

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL

INSTITUTO DE BIOCÊNCIAS

DEPARTAMENTO DE GENÉTICA

PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA E BIOLOGIA MOLECULAR

***"A Busca por Padrões Evolutivos em Genes
do Desenvolvimento"***

Vanessa Rodrigues Paixão-Côrtes

Tese submetida ao Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS como requisito parcial para a obtenção do grau de Doutor em Ciências (Genética e Biologia Molecular).

Orientadora: Maria Catira Bortolini

Porto Alegre, maio 2012.

Este trabalho foi realizado nas instalações do Laboratório de Evolução Humana e Molecular do Departamento de Genética do Instituto de Biociências da Universidade Federal do Rio Grande do Sul, com financiamento do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) e da Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul (FAPERGS).

“You OK with an answer based on evolution?”

Dr. Gregory House

AGRADECIMENTOS

- À Dra. **Maria Cátira Bortolini** pela orientação, confiança, incentivo e amizade. Por estar sempre disposta a ampliar as linhas de pesquisas, e integrar novas ideias aos objetivos do grupo;
- Ao Dr. **Francisco Mauro Salzano**, pelo fato de estar sempre presente e disposto a ajudar. Corrigindo resumos, manuscritos, verificando referência por referência, sempre com total dedicação. Um exemplo como cientista e pessoa;
- À **Laci**, por ser uma pessoa excepcional que, junto com a Professora Cátira, me buscava e levava em casa quando eu tive com problemas de locomoção;
- A **Tatiana Braga, Karina Mundstock, Carlos Alberto Mundstock** da Faculdade de Odontologia da UFRGS, pela colaboração, e por serem responsáveis pelas coletas de informações e amostras;
- Aos professores **Diogo Meyer, Tiago V. Pereira, Stéphane Mazières, Jacques Elion, Rajagopal Krishnamoorthy, Marco A. Zago⁷ e Wilson A. Silva Jr** pela colaboração, nas análises e pelas amostras dos Africanos do Sul do Saara;
- Aos Professores Dr **Claiton Bau** e Dra **Patricia Ashton-Prolla** e aos colegas **Nina Mota** e **Diego Paskulin** e por me convidarem para colaborar em seus estudos.
- À Dra. **Vera Gayesky**, ao Dr. **Aldo Mellender**, à Dra. **Sídia Callegari-Jacques**, à Dra **Lavínia Schüler-Faccini** e a todos os **professores** do Programa de Pós-Graduação em Genética e Biologia Molecular da UFRGS, que contribuíram muito para o meu aperfeiçoamento;
- À Dra. **Eliane Bandinelli**, Dra **Loreta Freitas** pelas conversas e companhia e pelo café;
- Ao **Elmo** e a **Lúcia** por ajudarem sempre;
- À Dra **Tábita** Hunemeier pela amizade. Fomos colegas de laboratório, e atualmente sou parte da sua rede de colaboradores. Estamos sempre discutindo ideias, programando o próximo paper, ou simplesmente fofocando;
- À **Luciana** Tovo Rodrigues, que não está aqui; está em Baltimore... Por ser minha amiga, companheira de almoços, e estar sempre disposta a me ajudar no trabalho e na vida pessoal;
- À Dra **Gislene** Gonçalves, sempre a disposição para conversar ou discutir um método;
- Aos companheiros de Sala e Laboratório: **Rafael**, são mais de 6 anos de convivência, com um colega dedicado e atencioso; ao **Caio**, futuro perito e

colaborador, sempre disposto a ajudar; ao **Eduardo**, pela companhia divertidíssima e a disposição de explicar os programas de coalescências; ao **Lucas** pela empolgação e comprometimento; à **Virgínia**; à **Luana**; ao **Carlos**; ao **Ely**; ao **Pedro**; ao **Álvaro** e todas as **meninas** novas que estão entrando agora no Lab.

- Aos meus **colegas** do PPGBM;
- Aos amigos: à **Karen** e às **Gêmeas**; ao **Dennis**; à **Claudia**; à **Vanessa** e todos que participam da minha vida;
- A toda a minha família e em especial: à **Vó**, a pessoa que me ensinou sobre bondade e a importância de ajudar o próximo; à **Aline**, minha super irmã caçula; ao **Cleber** e ao **Luan**, meu cunhado e sobrinho que chegou este ano para alegrar nossas vidas; à **Carla**, minha prima e amiga;
- A família que agora é minha também: à **Débora**, minha doce cunhada; à **Pietra**, minha sapeca sobrinha; ao **Bolívar**, meu sogro e agora, nestes últimos dias da tese, baba de plantão;
- Ao **Walter**, meu amor, meu amigo, o informata que sempre me ajuda quando estou com problemas, e que me apresenta soluções automatizadas; à **Lilian**, minha paixão e sucesso adaptativo: obrigada por completarem a minha vida;
- Em especial a minha **mãe**, a **“vó Cila”**, por deixar a vida dela em Florianópolis, e me ajudar com a Lilian. Por este motivo, eu consegui fazer este doutorado com dedicação e tranquilidade;
- As pessoas que foram muito importantes na minha vida, e que me ensinaram importantes lições, mas que agora já se foram: **Pai** (1986); **Rosinha** (1998); **Junior** (irmão -1998); **Vô** (2007); **Maria Luiza** (05/05/2012). Que a lembrança de todos permaneça viva em nossos corações.

SUMÁRIO

LISTA DE ABREVIATURAS	8
RESUMO.....	10
ABSTRACT	11
CAPÍTULO 1 - INTRODUÇÃO	12
1.1 - CONSIDERAÇÕES INICIAIS.....	12
1.2 - EVOLUÇÃO BIOLÓGICA	13
1.3 - PROCESSOS MACRO E MICROEVOLUTIVOS.....	15
1.4 - GENES MESTRES DO DESENVOLVIMENTO	16
1.5 A FAMÍLIA PAX	19
1.5.1 - PAX1/PAX9	23
1.5.2 - PAX2/PAX5/PAX8	27
1.5.3 – PAX3/PAX7	29
1.5.3 – PAX4/PAX6	31
1.6 – OS GENES MSX E A SUA INTERAÇÃO COM A FAMÍLIA PAX.....	34
CAPÍTULO 2 - OBJETIVOS	37
2.1 – OBJETIVO GERAL	37
2.2 – OBJETIVOS ESPECÍFICOS.....	37
CAPÍTULO 3 – RESULTADOS: “PAX9 AND MSX1 TRANSCRIPTION FACTOR GENES IN NON-SYNDROMIC DENTAL AGENESIS”	40
CAPÍTULO 4 – RESULTADOS - “GENETIC VARIATION AMONG MAJOR HUMAN GEOGRAPHIC GROUPS SUPPORTS A PECULIAR EVOLUTIONARY TREND IN PAX9”	62
CAPÍTULO 5 - RESULTADOS– “EVOLUTIONARY HISTORY OF VERTEBRATE PAX GENES –DYNAMICS OF CHANGE IN A COMPLEX GENE FAMILY”	81
CAPÍTULO 6 – CONCLUSÕES	113
REFERÊNCIAS BIBIOGRAFICAS	120

APÊNDICE 1 – “HOMO SAPIENS, HOMO NEANDERTHALENSIS AND THE DENISOVA SPECIMEN: NEW INSIGHT ON THEIR EVOLUTIONARY HISTORIES USING WHOLE-GENOME COMPARISONS”	127
APÊNDICE 2 – “PREDICTING HOMO PIGMENTATION PHENOTYPE THROUGH GENOMIC DATA: FROM NEANDERTHAL TO JAMES WATSON”	154
APÊNDICE 3 – “LINKING DOPAMINE NEUROTRANSMISSION AND NEUROGENESIS: THE EVOLUTIONARY HISTORY OF THE NTAD (NCAM1-TTC12-ANKK1-DRD2) GENE CLUSTER”	160
APÊNDICE 4 – “EVOLUTIONARY RESPONSES TO A CONSTRUCTED NICHE: ANCIENT MESOAMERICANS AS A MODEL OF GENE-CULTURE COEVOLUTION”	184
APÊNDICE 5 – “POPULATION DATA SUPPORT THE ADAPTIVE NATURE OF HACNS1 SAPIENS/NEANDERTAL-CHIMPANZEE DIFFERENCES IN A LIMB EXPRESSION DOMAIN”	220
APÊNDICE 6 – “THE TP53 FERTILITY NETWORK”	225
APÊNDICE 7 – “DISTRIBUTION OF Y-CHROMOSOME Q LINEAGES IN NATIVE AMERICANS”	248
APÊNDICE 8 - “A NEW SUBHAPLOGROUP OF NATIVE AMERICAN Y-CHROMOSOMES FROM THE ANDES”	253

LISTA DE ABREVIATURAS

A	=	Adenosina
ABCA1	=	ATP-binding cassette sub-family A member 1
Ala	=	Alanina
Arg	=	Arginina
C	=	Citosina
Cys	=	Cistidina
dN/dS ou ω	=	São as substituições não sinônimas divididas pelas substituições sinônimas por sítio nucleotídeo ou ômega
F_{ST}	=	Medida da diversidade de alelos escolhidos aleatoriamente dentro de uma subpopulação mesmo em relação ao que foi encontrado na população inteira, Índice de fixação ou estatística de fixação ou estatística F
G	=	Guanina
HACNS1	=	região localizada no gene CENTG2, classificada como HAR (Human Accelerated Region 2), atua como "enhancer" de vários genes inclusive o PAX9
HD	=	Homeodomínio
His	=	Histidina
HOX	=	Homeobox genes
HOX7	=	Antiga denominação do grupo de genes homeóticos <i>MSX</i>
HOX8	=	Antiga denominação do grupo de genes homeóticos <i>MSX</i>
Msh	=	muscle segment homeobox encontrado em <i>Drosophila</i>
Msh	=	<i>muscle segment homeobox</i>
Msx	=	Nomenclatura para muscle segment homeobox em ratos.
MSX	=	Nomenclatura para muscle segment homeobox em humanos.
MSX	=	<i>msh homeobox</i>
MSX	=	homeobox protein MSX
MSX1	=	<i>msh homeobox 1</i> ou <i>msh (Drosophila) homeo box homolog 1</i>
MSX1	=	homeobox protein MSX-1
Msx1/Msx2	=	<i>duplo nocaute para ambos genes em camundongos</i>
MSX2	=	<i>gene msh homeobox 2</i> ou <i>msh (Drosophila) homeo box homolog 2</i>
MSX2	=	homeobox protein MSX-2
MSX2P1	=	<i>msh homeobox 2 pseudogene</i> ou <i>MSX3</i>
MSX3	=	<i>MSX1 (Msh homeobox 1-like (HOX7), MSX2 (Msh homeobox 2-like (HOX8), Msx3 (Msh homeobox 3-like- em camundongos e MSX2P1- msh homeobox 2 pseudogene , em humanos).</i>
OC	=	Octapeptídeo
Paired	=	O domínio <i>paired</i> é composto por dois subdomínios hélice-gira-hélice (<i>helix-turn-helix</i>): o subdomínio N-terminal, <i>PAI</i> ; e o subdomínio C-terminal chamado RED. Ambos ligam-se ao DNA de forma independente.
PAX	=	Paired homeobox genes
PAX1	=	<i>paired box gene 1</i>
PAX1	=	<i>paired box protein Pax-1</i>
PAX2	=	<i>paired box gene 2</i>
PAX2	=	<i>paired box protein Pax-2</i>
PAX3	=	<i>paired box gene 3</i>

