impressionante ação da deriva genética na história dessas populações. Há um excesso de variantes raras nos ameríndios que indica claramente que em algum momento na história destas populações, fatores casuísticos possam ter sido preponderantes às restrições funcionais da proteína. Vale lembrar que a vantagem adaptativa pode não estar diretamente relacionada ao número de dentes, mas sim a características associadas que foram determinantes no complexo adaptativo selecionado para climas diferentes daqueles encontrados na África.

Considerando os mamíferos como um todo (incluindo humanos), poderosa ação de seleção purificadora mantém o exon 2 é extremamente conservado. Por outro lado, foi possível detectar um relaxamento da pressão seletiva em alguns códons, tanto do exon 3 quanto do exon 4. Foi observada também uma maior variabilidade em nível de aminoácidos no exon 3, do que no exon 4. Neste último, há indícios estatísticos da ação de seleção positiva, particularmente em dois códons específicos. Embora as análises preliminares não detaram nenhum sinal de seleção positiva envolvendo o exon 3 ela não pode ser descartada, visto o extraordinário caso da convergência molecular encontrado em espécies com morfologia dental e craniofacial similares, mas filogeneticamente distantes. Vale lembrar que a probabilidade de evolução paralela é quase duas vezes maior sobre a ação da seleção natural do que sobre um regime neutro (Orr, 2005). Carroll (2006), por sua vez, afirma que uma das maiores evidências da ubiquidade da seleção natural é que se dermos tempo suficiente, diferentes espécies, incluindo aquelas que pertencem a grupos taxonômicos completamente diferentes, irão responder da mesma maneira a uma condição seletiva similar.

Sendo assim, a variabilidade no exon 3 pode constituir o que Pereira e cols. (2006) chamou de "janela evolutiva do gene *PAX9*", ou seja, um segmento que estaria sujeito a mudanças inicialmente aleatórias com

possibilidade de posterior varredura adaptativa, selecionando mutações específicas que resultariam em algumas mudanças nas características morfológicas dependentes deste gene. Tomando como um todo o conjunto de dados apresentados aqui, é possível especular que o exon 3 seria um exemplo notável que conferiria “evolvabilidade” (Kirschner e Gerhart, 1998; Pigliucci, 2008) ao *PAX9*.

CAPÍTULO V - Referências Online

(sites de busca, ferramentas *online* e programas)

Genbank/NCBI

<http://www.ncbi.nlm.nih.gov/Genbank>

BLAST/NCBI

<http://www.ncbi.nlm.nih.gov/blast/Blast.cgi>

TaxBrowser/NCBI

<http://www.ncbi.nlm.nih.gov/sites/entrez?db=taxonomy>

dbSNP/NCBI

<http://www.ncbi.nlm.nih.gov/projects/SNP/>

HAPMAP

http://www.hapmap.org/cgi-perl/gbrowse/hapmap_B35/

Ensembl

<http://www.ensembl.org/index.html>

FastPCR

<http://www.biocenter.helsinki.fi/bi/Programs/fastpcr.htm>

PyPop

<http://www.pypop.org/>

Phred, Phrap and Consed

<http://www.genome.washington.edu>

Arlequin 3.11 software

<http://lgb.unige.ch/arlequin>

Mega 4.0

<http://www.megasoftware.net/>

DnaSP 4.10.9

<http://www.ub.es/dnasp/>

Jalview

<http://www.jalview.org/download.html>

PSIFRED

<http://bioinf.cs.ucl.ac.uk/psipred/psiform.html>

The ExpASy (Expert Protein Analysis System)

<http://www.expasy.org/>

DFSC 1.1 software

www.xmission.com/~wooding/DFSC/

PAML (Phylogenetic Analysis by Maximum Likelihood)

<http://abacus.gene.ucl.ac.uk/software/paml.html>

Animal Diversity Web - University of Michigan Museum of Zoology

<http://animaldiversity.ummz.umich.edu/site/index.html>

Mammalian Crania Photographic Archive

<http://taxa.soken.ac.jp/mcpa2/en/mammal.html>

GENECARDS

<http://www.genecards.org/index.shtml>

GENEATLAS

<http://www.dsi.univ-paris5.fr/genatlas/>

UCSC Genome Browser on Human

<http://genome.ucsc.edu/index.html?org=Human>

CAPÍTULO VI - Referências Bibliográficas

- Abzhanov A, Kuo WP, Hartmann C, Grant BR, Grant PR, Tabin CJ. (2006) The calmodulin pathway and evolution of elongated beak morphology in Darwin's finches. *Nature*;442(7102):563-7.
- Anisimova M, Bielawski J, Dunn K, Yang Z. Phylogenomic analysis of natural selection pressure in *Streptococcus* genomes. *BMC Evol Biol*;7:154.
- Alappat Silvia, Zhang Zun Yi, Chen Yi Ping (2003) Msx homeobox gene family and craniofacial development. *Cell research*; 13(6):429-442.
- Alonso Claudio R. and Wilkins Adam S (2005) The molecular elements that underlie developmental evolution. *Nat Rev Genet*; 6(9):709-715.
- Arnason U, Gullberg A, Janke A (1999) The mitochondrial DNA molecule of the aardvark, *Orycteropus afer*, and the position of the Tubulidentata in the eutherian tree. *Proc Biol Sci*; 266(1417):339-345.
- Artur Wallace (2000) The concept of developmental reprogramming and the quest for an inclusive theory of evolutionary mechanisms. *Evolution & Development* 2(1): 49 -57.
- Balczarek KA, Lai ZC, Kumar S (1997) Evolution of functional diversification of the paired box (Pax) DNA-binding domains. *Mol Biol Evol*; 14(8):829-42.
- Brakefield PM (2006) Evo-devo and constraints on selection. *Trends Ecol Evol*; 21(7):362-8.
- Breitling R, Gerber JK (2000) Origin of the paired domain. *Dev Genes Evol*; 210(12):644-50.
- Carroll SB (2006) *The Making of the Fittest – DNA and the ultimate forensic record of evolution*. 1ª ed. W. W. Norton & Company, Inc.

- Carroll SB (2005) Evolution at two levels: on genes and form. *PLoS Bio*; 3(7):e245.
- Carroll S B (2003) Genetics and the making of Homo sapiens. *Nature*; 422: 849-857.
- Carroll SB (2000) Endless forms: the evolution of gene regulation and morphological diversity. *Cell*; 101(6):577-80.
- Chen Y, Zhang Y, Jiang TX, Barlow AJ, St Amand TR, Hu Y, Heaney S, Francis-West P, Chuong CM, Maas R (2000) Conservation of early odontogenic signaling pathways in Aves. *Proc Natl Acad Sci U S A*; 9:10044-10049.
- Chi N. and Epstein J. A (2002) Getting your Pax straight: Pax proteins in development and disease. *Trends Genet.*; 18(1):41-7.
- Clamp, M., Cuff, J., Searle, S. M. and Barton, G. J. (2004), "The Jalview Java Alignment Editor", *Bioinformatics*, 20, 426-7.
- Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G Jr, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM (2005) Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science*; 307(5717): 1928-1933.
- Crow KD, Stadler PF, Lynch VJ, Amemiya C, Wagner GP (2006) The "fish-specific" Hox cluster duplication is coincident with the origin of teleosts. *Mol Biol Evol*; 23(1):121-36.
- Davidson EH, Erwin DH (2006) Gene regulatory networks and the evolution of animal body plans. *Science*;311(5762):796-800.
- Dahl E, Koseki H, Balling R(1997) Pax genes and organogenesis. *Bioessays* ; 19: 755-765.
- Delsuc F, Brinkmann H, Philippe H (2005) Phylogenomics and the reconstruction of the tree of life. *Nat Rev Genet*;6(5):361-75.
- Delsuc F, Scally M, Madsen O, Stanhope MJ, de Jong WW, Catzeflis FM, Springer MS, Douzery EJ (2002) Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. *Mol Biol Evol*; 19(10):1656-71.

- Dorus S, Anderson JR, Vallender EJ, Gilbert SL, Zhang L, Chemnick LG, Ryder OA, Li W, Lahn BT (2006) Sonic Hedgehog, a key development gene, experienced intensified molecular evolution in primates. *Hum Mol Genet*; 15(13):2031-7.
- Ebersberger I, Metzler D, Schwarz C, Paabo S (2002) Genome wide comparison of DNA sequences between humans and chimpanzees. *Am J Hum Genet*; 70:1490-1499.
- Eizirik E, Murphy WJ, O'Brien SJ.(2001) Molecular dating and biogeography of the early placental mammal radiation. *J Hered.*92(2):212-9.
- Enard W, Paabo S (2004) Comparative primate genomics. *Annu Rev Genomics Hum Genet*; 5:351-78.
- Enard W, Khaitovich P, Klose J, Zöllner S, Heissig F, Giavalisco P, Nieselt-Struwe K, Muchmore E, Varki A, Ravid R, Doxiadis GM, Bontrop RE, Pääbo S (2002) Molecular evolution of FOXP2, a gene involved in speech and language. *Science*; 296: 340-342.
- Espinoza RE, Wiens JJ, Tracy CR (2006) Recurrent evolution of herbivory in small, cold-climate lizards: breaking the ecophysiological rules of reptilian herbivory. *Proc Natl Acad Sci U S A* ;101(48):16819-24.
- Epstein J, Cai J, Glaser T, Jepeal L, Maas R (1994) Identification of a Pax paired domain recognition sequence and evidence for DNA-dependent conformational changes. *J Biol Chem*; 269(11):8355-61.
- Ewens, W. J. The sampling theory of selectively neutral alleles. *Theor Popul Biol*, v.3, n.1, Mar, p.87-112. 1972.
- Fagundes NJ, Ray N, Beaumont M, Neuenschwander S, Salzano FM, Bonatto SL, Excoffier L (2007) Statistical evaluation of alternative models of human evolution. *Proc Natl Acad Sci U S A* 104(45):17614-17619.
- Fernald RD (2004). Evolving eyes. *Int J Dev Biol*;48(8-9):701-5.
- Ferrier DE, Minguillon C (2003) Evolution of the Hox/ParaHox gene clusters. *Int J Dev Biol*; 47(7-8):605-11.