PAX3	=	<i>paired box protein Pax-3</i>
PAX4	=	<i>paired box gene 4</i>
PAX4	=	<i>paired box protein Pax-4</i>
PAX5	=	<i>paired box gene 5</i>
PAX5	=	<i>paired box protein Pax-5</i>
PAX6	=	<i>paired box gene 6</i>
PAX6	=	<i>paired box protein Pax-6</i>
PAX7	=	<i>paired box gene 7</i>
PAX7	=	<i>paired box protein Pax-7</i>
PAX8	=	<i>paired box gene 8</i>
PAX8	=	<i>paired box protein Pax-8</i>
PAX9	=	<i>paired box gene 9</i>
PAX9	=	<i>paired box protein Pax-9</i>
Pb	=	Pares de bases
PCR	=	Reação da polimerase em cadeia
PD	=	<i>Domínio paired</i>
PHT	=	homeodomínio <i>Paired-type tail</i>
Pox meso	=	<i>Paired box mesodermal</i> , gene PAX encontrado na <i>Drosophila</i>
Pro	=	Prolina
SME	=	sistema muscular esquelético
SNC	=	Sistema nervoso central
SNP	=	<i>gene msh homeobox 1</i> ou <i>msh (Drosophila) homeo box homolog 1</i>
TP53	=	<i>tumor protein p53</i>

Obs: para facilitar o entendimento sempre que se estiver referindo-se aos genes parálogos PAX a grafia irá obedecerá as recomendações do HUGO Gene Nomenclature Committee (HGNC) - <http://www.genenames.org/>; Demais informações para a nomenclatura foram consultadas nos bancos de dados: Uniprot - <http://www.uniprot.org/>; The Protein Naming Utility - http://www.jcvi.org/pn-utility/web/smarty_wrapper/smarty_import.php

RESUMO

Um dos maiores objetivos da biologia atualmente é poder relacionar alterações nos genes a mudanças fenotípicas em organismos complexos. O desafio é poder interpretar estas mudanças em vários níveis: (1) no organismo; (2) entre/dentro de populações; e (3) entre taxas. Os genes mestres do desenvolvimento seriam bons candidatos para o estudo do surgimento de novidades adaptativas. Então, foram escolhidos os genes da família *PAX* (genes *Paired box*), porque estes codificam fatores de transcrição que têm um papel central na determinação celular, atuando assim na morfogênese de diversos tecidos (sistema nervoso central, epiderme, músculos, ossos, células do sistema imune). Os enfoques desenvolvidos no presente estudo podem ser assim resumidos: (1) Para investigar o papel de um membro desta família, o gene *PAX9* que juntamente com o gene *MSX1* é um dos genes chave da cascata dentária, bem como do fenótipo de agenesia de terceiros molares, foi realizado um estudo piloto com 35 pacientes, portadores de agenesias dentárias não sindrômicas. Neste estudo foi possível observar que a mutação *240Pro/PAX9*, estaria causando a agenesia dos terceiros molares com um padrão de herança autossômico recessivo e expressividade variada. (2) Com o padrão de herança estabelecido pelo estudo piloto, verificou-se a distribuição do alelo conjuntamente com a distribuição da agenesia dos terceiros molares nos grandes grupos geográficos humanos. Para tanto, foi investigada a diversidade do éxon 3 e regiões adjacentes do gene *PAX9* em 125 indivíduos nativos americanos e 77 africanos do sul do Saara; estes dados foram integrados com a informação disponível para a mesma região genética de 115 indivíduos de diferentes origens geográficas. Adicionalmente, dados coletados do Hapmap de 6001 SNPs (do *PAX9* e de mais 101 genes envolvidos no desenvolvimento craniofacial) foram analisados quanto à sua distribuição em africanos, asiáticos e europeus. A distribuição do alelo *240Pro/PAX9*, e conseqüentemente do genótipo homocigoto *240Pro/Pro* poderia explicar toda a incidência de agenesia de terceiros molares na África (~2%). Para as outras populações estudadas outros genes da cascata dentária devem estar também influenciando na determinação do fenótipo de agenesia. Confirmando esta sugestão uma proporção significativamente alta (1,3%) de valores de F_{ST} extremos ($\geq 0,60$) é encontrada quando comparamos os SNPs dos genes do desenvolvimento craniofacial entre europeus e asiáticos com os africanos. O cenário proposto é que uma forte seleção purificadora estaria agindo sobre o fenótipo craniofacial na África; mas com a saída deste continente, poderiam estar atuando um relaxamento funcional ou mesmo uma possível seleção positiva nas populações da Europa e da Ásia. Testes de neutralidade e coalescência indicam que nos ameríndios fatores randômicos parecem ser as principais forças atuando na história evolutiva destas populações. (3) Para entender como é a dinâmica evolutiva da família *PAX* entre os vertebrados, os nove membros desta família foram procurados nos genomas de 53 cordados. Há evidência que a principal força atuando nesta família é a seleção purificadora, mas em alguns membros parece existir um relaxamento da pressão funcional (*PAX4*, *PAX7* e *PAX8*). O principal achado deste último estudo é que na linhagem Sauropsida (répteis e aves) houve a perda de dois genes *PAX* (*PAX4* e *PAX8*). Tanto o relaxamento funcional quanto a perda de genes poderia ser explicada pela sobreposição da expressão e redundância funcional de genes parálogos dentro da família *PAX*. Desta forma, a conservação ou presença de um gene contrabalançaria variação ou ausência de outro. O gene *PAX9* e a família *PAX* seriam um exemplo notável de como se processa a evolvibilidade.

ABSTRACT

One of the main goals in biology today is to establish the relationship between gene changes and phenotypic variations in multicellular organisms. The challenge is to understand these changes on several levels: (1) in the organism; (2) between/within populations; and (3) between taxa. Development master genes would be good candidates for studying the emergence of adaptive novelties. Therefore the *PAX* (Paired box) gene family was chosen, since it encodes transcription factors that act as genetic switches for cell lineage decisions, thus acting in the morphogenesis of various tissues (central nervous system, skin, muscle, bone, immune system cells). The approaches developed in the present study can be summarized as follows: (1) To investigate the role of a *PAX* family member (*PAX9* and *MSX1* are key factors in the development of animal dentition) and the third molars agenesis phenotype a pilot study was conducted with 35 patients suffering from non-syndromic teeth agenesis. Our results suggest that the *240Pro/PAX9* mutation was causing third molars agenesis with an autosomal recessive pattern of inheritance and variable expressivity. (2) After setting the inheritance pattern by this pilot study, we verified this allele distribution in conjunction with the distribution of third molars agenesis in the major human geographic groups. With this objective, a total of 125 Amerindians and 77 individuals from southern Sahara Africa were studied by sequencing the *PAX9* gene exon 3 and adjacent regions and the data integrated with the information available for the same genetic region from 115 individuals of different geographical origins. In addition, we collected data from the 6001 Hapmap SNPs (in *PAX9* plus 101 genes involved in craniofacial development) for their distribution in Africans, Asians and Europeans. The *240Pro/PAX9* allele distribution, and consequently the *240Pro/Pro* homozygotes, could account for all the incidence of third molars agenesis in Africa (~ 2%). But, for other populations other genes of the tooth development network must be influencing the agenesis phenotype. Confirming this suggestion a significantly higher proportion of F_{ST} extremes values (1.3%; ≥ 0.6) were observed when craniofacial development SNPs in Africans were compared with Europeans or Asians. The scenario proposed is of strong purifying selection that would be acting on the craniofacial phenotype in Africa; but after the out-of-Africa migration a functional relaxation and even a possible positive selection might be acting in the populations of Europe and Asia. Neutrality and coalescence tests indicate that, for the Amerindians, random factors seem to be the main forces acting in the evolutionary history of these populations. (3) To understand the evolutionary dynamics of the *PAX* family among vertebrates, the nine members of this family were searched in 53 chordate genomes. There is evidence that purifying selection has been acting in the majority of vertebrate *PAX*, but a possible relaxation of the selective constraint can be assumed for some (*PAX4*, *PAX7* and *PAX8*) members. The main finding of this study is that the Sauropsida lineage (reptiles and birds) have lost two *PAX* (*PAX4* and *PAX8*) genes. Since generally *PAX* genes preserved most of their ancestral functions, the constrained or preserved genes could substitute those variables or loss by deletion. *PAX9* and the *PAX* family would be a remarkable example of how evolvability works.