- Fisher SE, Marcus GF (2006) The eloquent ape: genes, brains and the evolution of language. *Nat Rev Genet*; 7(1):9-20.
- Fondon JW 3rd, Garner HR (2004) Molecular origins of rapid and continuous morphological evolution. *Proc Natl Acad Sci U S A*; 101(52):18058-63.
- Foster SA, Baker JA (2004) Evolution in parallel: new insights from a classic system. *Trends Ecol Evol* ;19(9):456-9.
- Frazier-Bowers SA, Guo DC, Cavender A, Xue L, Evans B, King T, Milewicz D, D'Souza RN (2002) A novel mutation in human PAX9 causes molar oligodontia. *J Dent Res*; 81:129-133.
- Galliot B, de Vargas C, Miller D (1999) Evolution of homeobox genes: Q50 Paired-like genes founded the Paired class. *Dev Genes Evol*;209(3):186-197.
- Gehring WJ (2001) The genetic control of eye development and its implications for the evolution of the various eye-types. *Zoology* 104: 171-183.
- Gilbert SF (2005). Mechanisms for the environmental regulation of gene expression: ecological aspects of animal development. *J Biosci*; 30(1):65-74.
- Gilbert SF (2005) Opening Darwin's black box: teaching evolution through developmental genetics. *Nat Rev Genet*;4(9):735-41.
- Gnarra JR, Dressler GR (1995) Expression of Pax-2 in human renal cell carcinoma and growth inhibition by antisense oligonucleotides. *Cancer Res*; 55: 4092-4098.
- Hall BK (2003) Evo-Devo: evolutionary developmental mechanisms. *Int J Dev Biol*; 47(7-8):491-5.
- Harris MP, Hasso SM, Ferguson MW, Fallon JF (2006) The development of archosaurian first-generation teeth in a chicken mutant. *Curr Biol*;16(4):371-7.
- Harris WA (1997) Pax-6: where to be conserved is not conservative. *Proc Natl Acad Sci U S A*; 94(6):2098-100.

- Hetzer-Egger C, Schorpp M, Boehm T (2000) Evolutionary conservation of gene structures of the Pax1/9 gene family. *Biochim Biophys Acta*;1492(2-3):517-521.
- Hoekstra HE, Price T (2004) Evolution. Parallel evolution is in the genes. *Science*; 303(5665):1779-81
- Horvath JE, Weisrock DW, Embry SL, Fiorentino I, Balhoff JP, Kappeler P, Wray GA, Willard HF, Yoder AD. (2008). Development and application of a phylogenomic toolkit: Resolving the evolutionary history of Madagascar's lemurs. *Genome Res.* 2008
- Horvath JE, Willard HF. (2007) Primate comparative genomics: lemur biology and evolution. *Trends Genet*;23(4):173-82
- Hunter JP, Jernvall J (1995) The hypocone as a key innovation in mammalian evolution. *Proc Natl Acad Sci U S A*; 92(23):10718-22.
- International Human Genome Consortium (2001) Initial sequencing and analysis of the human genome. *Nature*; 409: 860-921.
- Jernvall J, Keranen SV, Thesleff I (2000) Evolutionary modification of development in mammalian teeth: quantifying gene expression patterns and topography. *Proc Natl Acad Sci U S A*; 97(26):14444-8.
- Jernvall J (2000) Linking development with generation of novelty in mammalian teeth. *Proc Natl Acad Sci U S A*; 97(6):2641-5.
- Jobling M A, Hurles M and Tyler-Smith C. (2004) *Human Evolutionary Genetics – Origins, Peoples e Disease.* Garland Publishing.
- Jonker L, Kist R, Aw A, Wappler I, Peters H (2004) Pax9 is required for filiform papilla development and suppresses skin-specific differentiation of the mammalian tongue epithelium. *Mech Dev*; 121(11):1313-22.
- Jumlongras D, Lin JY, Chapra A, Seidman CE, Seidman JG, Maas RL, Olsen BR (2003) A novel missense mutation in the paired domain of PAX9 causes non-syndromic oligodontia. *Hum Genet*; 114(3):242-9.

- Kangas AT, Evans AR, Thesleff I, Jernvall J (2004) Nonindependence of mammalian dental characters. *Nature*; 432(7014):211-4.
- Kappen C (2000) Analysis of a complete homeobox gene repertoire: implications for the evolution of diversity. *Proc Natl Acad Sci U S A*; 97(9):4481-6.
- Karplus, K., Barrett, C. & Hughey, R (1998) Hidden Markov models for detecting remote protein homologies. *Bioinformatics* 14, 846-856.
- Kawasaki K, Suzuki T, Weiss KM (2005). Phenogenetic drift in evolution: the changing genetic basis of vertebrate teeth. *Proc Natl Acad Sci U S A*; 102(50):18063-8.
- Keranen SVE, Aberg T, Kettunen P, Thesleff I, Jernvall J (1999) Gene expression patterns associated with suppression of odontogenesis in mouse and vole diastema regions. *Dev Genet Evol*; 209: 495-506.
- King MC, Wilson AC (1975) Evolution at two levels in humans and chimpanzees. *Science*; 188: 107-116.
- Kirschner M, Gerhart J (1998) Evolvability. *Proc Natl Acad Sci U S A*; 95(15):8420-7.
- Kist R, Watson M, Wang X, Cairns P, Miles C, Reid DJ, Peters H (2005) Reduction of Pax9 gene dosage in an allelic series of mouse mutants causes hypodontia and oligodontia. *Hum Mol Genet*; 14(23): 3605-3617.
- Kozmik Z (2005). Pax genes in eye development and evolution. *Curr Opin Genet Dev*; 15(4):430-8.
- Kullberg M, Nilsson MA, Arnason U, Harley EH, Janke A (2006) Housekeeping genes for phylogenetic analysis of eutherian relationships. *Mol Biol Evol*; 23(8):1493-503.
- Kumar S, Hedges SB (1998) A molecular timescale for vertebrate evolution. *Nature*; 392: 917-920.
- Lancaster K., Single R. M., Solberg O. D., Nelson M. P. and Thomson g. (2007) "PyPop update - a software pipeline for large-scale multilocus population genomics" *Tissue Antigens* 69 (s1), 192-197.

- Levine M, Davidson EH (2005) Gene regulatory networks for development. *Proc Natl Acad Sci U S A*; 102(14):4936-42.
- Lewin R (1999) *Evolução Humana*. Atheneu, São Paulo.
- Line SR. (2001a) Molecular strategies in the evolution of mammalian dental patterning. *Evol Ecol*; 15: 73-79.
- Line SR (2001b) Molecular morphogenetic fields in the development of human dentition. *J Theor Biol*; 211:67-75.
- Line SR (2003) Variation of tooth number in mammalian dentition: connecting genetics, development, and evolution. *EVOLUTION & DEVELOPMENT* 5:3, 295-304.
- Majerus EN, Mundy NI (2003) Mammalian melanism: natural selection in black and white. *Trends in Genet*; 19: 585-588.
- Mansouri A, Hallonet M, Gruss P (1996) Pax genes and their roles in cell differentiation and development. *Curr Opin Cell Biol*; 8(6):851-7.
- Marroig G, Cheverud JM (2005) Size as a line of least evolutionary resistance: diet and adaptive morphological radiation in New World monkeys. *Evolution Int J Org Evolution*; 59(5):1128-42.
- Mattiangeli V, Ryan AW, McManus R, Bradley DG. (2007) A genome-wide approach to identify genetic loci with a signature of natural selection in the Irish population. *Genome Biol*;7(8):R74.
- McVean G. (2002) Natural selection. Student Folder. Department of Statistics, University of Oxford
- McKeown HF, Robinson DL, Elcock C, Al-Sharood M, Brook AH (2002) Tooth dimension in hypodontia patients, their unaffected relatives and a control group measured by a new image analysis system. *Eur J Orth*; 24: 131-141.
- McGuffin, L J, Bryson, K & Jones, D T (2000) The PSIPRED protein structure prediction server. *Bioinformatics* 16, 404-405.
- Melville J, Harmon LJ, Losos JB (2006) Intercontinental community convergence of ecology and morphology in desert lizards. *Proc Biol Sci*; 273(1586):557-563.

- Mendes-Junior CT, Castelli EC, Simões AL, Donadi EA (2007). Absence of the HLA-G*0105N allele in Amerindian populations from the Brazilian Amazon Region: a possible role of natural selection. *Tissue Antigens*;70(4):330-4.
- Miller DJ, Hayward DC, Reece-Hoyes JS, Scholten I, Catmull J, Gehring WJ, Callaerts P, Larsen JE, Ball EE (2000) Pax gene diversity in the basal cnidarian *Acropora millepora* (Cnidaria, Anthozoa): implications for the evolution of the Pax gene family. *Proc Natl Acad Sci U S A*; 97(9):4475-80.
- Monsoro- Burq A H(2005) Sclerotome development and morphogenesis: when experimental embryology meets genetics. *Int J Dev Biol*; 49(2-3):301-8.
- Mostowwska A, Biedziak B, Jagodzinski PP (2006) Axis inhibition protein 2 (AXIN2) polymorphisms may be a risk factor for selective tooth agenesis. *J Hum Genet*; 51: 262-266.
- Muller TS, Ebensperger C, Neubuser A, Koseki H, Balling R, Christ B, Wilting J (1996) Expression of avian Pax1 and Pax9 is intrinsically regulated in the pharyngeal endoderm, but depends on environmental influences in the paraxial mesoderm. *Dev Biol*; 178(2):403-17.
- Murphy WJ, Eizirik E, O'Brien SJ, Madsen O, Scally M, Douady CJ, Teeling E, Ryder OA, Stanhope MJ, de Jong WW, Springer MS. (2001) Resolution of the early placental mammal radiation using Bayesian phylogenetics. *Science*;294(5550):2348-51.
- Nei M (2007) The new mutation theory of phenotypic evolution. *Proc Natl Acad Sci U S A* 104(30):12235-12242.
- Nei M and Kumar S (2000) *Molecular Evolution and Phylogenetics*. 1^a ed. Oxford University Press.
- Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol*. 3(5):418-426.
- Neubuser A, Koseki H, Balling R (1995) Characterization and developmental expression of Pax9, a paired-box-containing gene related to Pax1. *Dev Biol*; 170(2):701-16.