CAPÍTULO 1 - INTRODUÇÃO

1.1 - CONSIDERAÇÕES INICIAIS

A humanidade está vivendo, atualmente, um momento ímpar na história da ciência. O desenvolvimento tecnológico e social permite uma interação, até então inédita, entre cientistas e pesquisadores do mundo inteiro. Os avanços nos campos das ciências da vida e da saúde são imensos, sendo estas conquistas o resultado de um esforço em conjunto de governos, universidades e também empresas privadas que se juntaram em grandes consórcios com o objetivo de atingir diversas metas históricas propostas por pesquisadores há décadas, como: o genoma humano (*Human Genome Project* -HGP - http://www.ornl.gov/sci/techresources/Human_Genome/project/about.shtml – em fevereiro de 2001 as revistas *Science* e *Nature* tiveram edições especiais apresentando o primeiro rascunho do genoma); a publicação de mapas da variabilidade genética populacional humana (International HapMap Project - <http://hapmap.ncbi.nlm.nih.gov/> - 2005; 1000 Genomes Project <http://www.1000genomes.org> - 2008; Projeto Candela- 2011); a publicação dos genomas de diversos animais ditos “modelos” (FlyBase -<http://flybase.org/> - 1992; Mouse Genome Informatics <http://www.informatics.jax.org/> -2004; Ensembl Database - <http://www.ensembl.org> – 2009) ; bem como, a disponibilização de genomas de espécies extintas a milhares de anos (UCSC Genome Browser - <http://genome.ucsc.edu> – 2000 [*Homo sapiens*] – 2010 [Neandertal e o homínídeo de Denisova]). Estes e ainda

outros avanços tecnológicos inovadores impulsionaram as fronteiras da pesquisa evolutiva nesta década, gerando, deste modo, uma imensa gama de informação que necessita ser integrada com dados de expressão, e/ou associação, e/ou funcionalidade, bem como em um contexto evolutivo para que se encontre um sentido biológico para estas descobertas (ver como exemplos da utilização de dados genômicos para estudos desta natureza nos apêndices 1, 2 e 3 desta tese).

1.2 - EVOLUÇÃO BIOLÓGICA

A palavra evolução é utilizada muitas vezes como sinônimo de mudança. No entanto, do ponto de vista biológico, evolução é um processo onde há lugar para mudanças através do tempo, mas não necessariamente, já que alguns mecanismos evolutivos podem atuar para a manutenção de um determinado *status quo* biológico através, por exemplo, da ação da seleção purificadora. Deste modo, evolução biológica é um fenômeno complexo que envolve um conjunto de mecanismos que promovem tanto a mudança, muitas destas revolucionárias, quanto à estabilidade. Os mais poderosos agentes deste processo podem, resumidamente, serem assim definidos: mutação, a fonte primária da variação, fluxo gênico, seleção natural e os processos estocásticos, que agem eliminando ou promovendo a variação independentemente de seu significado adaptativo. O padrão de variação, deste modo, pode ser observado em diferentes níveis, desde alterações em nucleotídeos até diferenças entre táxons, estando também associado a diferentes contextos, adaptativos ou não.

Embora percebida a muito por pensadores e naturalistas, a variação biológica só começou a ser compreendida efetivamente a partir da publicação da icônica obra “A Origem das Espécies”, de Charles Darwin. Ele foi capaz de apresentar uma grande quantidade de evidências sobre os padrões de variação, oriundas de diferentes áreas: pecuária, biogeografia, paleontologia, morfologia e embriologia (Darwin, 2000). Mais que isso, ele foi capaz, a partir de suas minuciosas observações empíricas, estabelecer como a seleção natural atuou sobre a variação fenotípica. Com a redescoberta dos trabalhos de Mendel e o desenvolvimento da genética populacional, no início do século 20, foi possível ir além. A chamada “Síntese Evolutiva Moderna” surge de um esforço integrativo que teve como meta agregar conhecimentos de várias áreas para explicar o fenômeno evolutivo. O esforço de pesquisadores, não muitos, mas brilhantes, foi recompensado na medida em que a Síntese teve êxito em melhorar nosso conhecimento sobre o tema. No entanto, novos desafios e questionamentos foram surgindo ao longo do processo, como normalmente ocorre com o desenvolvimento científico.

Atualmente, na era do sufixo “ômica”, descobertas nas áreas do desenvolvimento (*Evodevo*), da genômica comparada e da biologia de sistemas estão contribuindo para que seja possível estabelecer, por exemplo, as conexões entre as variantes genéticas encontradas e suas consequências fenotípicas e como isso estaria relacionado com a trajetória evolutiva dos organismos. Alguns pesquisadores afirmam que fazer todas estas conexões é uma das tarefas principais da biologia evolutiva moderna (Nielsen, 2009; Crespi, & Swanson, 2011).

1.3 - PROCESSOS MACRO E MICROEVOLUTIVOS

As bases genéticas e moleculares da adaptação evolutiva ainda são pouco conhecidas. Existem alguns poucos exemplos, bem conhecidos e documentados, onde a variação genética pode ser relacionada diretamente a uma característica com valor adaptativo. Um exemplo clássico intraespecífico, envolvendo a espécie humana são as mutações no gene da Lactase, que conferem o fenótipo da persistência do funcionamento da enzima na idade adulta (Bersaglieri e cols, 2004; Tishkoff e cols, 2007; Lefèvre e cols, 2010). Este é um exemplo clássico de estudos em um contexto microevolutivo, onde a variação populacional, conjuntamente com estudos funcionais, explicam uma novidade adaptativa. Outro exemplo seriam os estudos realizados pelo nosso grupo de pesquisa, que recentemente descreveu o primeiro alelo nativo-americano, funcional e positivamente selecionado, o alelo *230Cys* do gene *ABCA1*. A variante *230Cys* estaria relacionada com a hipótese do alelo/genótipo econômico em nativos americanos e seria um exemplo de como os fatores evolutivos interconectados como hábitos culturais estão envolvidos na geração e manutenção da diversidade (Acuña-Alonzo e cols, 2010; Apêndice 4).

Do ponto de vista macroevolutivo as novidades adaptativas são estudadas em nível de classes filogenéticas/taxonômicas. Até pouco tempo atrás este tipo de abordagem evolutiva era exclusiva de paleontólogos, mas com o advento da genômica comparada, este enfoque também passou a ser desenvolvido por geneticistas. Um exemplo é o estudo da família de genes *HOX*, responsáveis por determinar a identidade no eixo anteroposterior de diversas partes do embrião dos animais. Embora variem

consideravelmente em número, estes genes possuem sequências ortólogas tanto em artrópodes quanto em tetrápodes, sendo que alterações nestes genes já foram relacionadas a grandes mudanças de planos corporais em metazoários (Ferrier & Minguillón, 2003; Galliot, 2005; Garcia-Fernández, 2005).

Como visto nos exemplos ilustrativos dados acima, uma visão integrada onde diferenças intraespecíficas (microevolutivas) e divergências interespecíficas (macroevolutivas) são levadas em consideração, pode ser uma metodologia apropriada para seguirmos construindo nosso conhecimento sobre evolução biológica.

1.4 - GENES MESTRES DO DESENVOLVIMENTO

A dimensão e amplitude das formas de vida durante o período Cambriano pode ser vislumbrada através de um rico registro fóssil. A chamada “explosão Cambriana” ilustra o surgimento de planos corporais complexos e de inúmeros e diferentes táxons. Durante muito tempo a ausência de formas intermediárias neste registro foi interpretada como sendo resultado da inconsistência do registro fóssil (Conway-Morris, 2003) ou de um possível *saltacionismo* (Equilíbrio Pontuado; Gould & Eldredge, 1993). Esta última proposta, claramente contestava algumas premissas da evolução Darwiniana clássica (defendida tanto por Darwin quanto pelos advogados da Síntese, como visto anteriormente), que teria sido lenta e gradual. Na época, no entanto, não havia bases para explicar como tais “saltos”, que segundo os últimos autores estariam associados a grandes novidades adaptativas e a eventos de especiação, ocorreriam do ponto de vista genético.

O estudo de genes chaves do desenvolvimento sugere que entre 1 a 0,53 bilhões de anos atrás, na mesma época da explosão Cambriana, surgiu a chamada “caixa de ferramentas genéticas básicas” dos metazoários (Kirschner & Gerhart, 1998; Gerhart & Kirschner, 2007). Tal conjunto de genes teria permitido a imensa variação de formas encontrada neste período geológico. Por tanto, uma das mais surpreendentes descobertas da atualidade é que a maioria dos animais sejam eles vertebrados ou invertebrados, não importando como são diferentes em sua 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 (ou conjuntos de organismos) apresentarem processos moleculares conservados, embora tenham passado por rápidas e diversas mudanças em sua morfologia, na organização de tecidos, na fisiologia e no desenvolvimento. Um consenso é que houve várias rodadas iniciais de duplicação e diversificação destes genes mestres ou controladores, normalmente envolvendo genes codificadores de fatores de transcrição ou de crescimento. Vários estudos reportam o surgimento de novas redes regulatórias, por duplicação, divergência e posterior ganho de função. A já citada família *HOX* e ainda várias outras famílias de fatores de transcrição, são frequentemente citadas como modelos de como é complexa esta dinâmica. Estima-se que no genoma humano cerca de 1500 fatores de transcrição estejam atuando (Lynch & Wagner, 2008; Lynch e cols, 2008; Boyle e cols, 2010), dentre estes os genes das famílias *PAX* e *MSX*, investigados no presente estudo (ver itens abaixo).

Uma questão ainda controversa é se as alterações morfológicas encontradas entre e dentro táxons são produto de alterações em regiões regulatórias (Abzhanov e cols, 2007; Apêndice 5) ou resultado de alterações diretas em genes estruturais e/ou reguladores (Nei, 2007).

O estudo de um gênero de peixes *Gasterosteus aculeatus* mostrou que esta espécie é um excelente modelo evolutivo, pois após o último grande período glacial, no pleistoceno, estes peixes de vida marinha colonizaram e se adaptaram a muitos habitats de água doce, evoluindo com repetidas mudanças na forma do corpo, presença ou não de uma armadura óssea, especializações tróficas, pigmentação, adaptação a diferentes ambientes salinos, preferências para o acasalamento, e história de vida (Jones e cols, 2012). Estas mudanças foram estudadas em nível genômico populacional, e foi concluído pelos últimos autores que aproximadamente 41% da variação das características morfológicas poderia ser explicada por mudanças em regiões comprovadamente regulatórias, enquanto algo em torno de 42 % da variação estaria relacionada com alterações que são provavelmente regulatórias, e que finalmente 17 % da variação seriam produto de pequenas alterações de grande efeito em genes estruturais importantes.

Deste modo, genes que codificam fatores de transcrição envolvidos nos estágios iniciais do desenvolvimento são altamente conservados (Vogel & Chothia, 2006; Nei, 2007). Entretanto, mutações que levam a 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). Deste modo, pode-se dizer que a maioria dos eventos evolutivos aconteceria quando “velhos” genes “aprendem” novos truques (Brakefield, 2006). Da mesma forma, mutações em regiões genômicas que funcionam como aceleradores da transcrição, os chamados *enhancers*, podem alterar de maneira significativa a ação de genes, incluindo dos próprios genes reguladores, tal como é o caso do *enhancer HACNS1* e o gene *PAX9*, ambos estudados nesta tese como será visto mais adiante.

1.5 A FAMÍLIA PAX

Os genes *PAX* (*Paired box genes*) constituem uma família de genes que codificam fatores de transcrição de importância fundamental no desenvolvimento de órgãos e tecidos. Inicialmente foram descobertos em *Drosophila* (*Pox meso*). Subsequentemente foram encontrados ortólogos em diversos animais, sejam eles vertebrados ou invertebrados. Esta família de genes caracteriza-se por compartilhar um domínio conservado, chamado de domínio *Paired* que apresenta 128 aminoácidos, com propriedades de se ligar a uma região promotora do DNA alvo (Epstein e cols, 1994; Balczarek e cols, 1997; Chi & Epstein, 2002). 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). 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 são totalmente conhecidos (veja a Tabela 1).

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, 1999; Ogasawara e cols, 2000). Nos mamíferos e nos peixes foram encontrados nove genes *PAX* (*PAX 1,2,3,4,5,6,7,8 e 9*), sendo que, para os vertebrados, estes podem ser agrupados em quatro grupos parálogos, que teriam se originado antes da radiação Cambriana (Breitling & Gerber, 2000; Hetzer-Egger e cols, 2000; Miller e cols, 2000). A classificação nestes quatro grupos é baseada na presença ou ausência de três outras sequências conservadas, um homeodomínio (HD) e um octapeptídeo (OP), e um homeodomínio *Paired-type tail* (PHT; veja a Figura 1).



FIGURA 1. Esquema de classificação das subfamílias *PAX*. (Modificado de (Balczarek e cols, 1997; Breitling & Gerber, 2000; Miller e cols, 2000).

A descoberta da família *PAX* reforçou a hipótese de uma evolução em redes gênicas, onde há uma integração funcional, e a implementação de novas funções pode acontecer pelo recrutamento de novos domínios por genes chave. Esta hipótese parece confirmar-se, principalmente quando se estuda genes ativos no desenvolvimento, como os experimentos clássicos, do resgate de fenótipo da mosca com o gene *PAX6* do camundongo (Xu e cols, 1995, 1999; Xue & Noll, 1996; Crombach & Hogeweg, 2008). Entretanto, genes que possuem uma expressão ubíqua e que participam de centenas de redes funcionais, como o gene *TP53*, comportam-se diferentemente, a proteína é conservada, mas algumas redes funcionais em que está participa não o são (Paskulin e cols, 2012, no prelo; apêndice 5).

1.5.1 - *PAX1/PAX9*

PAX1 e *PAX9* constituem o Grupo I da família *PAX*, 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 & Gerber, 2000). Os 128 aminoácidos do domínio *Paired* dos genes *PAX1* e *PAX9* apresentam alta similaridade (98%). As investigações sobre o gene *PAX1* têm mostrado o papel essencial deste, especificamente na determinação da segmentação da coluna vertebral (Giampietro e cols, 2005). O gene *PAX1* está envolvido ainda, conjuntamente com o *PAX9*, na determinação do primeiro arco branquial que dá origem à maxila proximal (Sharpe, 2001). Alguns estudos têm mostrado também que *PAX1* e *PAX9* são

coexpressos e podem ter funções similares e complementares (Wilm, 1998; Peters e cols, 1999), o que seria um reflexo da uma origem evolucionária comum.

O fator de transcrição codificado pelo *PAX9* tem função considerada fundamental durante o desenvolvimento embrionário, atuando além da formação dos arcos branquiais (conjuntamente com o *PAX1*), na formação do timo, glândulas paratireoides e ossos. 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. Além do *PAX9*, a odontogênese 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 cols, 1997, 1998, 1999; Hetzer-Egger e cols, 2000; Santagati e cols, 2001; Frazier-Bowers e cols, 2002).

Vários outros estudos têm mostrado 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 de modularidade (Chen e cols, 2000; Kist e cols, 2005; Line, 2003; Peres e cols, 2005; Peters e cols, 1999).

Os estudos de mutações que provocam agenesia de dentes em humanos e em camundongos 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 (Peters e cols, 1998; Chen e cols, 2000; Kist e cols, 2005).

Como visto anteriormente nesta mesma introdução, mutações em genes regulatórios podem ser responsáveis por grandes mudanças morfológicas. Partindo dessa premissa, um estudo anterior de nosso grupo de pesquisa avaliou a possibilidade do gene *PAX9* estar envolvido em alterações dentárias encontradas na nossa e em outras espécies de primatas (Pereira e cols, 2006). Para isso foram sequenciados todos os 4 éxons 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 éxon 3. Análises envolvendo sequências 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 éxon 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 prejudicial 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 *Ala240Pro* resultaria numa redução no número de dentes e isso poderia ter tido alguma vantagem ao longo da trajetória evolutiva dos homínídeos, 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 éxon 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 éxon específico, pois esta região não estaria sujeita a restrições funcionais. O relaxamento da seleção natural sob este éxon 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 de genes. Acredita-se que o processo de formação das 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 algumas circunstâncias, poderiam ser selecionadas e fixadas (Wang & Gu, 2001). Foi sugerido também que determinados éxons de um gene específico também poderiam estar “livres” para acumular mutações, desde que os outros éxons do gene preservassem algumas das funções primordiais do gene como um todo (Pereira e cols, 2006). Ou seja, o éxon 3 seria um exemplo notável que conferiria “evolubilidade” (Kirschner & Gerhart, 1998; Pigliucci, 2008) ao *PAX9*. Como resultado, fenótipos específicos poderiam ser gerados.

Esse conjunto de estudos envolvendo o *PAX9* 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. Sendo assim, estudos que buscam desvendar mecanismos regulatórios, dentre os quais investigações que levem a identificação e caracterização de genes codificadores de fatores de transcrição, podem ser extremamente importantes para o entendimento de vários fenômenos relacionados à história evolutiva dos primatas e de outros animais.

1.5.2 - *PAX2/PAX5/PAX8*

Os genes do Grupo II (*PAX2*, *PAX5* e *PAX8*) caracterizam-se por terem em sua estrutura os domínios Paired (PD), o octapeptídeo (OC) e um homeodomínio incompleto (HD). A Comparação dos padrões de expressão deste grupo de genes (*PAX2*, *PAX5* e *PAX8*) entre peixes, anfíbios e mamíferos revelou que a especificidade da expressão destes genes, em geral é evolutivamente conservada (Torlakovic e cols, 2006). Estes genes são expressos durante o desenvolvimento do sistema nervoso central (*PAX2* e 8), do nervo ótico (*PAX2*), dos ouvidos (*PAX2*), dos rins (*PAX2* e 8), dos sistema urogenital (*PAX2* e 8), das células B (*PAX5*), da medula óssea (*PAX5*) e da tireóide (*PAX8*). E também são expressos em vários tecidos adultos, para detalhes ver tabela 1 (Li & Eccles, 2012; Dressler, 2011).

O gene *PAX2*, como os outros membros desta família, parece ter um papel central no desenvolvimento, pois mutações neste gene em humanos foram associadas a

defeitos no desenvolvimento dos rins e do nervo ótico (Síndrome Renal-Coloboma, Síndrome Papilo renal, hipoplasia renal isolada). Mutações espontâneas em camundongos, por sua vez, afetam gravemente o desenvolvimento do nervo óptico e do ouvido interno, levando a malformações do sistema nervoso e urogenital. A expressão recorrente dos genes *PAX* parece ser importante para a reparação e regeneração de tecidos. Um exemplo é o que acontece com o gene *PAX2* nos rins, pois em estudos com camundongos, foi comprovado que danos nos rins, provocam a expressão deste gene restrita temporariamente e espacialmente a região danificada (Li & Eccles, 2012; Ritz-Laser e cols, 2000).

O fator de transcrição codificado pelo *PAX5*, denominado BSAP, é expresso durante a embriogênese transitória no mesencéfalo e na medula espinal, com um padrão de expressão espacial e temporal distinto dos outros genes *PAX* no desenvolvimento do sistema nervoso central. Mais tarde, a expressão deste gene desloca-se para o fígado do embrião, estando correlacionada com o aparecimento dos linfócitos B. Também foi evidenciado que a expressão deste gene é responsável pela diferenciação das células hematopoiéticas em linfócitos B.

O terceiro membro do grupo II, o *PAX8*, é expresso no embrião de camundongos durante o desenvolvimento do sistema nervoso central, rim e da tireóide. No organismo adulto o gene *PAX8* se expressa exclusivamente na tireóide, nos rins e no sistema reprodutivo. *PAX8* é o único membro da família *PAX* que é expresso no tecido da tireóide. Em camundongos tem sido demonstrado que a proteína *PAX8* está envolvida tanto na organogênese/morfogênese da glândula tireóide. Camundongos

nocauteados para este gene apresentam defeitos graves na tireóide, sendo a glândula tireóide pouco visível e sem as células foliculares. Estes estudos levaram à conclusão de que o PAX8 é necessário para a sobrevivência e diferenciação adequada de tireócitos. Nos seres humanos, mutações no *PAX8* têm sido associadas com alguns casos de hipotireoidismo.

Finalmente, tem sido evidenciado que a superexpressão do gene *PAX8* e outros deste grupo (*PAX2* e *PAX5*) está associada a diversos tipos de tumores (Damante, 1998, 1998; Pasca di Magliano e cols, 2000; Ritz-Laser e cols, 2000; Robson e cols, 2006; Li & Eccles, 2012).