- Nilsson DE (2004) Eye evolution: a question of genetic promiscuity. *Curr Opin Neurobiol*; 14(4):407-14
- Nüsslein-Volhard C(2006) *Coming to Life: How Genes Drive Development*. Kales Press, Carlsbad.
- Ogasawara M, Wada H, Peters H, Satoh N (1999) Development expression of Pax1/9 genes in urochordate and hemichordate gills: insight into function and evolution of the pharyngeal epithelium. *Development*; 126: 2539-2550.
- Ogasawara M, Shigetani Y, Hirano S, Satoh N, Kuratani S (2000) Pax1/PAX9-Related genes in an agnathan vertebrate, *Lampetra japonica*: expression pattern of LjPAX9 implies sequential evolutionary events toward the gnathostome body plan. *Dev Biol*; 223: 399-410.
- Ogawa T, Kapadia H, Feng JQ, Raghov R, Peters H, D'Souza RN (2006) Functional consequences of interactions between Pax9 and Msx1 genes in normal and abnormal tooth development. *J Biol Chem*; 281(27):18363-9.
- Orr HA. (2005)The probability of parallel evolution. *Evolution Int J Org Evolution*;59(1):216-20.
- Palumbi SR (1996) *Nucleic acids II: the polymerase chain reaction*. Em: *Molecular Systematics*. Sinauer Associates, Inc. Publishers, Massachusetts.
- Patterson C (1988) Homology in classical and molecular biology. *Mol Biol Evol*; 5: 603-625.
- Peichel CL, Nereng KS, Ohgi KA, Cole BL, Colosimo PF, Buerkle CA, Schluter D, Kingsley DM (2001) The genetic architecture of divergence between threespine stickleback species. *Nature*; 414(6866):901-5.
- Pena SDJ. (2006). *Acidentes acontecem*. *Ciência Hoje On-Line* (<http://cienciahoje.uol.com.br/45989>).
- Pereira T V, Salzano F M, Mostowska A, Trzeciak W H, Ruiz-Linares A, Chies J A B, Saavedra C, Nagamachi, Hurtado A M, Hill K, Castro-de-Guerra D, Silva-Junior W A, and Bortolini M C (2006) Natural selection and molecular evolution in primate PAX9 gene, a

- major determinant of tooth development. *Proc Natl Acad Sci USA*; 15:5676-5681.
- Peres RC, Scarel-Caminaga RM, do Espirito Santo AR, Line SR (2005) Association between PAX-9 promoter polymorphisms and hypodontia in humans. *Arch Oral Biol*; 50(10):861-71.
- Perry GH, Verrelli BC, Stone AC (2006) Molecular evolution of the primate developmental genes MSX1 and PAX9. *Mol Biol Evol*; 23: 644-654.
- Peters H, Balling R (1999) Teeth. Where and how to make them. *Tooth Devel*; 15: 59-65.
- Peters H, Wilm B, Sakai N, Imai K, Maas R, Balling R (1999) Pax1 and PAX9 synergistically regulate vertebral column development. *Development*; 126: 5399-5408.
- Peters H, Neubuser A, Balling R (1998) Pax genes and organogenesis: PAX9 meets tooth development. *Eur J Oral Sci*; 106 (Suppl 1): 38-43.
- Peters H, Schuster G, Neubuser A, Richter T, Hofler H, Balling R. (1997) Isolation of the Pax9 cDNA from adult human esophagus. *Mamm Genome*;8(1):62-4.
- Pigliucci M. (2008) Is evolvability evolvable? *Nat Rev Genet*;9(1):75-82.
- Pough J H, Janis, C M, Heiser J B (2003) A vida dos vertebrados. 6ª ed. São Paulo, Atheneu.
- Protas ME, Hersey C, Kochanek D, Zhou Y, Wilkens H, Jeffery WR, Zon LI, Borowsky R, Tabin CJ (2006) Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nat Genet*; 38: 107-111.
- Prud'homme B, Gompel N, Rokas A, Kassner VA, Williams TM, Yeh SD, True JR, Carroll SB (2006) Repeated morphological evolution through cis-regulatory changes in a pleiotropic gene. *Nature*; 440(7087):1050-3.
- Ramalho R F (2007) Análise do polimorfismo da região cis-reguladora do gene CCR5 em populações ameríndias. Dissertação (Mestrado)