1.5.3 – PAX3/PAX7

Os genes *PAX3* e *PAX7* possuem em sua estrutura conservada, além do domínio *Paired*, todos os domínios relatados para a família *PAX*, ou seja, um homeodomínio (HD), um octapeptídeo (OP), e um homeodomínioo *Paired-type tail* (PHT). Esta conformação é considerada a estrutura original da família *PAX*, denominada proto-*PAX*. A hipótese é que houve um recrutamento de uma forma ancestral do gene codificador de uma *transposase marine*, que continha o domínio de ligação ao DNA (proto-Pax transposase), e que este foi fusionado a um homeodomínio logo após a emergência dos metazoários (Breitling & Gerber, 2000; Vorobyov & Horst, 2006). No entanto, ainda existem, pelo menos, duas propostas conflitantes de como foi a formação dos quatro grupos parálogos da família *PAX*, veja a figura 2 (Breitling &

Gerber, 2000; Gröger e cols, 2000; Vorobyov & Horst, 2006; Hill e cols, 2010; Wang e cols, 2010)

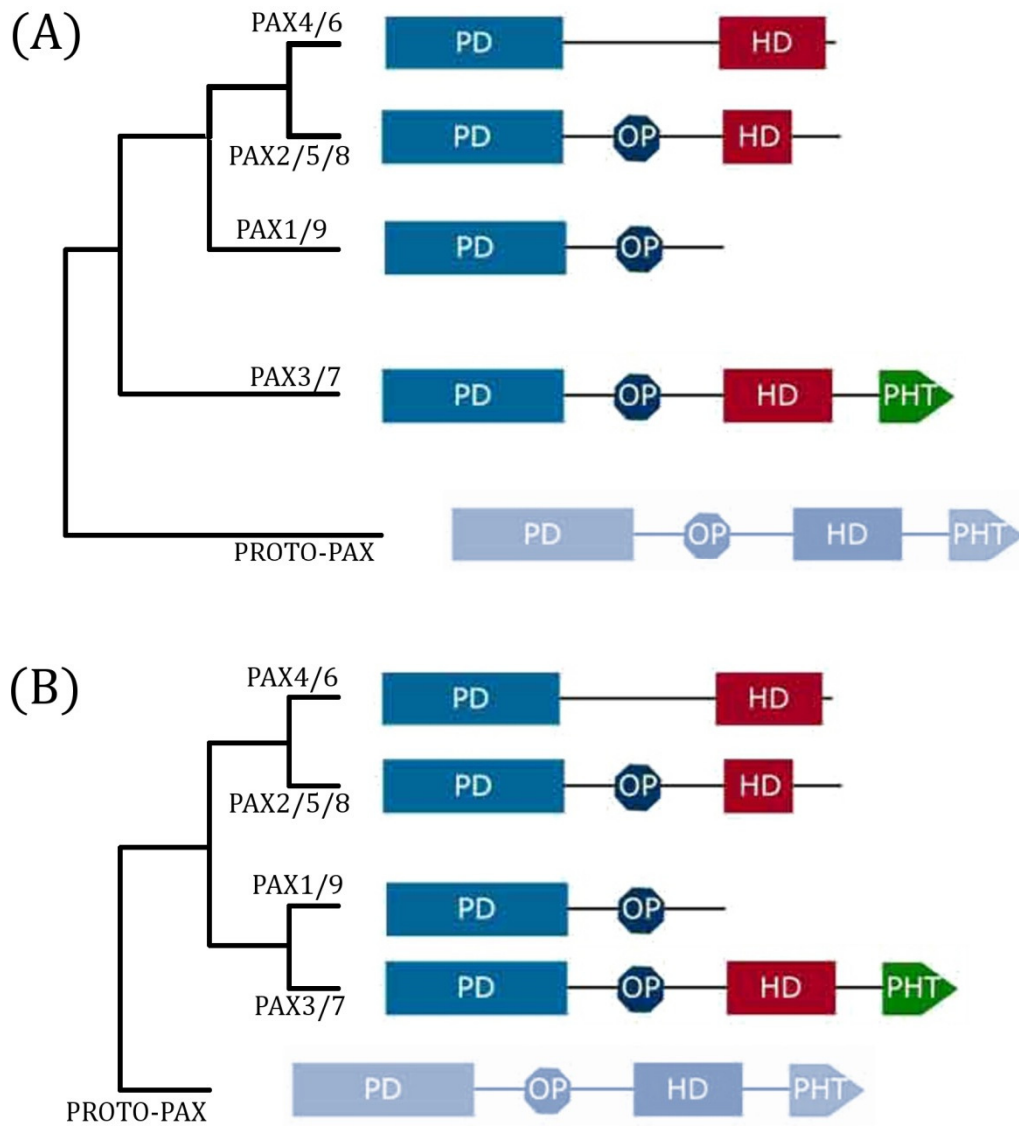


FIGURA 2. Propostas da história evolutiva da família *PAX*. A: (Gröger e cols, 2000; Chen e cols, 2010; Wang e cols, 2010) B: (Breitling & Gerber, 2000; Vorobyov & Horst, 2006);

Os fatores de transcrição *PAX3/7* possuem importantes funções durante o desenvolvimento do sistema nervoso central (SNC) e do sistema muscular esquelético (SME). O gene *PAX3* tem um interessante papel, pois coordena o desenvolvimento do

SNC, somitos, músculos esqueléticos e várias células derivadas da crista neural, tais como: tecido cardíaco, gastrointestinal, e melanócitos (Lang e cols, 2007; Kubic e cols, 2008). Outra característica notável do gene *PAX3* é o fato de possuir vários transcritos, expressos de maneira temporal e tecido específica, o que lhe confere a habilidade de controlar vários eventos durante o desenvolvimento do embrião (Short & Holland, 2008; Holland & Short, 2010). Embora, existam vários transcritos do *PAX3*, a sua sequência de aminoácidos é altamente conservada ficando em torno de 95% entre os mamíferos. A proteína *PAX3* esta presente nas células indiferenciadas do tecido epidérmico os melanoblastos, e também é encontrada nos melanócitos em tecidos adultos.

O gene *PAX7* é expresso durante o desenvolvimento no SNC e do SME conjuntamente com o *PAX3*. Além disso, ele também tem um papel central na gastrulação. Segundo Bach e colaboradores a especificação da crista neural inicia mais cedo do que anteriormente se pensava. Além disso, o gene *PAX7* também participa na formação do palato, maxila e cavidade nasal (Basch e cols, 2006; Buckingham & Relaix, 2007; Blake e cols, 2008). Foi sugerido também que tanto o gene *PAX7*, diretamente, quanto o *PAX3*, de uma forma indireta, participam na diferenciação de células satélites musculares, ativando o processo de miogênese depois de um processo de injúria no indivíduo adulto (Blake e cols, 2008).

1.5.3 – *PAX4/PAX6*

O Grupo IV (*PAX4* e *PAX6*) apresenta dois elementos de ligação ao DNA: o domínio *Paired* e o homeodomínio (figura 1). Os dois genes são essenciais para o desenvolvimento do pâncreas, e parecem ser necessários para a diferenciação das células pancreáticas. Camundongos duplo nocaute para estes genes não continham nenhuma célula endócrina madura. A função do gene *PAX4* no sistema endócrino está associado a diferenciação das ilhotas pancreáticas beta e, portanto, este gene é expresso constitutivamente após o nascimento, sendo indispensável para a expressão normal da insulina. Alguns estudos, ainda que contraditórios, encontraram associação deste de variantes no gene *PAX4* com diabetes tipo 1 e 2 (Smith e cols, 2000; Tiedge, 2005; Lang e cols, 2007; Laakso, 2009; Holland & Short, 2010). Enquanto que no pâncreas a proteína *PAX4* é importante para a diferenciação das células beta, a proteína *PAX6* é necessária para a maturação das ilhotas alfa.

O gene *PAX6* tem ainda uma função diferenciada no desenvolvimento do SNC, nas células do sistema olfativo e na diferenciação das células formadoras dos olhos. Diferentes estudos sugerem que na formação de todos os olhos complexos, há o compartilhamento de uma cascata do desenvolvimento similar, visto que mutações no gene *PAX6* interrompem o desenvolvimento de olhos tanto em mamíferos quanto em insetos. E ainda, o mais surpreendente, é que *PAX6* clonados de diversos bilatérios podem ativar ectopicamente a diferenciação de células e iniciar o desenvolvimento de olhos, tanto em drosófila quanto em sapos.

Mutações no gene *PAX6* em humanos, bem como camundongos nocautes para este gene, compartilham malformações no sistema nervoso, olfativo e defeitos no

desenvolvimento dos olhos (ausência de íris (Xu e cols, 1999; Kleinjan e cols, 2006; Laakso, 2009, 2009; Suga e cols, 2010)).

Como acima descrito, 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, de um modo tecido específico, a expressão de um *PAX* contrabalançaria a não expressão de outro, levando a um fenótipo normal ou próximo do normal, pelo menos em algumas circunstâncias. Deste modo, o estudo comparativo desta família pode fornecer importantes pistas sobre como as forças evolutivas agem após a duplicação e especialização destes genes mestres do desenvolvimento (Peters e cols, 1999; Short & Holland, 2008; Holland & Short, 2010).

1.6 – OS GENES *MSX* E A SUA INTERAÇÃO COM A FAMÍLIA *PAX*

Os genes *MSX* (*msh* (*Drosophila*) *homeo box homolog*) são genes homeóticos, e fazem parte da família *homeobox* (Hewitt e cols 1991), sendo expressos principalmente nos estágios iniciais do desenvolvimento embrionário, modulando a transcrição de outros genes. Os membros da classe de genes *MSX* compartilham grande homologia com o gene *msh* (*muscle specific homeobox*) de *Drosophila*, e são encontrados fisicamente separados no genoma dos vertebrados. A família *MSX* é aparentemente composta por três membros: *MSX1* (*Msh homeobox 1-like* (*HOX7*), *MSX2* (*Msh homeobox 2-like* (*HOX8*), *Msx3* (*Msh homeobox 3-like*- em camundongos e *MSX2P1*- *msh homeobox 2 pseudogene* , em humanos). Os genes *MSX1* e *MSX2* são os mais estudados até o momento, sendo que ambos são expressos em domínios sobrepostos, tanto espacialmente quanto temporalmente, durante o desenvolvimento de regiões do primórdio facial, membros, tubo neural e outras regiões embrionárias. Adicionalmente, esses genes são fortemente expressos na região craniofacial durante desenvolvimento, indicando um papel central dos genes *MSX* na formação dessas estruturas (Hu e cols, 2001; Alappat e cols, 2003; Finnerty e cols, 2009).

Camundongos nocauteados para ambos os genes, *Msx1/Msx2*, exibem defeitos na ossificação das suturas calvárias, dentes, orelhas, membros, pelos e glândulas mamárias, consistentes com o fato de estes genes possuírem o mesmo padrão de expressão, esta redundância parece estar restrita às fases iniciais do desenvolvimento, já que mutantes para cada um dos genes apresentam fenótipos específicos. Enquanto camundongos mutantes para o *MSX2* apresentam a ossificação das suturas calvárias

atrasadas ou incompletas, enquanto os defeitos nos dentes parecem estar relacionados ao esmalte. Camundongos mutantes para o *MSX1* são incapazes de levar a termo o desenvolvimento dos dentes, além de apresentarem defeitos de fechamento do palato (Alappat e cols, 2003; Fleischmannova e cols, 2008; Finnerty e cols, 2009).

Humanos com mutações no gene *MSX1* apresentam o palato secundário e/ou lábio fendido e diversas agenesias de dentes permanentes, enquanto que mutações do *MSX2* causam craniosinostoses, do tipo Boston, que resultam do fechamento precoce das suturas cranianas, além de outras más formações orofaciais. Esses dados confirmam o importante papel que os genes *MSX* possuem na determinação das estruturas craniofaciais (Alappat e cols, 2003; Ogawa e cols, 2005; Fleischmannova e cols, 2008; Finnerty e cols, 2009).

Os genes *PAX* e *MSX*, interagem e são expressos simultaneamente em diversos tecidos durante o desenvolvimento. Existem diversos relatos de coexpressão dos genes *PAX/MSX*, nos mais diversos tipos de tecidos embrionários: *PAX6/MSX1/MSX2* são expressos conjuntamente em eventos iniciais da morfogênese craniofacial e do tubo neural/somitoss, além dos olhos (Foerst-Potts & Sadler, 1997); *PAX3/MSX2* são expressos na região dorsal da espinha dorsal em desenvolvimento e próximos as margens do tubo neural (Bang e cols, 1999). Adicionalmente existem evidências que o domínio *Paired* do *PAX3* e o homeodomínio do *MSX1* são essenciais para a interação das duas proteínas codificadas. *PAX9/MSX1*, por sua vez, são expressos conjuntamente e são responsáveis pela formação e possível diferenciação dos tipos dentários em mamíferos. Existem evidências de uma relação funcional entre *PAX9* e *MSX1*, com

interações em 2 níveis, o fator de transcrição PAX9 se liga diretamente ao promotor do gene *MSX1* e as proteínas PAX9/MSX1 formam um complexo heterodímero. Mutações no domínio *Paired* do PAX9, e no homedomínio do *MSX1* afetam diretamente estas interações e são associadas a diversos tipos de agenesia dentárias (Ogawa e cols, 2005, 2006).

CAPÍTULO 2 - OBJETIVOS

2.1 – OBJETIVO GERAL

Buscar uma visão integrada das questões evolutivas sobre os genes do desenvolvimento considerando diferenças intraespecíficas e divergências interespecíficas, para tentar compreender o surgimento de padrões (moleculares e fenotípicos) durante a evolução dos vertebrados.

2.2 – OBJETIVOS ESPECÍFICOS

2.2.1 – Analisar a variação intraespecífica do gene *PAX9*, e de outros genes relacionados ao desenvolvimento, bem como suas participações na determinação de fenótipos craniofaciais, contribuindo para elucidar alguns aspectos da evolução molecular desses genes em diferentes populações humanas.

2.2.1.a - Identificar variantes nos genes *PAX9* e *MSX1*, bem como seu padrão de herança, que estariam relacionadas com diferentes formas de agenesia dentária não-sindrômicas humanas, mais especificamente com a agenesia dos terceiros molares, através de um estudo caso-controle. Resultados que atingem este objetivo são descritos no item Resultados-Capítulo 3;

2.2.1.b - Testar a hipótese de que na espécie humana poderia ser detectada a ação da seleção positiva no gene *PAX9*, bem como em outros genes relacionados ao desenvolvimento craniofacial, devido

particularmente à dispersão geográfica do *Homo sapiens*. Resultados que atingem este objetivo são descritos no item Resultados-Capítulo 4;

2.2.2 - Estudar a evolução da família *PAX*, por meio da genômica comparada, em diferentes grupos filogenéticos de cordados. Tentar relacionar as mudanças moleculares às novidades adaptativas que surgem quando emergem os vertebrados, bem como em linhagens particulares destes. Contribuir para o delineamento do cenário evolutivo em nível molecular dos genes parálogos desta família de genes mestres do desenvolvimento. Resultados que atingem este objetivo são descritos no item Resultados-Capítulo 5.

Vale ressaltar que, adicionalmente, seis apêndices são ainda apresentados. Os mesmos reportam seis artigos científicos (2 publicados, 4 aceitos para publicação) desenvolvidos durante este doutorado que de uma forma ou de outra envolveram conhecimentos relacionados aos objetivos desta tese como pode ser visto ao longo de citações no item Introdução-Capítulo 1. Adicionalmente, foram ainda desenvolvidas atividades que resultaram em mais dois artigos científicos já publicados, cujo tema envolve populações ameríndias, e que foram citados no item Conclusões-Capítulo 6 para ilustrar o particular padrão de diversidade intra e interpopulacional deste grupo nativo continental e o quanto isso está relacionado aos resultados encontrados com os genes da família *PAX*.

RESULTADOS

***CAPÍTULO 3 – RESULTADOS: “PAX9 AND MSX1 TRANSCRIPTION
FACTOR GENES IN NON-SYNDROMIC DENTAL AGENESIS”***

***CAPÍTULO 4 – RESULTADOS - “GENETIC VARIATION AMONG
MAJOR HUMAN GEOGRAPHIC GROUPS SUPPORTS A PECULIAR
EVOLUTIONARY TREND IN PAX9”***

Genetic Variation among Major Human Geographic Groups Supports a Peculiar Evolutionary Trend in PAX9

Vanessa R. Paixão-Côrtés¹, Diogo Meyer², Tiago V. Pereira³, Stéphane Mazières^{4,5}, Jacques Elion⁶, Rajagopal Krishnamoorthy⁶, Marco A. Zago⁷, Wilson A. Silva Jr.⁸, Francisco M. Salzano¹, Maria Cátira Bortolini^{1*}

1 Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Porto Alegre, Rio Grande do Sul, Brazil, **2** Departamento de Genética e Biologia Evolutiva, Instituto de Biociências, Universidade de São Paulo, São Paulo, São Paulo, Brazil, **3** Departamento de Bioquímica, Universidade Federal de São Paulo, São Paulo, São Paulo, Brazil, **4** Laboratoire d'Anthropobiologie, FRE2960, CNRS, Toulouse, France, **5** Anthropologie Bioculturelle, Faculté de Médecine, CS80011, Marseille, France, **6** Inserm, UMR 763, Université Paris Diderot, Hôpital Robert Debré, Paris, France, **7** Departamento de Clínica Médica, Faculdade de Medicina, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brazil, **8** Departamento de Genética, Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brazil

Abstract

A total of 172 persons from nine South Amerindian, three African 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 bp and 220 bp, respectively) and integrated with the information available for the same genetic region from individuals of different geographical origins. Nine mutations were scored in exon 3 and six in its flanking regions; four of them are new South American tribe-specific singletons. Exon3 nucleotide diversity is several orders of magnitude higher than its intronic regions. Additionally, a set of variants in the *PAX9* and 101 other genes related with dentition can define at least some dental morphological differences between Sub-Saharan Africans and non-Africans, probably associated with adaptations after the modern human exodus from Africa. Exon 3 of *PAX9* could be a good molecular example of how evolvability works.

Citation: Paixão-Côrtés VR, Meyer D, Pereira TV, Mazières S, Elion J, et al. (2011) Genetic Variation among Major Human Geographic Groups Supports a Peculiar Evolutionary Trend in PAX9. PLoS ONE 6(1): e15656. doi:10.1371/journal.pone.0015656

Editor: Dennis O'Rourke, University of Utah, United States of America

Received: August 20, 2010; **Accepted:** November 19, 2010; **Published:** January 27, 2011

Copyright: © 2011 Paixão-Côrtés et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by Conselho Nacional de Desenvolvimento Científico e Tecnológico, Institutos do Milênio and Programa de Apoio a Núcleos de Excelência Programs, Fundação de Amparo à Pesquisa do Estado do Rio Grande do Sul, and Inserm-FAPESP 2010-2011 Joint Projects Agreement, Brazil and France. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interest exist.

* E-mail: maria.bortolini@ufrgs.br

Introduction

The central goal of genetics is to understand how heritable information encoded in the genome determines the phenotype of an organism [1]. Some results support the idea that morphological evolution can occur basically by a small proportion of major effect mutations in structural, developmental or regulatory key genes (major gene effect hypothesis [2,3]. Considering this possibility, we studied the entire coding region as well as exon-intron boundaries of the *Paired box homeotic gene 9* (*PAX9*; a transcription gene with a major role in early development, including dentition) to investigate the importance of small changes in major protein-coding genes that could be involved with the normal intra and interspecies dental variation [4].

The *PAX9* coding sequence is divided into four exons. The first exon acts uniquely as the start codon, whereas exon 2 codes for a highly conserved region of the *PAX9* protein, the paired domain, which is responsible for sequence-specific contacts with DNA [5, 6], <http://www.ensembl.org/index.html>. The vital importance of exon 2 is illustrated by an impressive absence of nucleotide sequence variation at this region among 44 species of mammals (V.R. Paixão-Côrtés et al., unpublished data), as well as within and between human populations, and the identification of mutations in the paired domain as etiologic for several cases of both nonsyndromic hypodontia and oligodontia in our species (<http://www.ncbi.nlm.nih.gov/omim> - OMIM#167416; OMIM#142893).

Similarly, although the fourth exon of *PAX9* lacks functional information, it is located in a region which is also relatively conserved across 44 mammalian species (V.R. Paixão-Côrtés et al., unpublished data), and is involved in molar oligodontia in humans, indicating a prominent role of this exon for *PAX9* function [7]. However, the expectation that a similar degree of evolutionary constraint might also be observed at exon 3 of *PAX9* was not confirmed by us [4]. We have obtained evidence not only for a high level of human intraspecific variation, but also that a common polymorphism (*Ala240Pro*; rs4904210) is probably functional and could be associated with third molar agenesis, since *240Pro* homozygotes might present a *PAX9* protein with a slightly reduced DNA-binding capacity [4]. Additionally, our recent [8] as well as other family studies [9,10], showed that the derived *240Pro* allele has a recessive pattern of inheritance with variable expressivity, since all homozygotes for it had congenital missing third molar(s); the number of missing teeth, however, was different for each affected individual, in agreement with Pereira et al's hypothesis [4].