- Instituto de Biociências da Universidade de São Paulo.
Departamento de Genética.
- Redondo RH, Santos FR (2006) Evolutionary studies on the amilase gene segment in bats and other mammals. *Genetica*; 126: 199-213.
- Richardson MK, Brakefield PM (2003) Developmental biology: hotspots for evolution. *Nature*; 424(6951):894-5.
- Robson EJ, He SJ, Eccles MR (2006) A PANorama of PAX genes in cancer and development. *Nat Rev Cancer*; 6(1):52-62.
- Rockman MV, Hahn MW, Soranzo N, Zimprich F, Goldstein DB, Wray GA (2005) Ancient and recent positive selection transformed opioid cis-regulation in humans. *PLoS Biol*; 3(12):e387.
- Rockman MV, Wray GA (2002) Abundant raw material for cis-regulatory evolution in humans. *Mol Biol Evol*; 19(11):1991-2004.
- Rodrigo I, Hill RE, Balling R, Munsterberg A, Imai K (2003) Pax1 and Pax9 activate Bapx1 to induce chondrogenic differentiation in the sclerotome. *Development*; 130(3):473-82.
- Rossi DL, Acebron A, Santisteban P (1995) Function of the homeo and paired domain proteins TTF-1 and Pax-8 in thyroid cell proliferation. *J Biol Chem*; 270:23139-23142 .
- Rozkovcová E, Marková M, Dolejší J (1999) Studies on agenesis of third molars amongst populations of different origin. *Sborník*; 100(2): 71-84.
- Salazar-Ciudad I, Jernvall J (2004) How different types of pattern formation mechanisms affect the evolution of form and development. *Evol Dev*; 6(1):6-16.
- Salazar-Ciudad I, Jernvall J, Newman SA (2003) Mechanisms of pattern formation in development and evolution. *Development*; 130(10):2027-37.
- Salazar-Ciudad I, Jernvall J (2002) A gene network model accounting for development and evolution of mammalian teeth. *Proc Natl Acad Sci U S A*; 99(12):8116-20.
- Santagati F, Abe K, Schmidt V, Schmitt-John T, Suzuki M, Yamamura K, Imai K (2003) Identification of Cis-regulatory elements in the

- mouse Pax9/Nkx2-9 genomic region: implication for evolutionary conserved syntenies. *Genetics*; 165(1):235-42.
- Santagati F, Gerber JK, Blusch JH, Kokubu C, Peters H, Adamski J, Werner T, Balling R, Imai K (2001) Comparative analysis of the genomic organization of Pax9 and its conserved physical association with Nkx2-9 in the human, mouse, and pufferfish genomes. *Mamm Genome*; 12(3):232-7.
- Santos FR, Guimarães PEM, Redondo RH (2002) Bancos de DNA: coleções estratégicas para estudos da biodiversidade. *Lundiana*; 3: 93-98.
- Semple CA, Maxwell A, Gautier P, Kilanowski FM, Eastwood H, Barran PE, Dorin JR. The complexity of selection at the major primate beta-defensin locus. *BMC Evol Biol*. 2005 May 18;5(1):32.
- Shapiro MD, Marks ME, Peichel CL, Blackman BK, Nereng KS, Jónsson B, Schluter D, Kingsley DM (2004) Genetic developmental basis of evolutionary pelvic reduction in threespine sticklebacks. *Nature*; 428: 717-720.
- Schneider I, Schneider H, Schneider M P, Silva A (2004) The prion protein and New World primate phylogeny. *Genetics and Molecular Biology*, 27, 4, 505-510.
- Slatkin M. (1994). "An exact test for neutrality based on the Ewens sampling distribution". *Genetical Research*. 64. 71-74.
- Stapleton P, Weith A, Urbanek P, Kozmik Z, Busslinger M (1993) Chromosomal localization of seven PAX genes and cloning of a novel family member, PAX9. *Nat Genet*; 3:292-298.
- Stockton DW, Das P, Goldenberg M, D'Souza RN, Patel PI (2000) Mutation of PAX9 is associated with oligodontia. *Nat Genet*; 24: 18-19.
- Sucena E, Delon I, Jones I, Payre F, Stern DL (2003) Regulatory evolution of shavenbaby/ovo underlies multiple cases of morphological parallelism. *Nature*; 424(6951): 935-938.
- Tavaré S, Marshall CR, Will O, Soligos C, Martin RD (2002) Using the fossil record to estimate the age of the last common ancestor of extant primates. *Nature*; 416: 726-729.

- The Chimpanzee Sequencing and Analysis Consortium (2005) Initial sequence of the chimpanzee genome and comparison with human genome. *Nature*; 437: 69-87.
- Thesleff I (2000) Genetic basis of tooth development and dental defects. *Acta Odontol Scand*; 58: 191-194.
- Thornton JW, DeSalle R (2000) Gene family evolution and homology: genomics meets phylogenetics. *Annu Rev Genomics Hum Genet*;1:41-73.
- Varki A, Altheide TK (2005) Comparing the human and chimpanzee genomes: searching for needles in a haystack. *Genome Research*; 15: 1746-1758.
- Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, et al. (2001) The sequence of the human genome. *Science*; 291: 1304-1351.
- Vogel C. and Chothia C. (2006) Protein family expansions and biological complexity. *PLoS Comput Biol*;2(5):e48.
- Wagner GP, Lynch VJ (2005) Molecular evolution of evolutionary novelties: the vagina and uterus of therian mammals. *J Exp Zool B Mol Dev Evol*; 304(6):580-92.
- Wang Y, Gu X (2001) Functional Divergence in the Caspase Gene Family and Altered Functional Constraints: Statistical Analysis and Prediction. *Genetics*; 138: 1311-1320.
- Weiss KM (2005) The phenogenetic logic of life. *Nat Rev Genet* ;6(1):36-45.
- West-Eberhard MJ (2005) Developmental plasticity and the origin of species differences. *Proc Natl Acad Sci U S A*; 102 Suppl 1:6543-9.