Here, we presented an analysis based on new *PAX9* exon 3 sequence information and its exon-intron border regions from Native American, African, and Eskimo populations, which were integrated with published information encompassing all major human geographic groups. We also used data from HapMap Phase II [11] to identify signals of possible different genetic

backgrounds involved in the dental development in Africans and non-Africans. Finally, we explored the evolutionary scenario related to the origin and prevalence of the *PAX9* polymorphisms, suggesting that exon 3 is an important agent in the ability to evolve (evolvability) of this extremely conserved gene.

Results

Mutations and polymorphic sites

A total of 590 base pairs, corresponding to *PAX9* exon 3 and its 5' and 3' flanking intronic regions were sequenced in 344 human chromosomes, and the data were analyzed together with other previously published sequences [4]. There are four polymorphic sites in the 5' flanking and two in the 3' flanking intronic regions in the Apalaí, Galibi, Mekranoti, European and African populations (**Table 1**). Only one of these, the variant allele *A* at position 36,205,241 was present in homozygous state in one Galibi individual. As for the sequence changes observed in exon 3, almost all were seen in heterozygosity, two exceptions being variants located at codons 239 and 240 (C→T, *His239His*; and G→C, *Ala240Pro*; see below).

We found several singletons in exon 3, representing new South American tribe-specific variants. They were the G→A (*Ser226Asn*) transition (Apalaí), which is located in the same position as a previously described polymorphism found in the Bari from Venezuela (G→T, *Ser226Ile*); the G→A (*Ala236Thr*) transition (Tenharim); the T→A (*Asn242Asp*) transversion (Arara); and the A→T (*Glu250Val*) transversion (Apalaí; **Table 1**). Another C→A (*Glu251Lys*) mutation, previously seen exclusively in the Guarani [4] was not observed in the present investigation. On the other hand, the derived *T* allele of the *His239His* single nucleotide polymorphism (SNP) previously observed only in the Kaingang of southern Brazil [4] was now also found among the Eskimo. Since this variant was also observed with relatively high frequencies in European (27%) and African (3%) populations, its presence among the Kaingang and Eskimo could be due to admixture with non-Amerindians.

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 Mundurucu, Tenharim and Xikrin. Curiously, in this same position, 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 scattered 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. *Ala240Thr*, on the other hand, 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 Jéan; split at least ~5 thousand years before present [12]). Finally, allele *T* at the *His239His* SNP is only present in populations where the *Ala240Pro* allele *C* is also found (**Table 1**), but they occur in different individuals, indicating that the derived alleles, *239-T* and *240-C*, are not in linkage disequilibrium.

To better explore the results obtained with the *Ala240Pro* SNP, detailed information on its genotype and allele frequencies is given in **Table 2**. Although some 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 and lowest *C/Pro* frequencies can be found in Asia (45%) and Africa (20%), respectively, while in Eskimos and Europeans (30%) the values are intermediate [4, 13, present study].

Genetic variability

Nucleotide diversity (π) values for *PAX9* exon 3 and its intronic flanking regions are summarized in **Table S1**. Considering exon 3 only, values for π ($\times 10^4$) for the Amerindian sample as a whole (16.4) are lower than those obtained for the Eskimos (37.2), Europeans (56.1), and Africans (28.1). Curiously the African value is lower than that observed for Europeans, contrasting with studies using neutral markers [14]. Table S1 also shows that the π values for exon 3 are generally several orders of magnitude higher than those calculated for the non-coding regions in all populations, indicating a reduced evolutionary constraint.

Demographic simulations and neutral/natural selection tests using *PAX9* human data

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, since population history and demography can result in both negative or positive values for these statistics. The strong negative values found for South Amerindians (**Table 3**) are probably determined by demographic processes, since their significance is lost when scenarios of bottlenecks with later expansion are simulated (**Table S2**). Note, however, that for the remaining groups several values are positive and statistically significant under the different population growth scenarios, considering the exon 3 datasets only (**Tables 3 and S2**). Positive values for these statistics generally indicate an excess of intermediate frequency alleles which is incompatible with a simple neutral model.

Third molar agenesis and the *Ala240Pro* polymorphism

The number of *CC/ProPro* homozygotes (2%) observed in our Sub-Saharan samples is compatible with the frequency of agenesis in people that live in this African region (~2%, excluding Khoi-San and Pygmy [15–17]). On the other hand, the frequency of *CC/ProPro* homozygotes alone (7%, 7%, 25%) cannot explain the level of third molar agenesis among Eskimos (~18% [18]), Europeans (~20%; [17,19,20]) and Asians (47% [21]). In fact, our familial study [8] indicated that *GG/ArgArg* homozygotes also display third molar agenesis. Anthropological studies conducted with Native Americans, on the other hand, reveal that it is difficult to establish a general third molar agenesis pattern for Native Americans as a whole, due to the large amount of variation found (from zero to ~40% [18,22–25]). It seems that the *Ala240Pro* polymorphism has little influence in the level of third molar agenesis detected in Amerindian populations.

The search for distinct gene dental complexes using the HapMap panel

As indicated above, other genetic variants besides those of *PAX9* should influence the dental, jaw and maxillary development in different human populations. We have therefore searched for possible distinct continental gene dental complexes using information from Africans, Europeans and Asians. Unfortunately, the available data for Native Americans are not sufficient to include them in the analysis. Data from 102 dental development genes and their adjacent regions were considered (release 21 - <http://hapmap.ncbi.nlm.nih.gov/> [11]). **Table S3** lists these genes and **Table S4** the traits that are associated to them. The frequencies of 6,001 SNPs distributed along a 10 Mbp of the indicated region was then obtained and F_{ST} values for African/European, African/East Asian, and European/East Asian comparisons were calculated. Higher F_{ST} values ($1.3\% \geq 0.6$) were observed when Africans were compared with Europeans or Asians. The former number is significantly higher than that estimated by Wu and Zhang [26] using a random gene panel. This excess of high F_{ST} values can be

Table 1. Mutations observed in PAX9 exon3^a.

Mutations	Populations (Numbers below the names indicate the total of chromosomes sequenced) ^b																			
	Aché	Apalai	Arara	Bari	Cayapo	Guarani	Galibi	Jamamadi	Kaingang	Mekranoti	Munducuru	Tenharim	Warao	Xikrin	Yucpa	Eskimo	Japanese	European	African	
Region	40	18	42	8	2	10	14	10	40	10	16	16	8	34	8	88	28	30	94	
Position in chrom. 14	SNP	Amino acid change		Nucleotide change																
5' Flanking	36. 205. 241																			
	36. 205. 341									1										
	36. 205. 356																		1	8
	36. 205. 362		1																	
Exon 3	36. 205. 463		1																	
	36. 205. 463																			
	36. 205. 492																			
	36. 205. 503	rs12881240																		
	36. 205. 504	rs4904210		2	3			4	5		1	2	0	2	0	26	4	7	19	
	36. 205. 504																			
	36. 205. 510																			
	36. 205. 535																			
	36. 205. 537																			
3' Flanking	36. 205.575									0										1
	36. 205.611									1										1

^aSequence of 588 bp corresponding to PAX9 exon 3 (137 bp) and its 5' and 3' flanking intronic segments (232 bp and 219 bp, respectively). Numbers in columns 6–24 are those observed for the derived allele in each population.
^bSources: Present study, Apalai, Arara, Cayapo, Galibi, Jamamadi, Mekranoti, Munducuru, Tenharim, Xikrin, Eskimo, and Africans. Aché, Bari, Guarani, Kaingang, Warao, Yucpa, Japanese, and Europeans [4].
 doi:10.1371/journal.pone.0015656.t001

Table 2. Genotype and allele distributions of the Ala240Pro polymorphism in 15 South Amerindian and other human populations^a.

Population	Number of individuals	Genotype frequency			Allele frequency		References
		Ala/Ala G/G (%)	Ala/Pro G/C (%)	Pro/Pro C/C (%)	G (Ala)	C (Pro)	
<i>South Amerindian</i>							
Aché	20	20 (100)	0	0	1	0	[4]
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	[4]
Cayapó	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	[4]
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	[4]
Mekranoti	5	5(100)	0	0	1	0	Present study
Mundurucu	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	[4]
Xikrin	17	15 (88)	2 (12)	0	0.94	0.06	Present study
Yucpa	4	4 (100)	0	0	1	0	[4]
Total	138	119 (86)	18 (13)	1 (<1)	0.91	0.08	
Eskimo	44	21(48)	20(45)	3 (7)	0.70	0.30	Present study
Asians ^b	116	40 (35)	47 (40)	29 (25)	0.55	0.45	[4,15]
Europeans ^c	365*	169 (46)	172 (47)	24 (7)	0.70	0.30	[4]
Africans	47	29 (62)	17 (36)	1 (2)	0.80	0.20	Present study

^aData obtained from RFLP or sequencing;

^bChinese and Japanese;

^cSpaniards and Polish.

doi:10.1371/journal.pone.0015656.t002

a signal for natural selection, since population-specific positive selection tends to increase F_{ST} [27].

Additionally, a high proportion (65–70%) of extreme integrated haplotype score (iHS) negative values (<−3.0), for the top score *PAX9* SNPs was observed for Asians and Europeans (Table S5), while in Africans, SNPs with extreme iHS negative values represented only 30% of the cases. An excess of extreme iHS negative values can indicate that the derived alleles (or haplotypes)

have swept up in frequency, an indication of positive selection [28]. On the other hand, the extreme iHS positive values for Africans occur in 70% of the cases. This indicates that the ancestral alleles have been fixed preferentially in Africa [28].

These results reinforce the idea that *PAX9* is not the unique gene involved in the dental development and that only an approach based on gene network can lead to a more comprehensive view of the complex biology involved.

Table 3. Statistical tests to detect departures from a standard neutral model ^a.