- West-Eberhard MJ (1998) Evolution in the light of developmental and cell biology, and vice versa. *Proc Natl Acad Sci U S A*; 95(15):8417-9.
- West-Eberhard MJ (1986) Alternative adaptations, speciation, and phylogeny (A Review). *Proc Natl Acad Sci U S A*; 83(5):1388-1392.
- Wiens JJ, Chippindale PT, Hillis DM (2003) When are phylogenetic analyses misled by convergence? A case study in Texas cave salamanders. *Syst Biol*; 52: 501-514.
- Wilm B, Dahl E, Peters H, Balling R, Imai K (1998) Targeted disruption of Pax1 defines its null phenotype and proves haploinsufficiency. *Proc Natl Acad Sci USA*; 95: 8692-7869.
- Wimmer R, Kirsch S, Rappold GA, Schempp W (2002) Direct evidence for the Homo-Pan clade. *Chromosome Res*; 10 :55-61.
- White, J. V., Stultz, C. M. & Smith, T. F. (1994) *Math. Biosci.* 119, 35-75.
- Wood TE, Burke JM, Rieseberg LH (2005) Parallel genotyping adaptation: when evolution repeats itself. *Genetica*; 123: 157-170.
- Wray GA (2006). Evolution: spot on (and off). *Nature*; 440(7087):1001-2.
- Wray GA, Kitazawa C, Miner B (2004) Culture of echinoderm larvae through metamorphosis. *Methods Cell Biol*; 74:75-86.
- Wray GA (2003) Transcriptional regulation and the evolution of development. *Int J Dev Biol*; 47(7-8):675-84.
- Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, Smith SF, North P, Callaway H, Kelly K, Walter K, Abnizova I, Gilks W, Edwards YJ, Cooke JE, Elgar G. (2005) Highly conserved non-coding sequences are associated with vertebrate development. *PLoS Biol.* 3(1):e7.
- Zhang J, Kumar S (1997) Detection of convergent and parallel evolution at the amino acid sequences level. *Mol Biol Evol*; 14; 527-536.

- Yang F, Graphodatsky AS, Li T, Fu B, Dobigny G, Wang J, Perelman PL, Serdukova NA, Su W, O'Brien PC, Wang Y, Ferguson-Smith MA, Volobouev V, Nie W (2006) Comparative genome maps of the pangolin, hedgehog, sloth, anteater and human revealed by cross-species chromosome painting: further insight into the ancestral karyotype and genome evolution of eutherian mammals. *Chromosome Res*; 14:283-296.
- Yang J, Wang ZL, Zhao XQ, Wang de P, Qi de L, Xu BH, Ren YH, Tian HF (2008) Natural selection and adaptive evolution of leptin in the ochotona family driven by the cold environmental stress. *PLoS ONE*. 3(1):e1472.
- Yang Z. (2007) PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*;24(8):1586-91.
- Yang Z. (2006) *Computacional Molecular Evolution*. 1ª ed. Oxford Univerity Press.
- Yang Z, Nielsen R. (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol*;19(6):908-17.
- Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics*. 155(1):431-449.

APÊNDICE I

APÊNDICE I

AI-1 Amostras

Como já foi salientado, visando atingir os objetivos 2 e 3 dessa dissertação, foram investigadas 54 amostras, representando 25 espécies de mamíferos (Tabela 3). Algumas destas foram cedidas pelo banco de DNA da Universidade Federal de Minas Gerais, que conta com uma coleção de amostras de mais de 400 espécies de vertebrados, principalmente aves e mamíferos, sob a responsabilidade do prof. Fabrício Santos.

As amostras de cetáceos foram disponibilizadas pelo GEMARS, Grupo de Estudo de Mamíferos Marinhos do Rio Grande do Sul, que possui uma vasta coleção de amostras de tecido e DNA de mamíferos marinhos.

Já as amostras de *Ctenomys* foram cedidas pelo Projeto Tuco-Tuco, sob a responsabilidade do prof. Thales de Freitas (Dep. Genética-UFRGS).

A maioria das amostras de primatas do Novo Mundo, foram cedidas pelo Centro de Primatologia do Rio de Janeiro –CPRJ/FEEMA, cujo curador e veterinário responsável é o Dr. Alcides Pissinatti que

conta com um plantel de 158 indivíduos de diversos gêneros, através de um sistema de colaboração firmado em agosto de 2007.

Ainda contamos com a colaboração do Dr. Fengtang Yang (*The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Inglaterra*) que dispõe de amostras da espécie *Manis javanica* (pangolin de Java).

O banco de dados foi ampliado através da busca de seqüências do gene *PAX9* de outros mamíferos depositadas no Genbank/NCBI (<http://www.ncbi.nlm.nih.gov/Genbank>) e no Ensembl (<http://www.ensembl.org/index.html>). A partir da seqüência de referência humana do *PAX9* foram realizadas buscas utilizando-se o sistema *Nucleotide-nucleotide* BLAST. Posteriormente foram escolhidos os resultados identificados ou previstos como sendo correspondentes ao gene *PAX9*, e ainda, seqüências que possuïrem uma identidade de >70%. Seqüências de 28 espécies de mamíferos foram selecionadas (Tabela 4).

AI-2 Reação de amplificação

Primers universais (Tabela 5) foram desenhados a partir dos dados publicados para as 28 espécies de mamíferos, além dos humanos, utilizando-se o programa FastPCR (<http://www.biocenter.helsinki.fi/bi/Programs/fastpcr.htm>). Como critério, foi utilizado a busca por regiões que fossem

pelo menos 98% idênticas entre todas as espécies. Ajustes nas condições de amplificação.

AI -TABELA 1. Relação dos *primers* utilizados para a amplificação;

Primers	Seqüências	Tamanho esperado em pb	Design
Exon 2V	Forward 5' - CCAGCCTTCGGGGAGGTGAA -3' Reverse 5' -GACGCTGCACATCCACACG -3'	~640	<i>FastPCR</i>
Exon 3V	Forward 5' - GTGAGCGACAGCTCCCCCTA -3' Reverse 5' - CTGACCGTACTTGGCTTCCT -3'	~140	<i>FastPCR</i>
Exon 4V	Forward 5' - AGGCACCAAATGGTCTCCCAGCTGT -3' Reverse 5' - GAAGCCGTGACAGAATGACTA -3'	~247	<i>FastPCR</i>

A amplificação dos exons 2, 3 e 4, foi feita em um volume total de 25 µl por reação, contendo 0,2 µM de cada nucleotídeo, 50 pmol de cada *primer*, 1.5 mM de MgCl₂, 2,5 mM buffer, 2 U de *Taq Platinum DNA polymerase* e 100 ng de DNA genômico. O *PCR* foi realizado nas seguintes condições: 35 ciclos de desnaturação a 94 °C, anelamento a 53 °C – 59 °C e a extensão a 72 °C, a duração de cada passo dos ciclos será de 50 segundos, com exceção do primeiro passo de desnaturação com 2 minutos, e do última etapa de extensão de 10 minutos. Ajustes

nas condições de amplificação, particularmente relacionadas a temperatura de anelamento, foram realizados sempre que necessários.

AI-3 Seqüenciamento

As purificações dos produtos amplificados foram feitas com as enzimas Exonuclease I e *Shrimp Alkaline Phosphatase* comercializadas pela *GE Healthcare* seguindo o protocolo recomendado pelo fabricante.

Seqüenciamento de ambas as fitas do DNA foi realizado através dos equipamentos MegaBACE 500 e MegaBACE 1000 de acordo com condições descritas pelos fabricantes.

AI-4 Análise de dados

Para testar se a variação encontrada nas regiões codificadas do *PAX9* nas 53 espécies poderia ser explicada pelo equilíbrio entre mutação e deriva (neutralidade) foi realizado testes usando os programas *YN00* e *CODEML*, parte do pacote PAML 3.15 (Anisimova e cols., 2007; Yang e cols., 2000, 2002, 2007). Os mesmos estimam a taxa de substituições sinônimas e não-sinônimas (dN/dS), também conhecida como ω , através dos métodos de Nei e Gojobori (1986), e de Yang e Nielsen (2000).

Valores das taxas de substituição $\omega < 1$ e $\omega > 1$ indicam respectivamente seleção purificadora (ou negativa), ou positiva (Darwiniana). Já valores de $\omega = 1$ estariam indicando que a hipótese de neutralidade não poderia ser descartada.

A implementação dos modelos de substituição de códons foi feita seguindo para seguintes os modelos: M0 (*one ratio*), onde um único valor de ω é assumido em todos os códons, além de pressupor a mesma e constante pressão seletiva em todos os sites através do tempo; e M3 (*discrete*) que admite valores diferentes de ω , e também variação entre diferentes ramos da filogenia; já os modelos M1a (*neutral*) admite diferentes valores de ω , mas nunca maiores do que um, e M2a (*selection*) admite valores de $\omega > 1$; M7 (*beta*) também admite diferentes valores de ω , mas nunca maiores do que um, mas diferentemente de M1a, considera 11 parâmetros; e finalmente M8 (*beta & ω*) que permite valores de $\omega > 1$. Foram feitas comparações por análises de Máxima Verossimilhança, entre os modelos M0 e M3, M1a e M2a, e M7 e M8 para determinar qual melhor modelo de substituição de códons, adequava-se aos dados. Abordagem bayesiana foi aplicada para estimar a significância entre as diferenças de dN/dS (Anisimova e cols., 2007; Semple e cols., 2005; Yang, 2006).

A comparação da conservação das seqüências de aminoácidos foi feita com o programa Jalview (Clamp *et al*, 2004). Já as predições das

estruturas protéicas secundárias foram realizadas utilizando-se *Jnet Secondary Structure Prediction*, componente do Jalview, e do programa PSIFRED (McGuffin e cols., 2000). Estes programas indicam a probabilidade da existência (ou não) de certos motivos moleculares, que à luz do conhecimento sobre suas propriedades químicas, podem ser associados a eventuais perdas ou ganhos funcionais. Abordagem similar foi feita no trabalho de Pereira e cols. (2006).