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**z
Eskimo	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
Africans	−1,19605	−0,01004	−1,97525	−2,02829	0,69032	0,55518

^aSequence data from the present study and [4]. Statistical significance:

*P<0.05;

**P<0.02, ND: Not determined.

doi:10.1371/journal.pone.0015656.t003

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 diversity is a result of variable selective pressures in different geographical regions, each with its own different climate, pathogens and sources of food [29–32]. Additionally, natural selection may act differently in distinct genome segments. While portions of a given gene, for example, are largely invariable due to a strong purifying selection, other regions may vary due to positive selection, relaxation of purifying selection, and/or balancing selection [33,34]. The present work illustrates this concept, demonstrating that the human *PAX9* might be a good example of these interpopulation and intragene evolutionary dynamics.

By analyzing genetic variation in several Native American tribes, one Eskimo population and other major human groups, we provide evidence for higher nucleotide diversity in exon 3 than in the non-coding sequences of the human *PAX9*. Given the features previously associated with other segments of *PAX9* (i.e. a high degree of conservation due to strong purifying selection in exons as well as in intronic regions across species [4], a pertinent question is why the third exon of this gene possesses such dissimilar pattern of nucleotide diversity when compared to other coding and non-coding regions.

In about 60,000 years, modern humans left Africa and colonized all other continents, with the exception of Antarctica [35,36]. Since then, human populations have lived in an extraordinary variety of habitats. Additionally, for many human populations, the diet and life style changed drastically due to cultural practices. In this context, a lower number of teeth could be advantageous in populations with progressive reduction of the dental arches and a diet based on processed food, since it would prevent impacted and defective third molars, as well as related-oral problems, an important cause of morbidity in non-industrial societies [4,37]. Since *PAX9* has a crucial role in skeleton, thymus and parathyroid glands organogenesis [38–40] it is valid to speculate that the high level of diversity in only one specific region of the gene might offer options for selection in a context as those resulting from dispersal to new habitats or rapid cultural changes. Biologists have increasingly been asking whether evolvability, the ability of a biological system to evolve, itself evolves, and whether this phenomenon is a result of natural selection or a by-product of other evolutionary processes [41]. *PAX9* exon 3 could be a good molecular example as how evolvability develops, since it may have been an alternative to promote functional adaptations in human dentition. Our results, based on demographic simulations and neutral selection tests, showed that the variation found in *PAX9* exon 3 is incompatible with a simple neutral model, being consistent with the fact that at least one variant at this exon (*Ala240Pro*) might be functional.

Finally, our databank analysis detected a complex involving *PAX9* and 101 other genes related to dentition that can be under different patterns of evolution in Africa and in other continents, defining at least some dental morphological differences between Sub-Saharan Africans and non-Africans, probably associated with adaptations after the modern human exodus from Africa.

Methods

Populations examined

PAX9 exon 3 and its flanking intronic sequences were obtained from 172 individuals from nine South Amerindian (N = 81) one Siberian Eskimo (N = 44) and a composite Sub-Saharan African sample (N = 47). Nucleotide sequences of the same DNA region from 15 Europeans, 14 Japanese and 57 South Amerindians were re-analyzed based on our previous investigations [4]. Additional

data for the *Ala240Pro* (rs4904210) polymorphism were obtained through HAPMAP ([http://www.hapmap.org/cgi-perl/\[11\]](http://www.hapmap.org/cgi-perl/[11])) and dbSNP (<http://www.ncbi.nlm.nih.gov/projects/SNP/>) databases. Relevant information about the new populations studied is provided in Supplemental Material **Table S6**.

Ethical approval for the present study was provided by the Brazilian National Ethics Commission (CONEP Resolution no. 123/98) for the Amerindian and Siberian samples, as well as by ethics committee of Hôpital Robert Debré, Paris France for the African ones. Informed oral consent was obtained from all participants, since they were illiterate, and they were obtained according to the Helsinki Declaration. The ethics committees approved the oral consent procedure.

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 S6**, using the QIAamp DNA MiniKit (Qiagen) and following the manufacturer's instructions. Non-Brazilian samples were collected and processed as indicated in the references given in **Table S6**.

PAX9 exon 3 (137 bp) and its 5' and 3' flanking intronic segments (232 bp and 219 bp, respectively) were amplified using primers and conditions described elsewhere [4]. The PCR products were purified following the Exonuclease I and Alkaline Phosphatase (Amersham Biosciences - GE Healthcare) protocols. Both DNA strands were sequenced using MegaBace500 and MegaBace1000 (Amersham Biosciences - GE Healthcare) sequencing machines. Whenever a putative mutation was identified, the sample was re-amplified and re-sequenced for confirmation.

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.

For all groups with exon 3 *PAX9* sequences available one population parameter was estimated using Arlequin 3.11 (<http://cmpg.unibe.ch/software/arlequin3/>): nucleotide diversity (π) [2]. To detect departures from a standard neutral model, we used Tajima's D [42] and Fu and Li's D* and F* [43] tests (implemented using Mega 4.0 [44] and DnaSP 4.10.9 [45]). Significantly positive and negative values for these statistics lead to the rejection of the null hypothesis of mutation-drift equilibrium. However, since it is well-known that these parameters are also affected by the demographic histories of the population considered [29], we incorporated in the analysis coalescence simulations using Rogers' algorithm [46] as implemented by [47] in the DFSC 1.1 software (www.xmission.com/wooding/DFSC/). Distinct scenarios on times of expansion and initial population sizes compatible with the available data were simulated as follows: 1) For all samples, initial populations of 1,000 or 10,000 individuals [48,49] starting 60,000 years ago [35,36,49]; 2) For all non-African populations, initial population of 10,000, date 40,000 years ago; 3) For Japanese only, 3,000 starting individuals, date 37,000 years ago [50]; 4) For the Native American and Eskimo dataset, initial populations of 80 or 500 individuals [51,52], starting 18,000 years ago [51,53]. For all scenarios described above we considered assumptions of population growth ranging from 1-fold (stationary) to 100-fold. The substitution rate (μ) used to obtain the θ values under the different scenarios was 1×10^{-9} [47].

Based on the approach used by Wu and Zhang [26], we also searched a set of genes related with dental development using the

Gene Ontology database (browser AmiGO - <http://amigo.geneontology.org/cgi-bin/amigo/browse.cgi> [54]). This procedure allowed the selection of 102 genes (including *PAX9*) located in the autosomes (**Table S3**). The function of these genes was verified using GeneDeks V3 software (<http://www.genecards.org/> [55]- **Table S4**). The major trait phenotype associated to these genes was growth/size (63,11%) followed by craniofacial characteristics (57,28%). A region covering 10 Mbp, upstream of the 5' end and downstream of the 3' terminus for each of the 102 genes was used to select SNPs, and a total of 6,001 SNPs were selected. Their allele frequencies were retrieved from HapMap Phase II (release 21 - <http://hapmap.ncbi.nlm.nih.gov/> [11]) from three populations: African (YRI panel including 60 Yoruban individuals from Ibadan), European (CEU panel including 60 Utah residents with northern and western Europe ancestry) and 89 East Asian (ASN panels including Han Chinese- HCB and Japanese from Tokyo-JPT). Unfortunately, the rs4904210 *Ala240Pro* polymorphism is not included in these panels. F_{ST} values between African (YRI)/European (CEU), African (YRI)/East Asian (EA) and European (CEU)/East Asian (EA), as well as iHS (integrated haplotype score statistic, used for scanning single nucleotide polymorphism data for signals of selection [28] were recovered from the Haplotter <http://haplotter.uchicago.edu/selection/> and SNP@Evolution (<http://bighapmap.big.ac.cn/>;[56]) databases.

Supporting Information

Table S1 Additional information's about mutations and nucleotide sequence diversity estimates of *PAX9* exon 3 and non-exon 3 bordering regions in several human populations. (DOC)

References

- 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.
- Nei M (1987) Molecular evolutionary genetics. New York: Columbia Univ Press.
- Nei M (2007) The new mutation theory of phenotypic evolution. *Proc Natl Acad Sci USA* 104: 12235–12242.
- Pereira TV, Salzano FM, Mostowska A, Trzeciak WH, Ruiz-Linares A, et al. (2006) Natural selection and molecular evolution in primate *PAX9* gene, a major determinant of tooth development. *Proc Natl Acad Sci USA* 103: 5676–5681.
- Xu W, Rould MA, Jun S, Desplan C, Pabo CO (1995) Crystal structure of a paired domain-dna complex α at 2.5 Å resolution reveals structural basis for Pax Developmental Mutations. *Cell* 80: 639–650.
- Chi N, Epstein JA (2002) Getting your Pax straight: Pax proteins in development and disease. *Trends Genet* 18: 41–47.
- Frazier-Bowers SA, Guo DC, Cavender A, Xue L, Evans B, et al. (2002) A novel mutation in human *PAX9* causes molar oligodontia. *J Dent Res* 81: 129–133.
- Paixão-Cortes VR, Braga TS, Salzano FM, Munsdtstock KS, Mundstock CA, et al. (2010) *PAX9* and *MSX1* transcription genes in non-syndromic dental agenesis. *Arch Oral Biol* (in press).
- Trimmell JB (2004) A mutation of *PAX9* associated with congenitally missing teeth. M. Sc. Dissertation, Kansas City: University of Missouri.
- Kula K, Trimmell J, lu Y, Briscoe P, Feng JQ (2008) Tooth agenesis in a family and homozygous *PAX9* mutation in exon 3: a case report. *World J Orthod* 9(4): e55–61.
- The International HapMap Consortium (2007) A second generation human haplotype map of over 3.1 million SNPs. *Nature* 449: 851–861.
- Urban G (1998) A história da cultura brasileira segundo as línguas nativas. In: Carneiro da Cunha MC, ed. História dos índios no Brasil. São Paulo: Companhia das Letras, pp 87–102.
- Pan Y, Wang L, Ma J, Zhang W, Wang M, et al. (2008) *PAX9* polymorphisms and susceptibility to sporadic tooth agenesis: a case-control study in southeast China. *Eur J Oral Sci* 116: 98–103.
- Wang S, Lewis CM, Jakobsson M, Ramachandran S, Ray N, et al. (2007) Genetic variation and population structure in Native Americans. *PLoS Genet* 3: e185.
- Irish JD (1993) Biological affinities of late Pleistocene through modern African aboriginal populations: the dental evidence. Ph.D. Dissertation, Tempe: Arizona State University.
- Chagula WK (1960) The age at eruption of third permanent molars in male East Africans. *Am J Phys Anthropol* 18: 77–82.
- Rozkvcová E, Marková M, Dolejší J (1999) Studies on agenesis of third molars amongst populations of different origin. *Sb Lek* 100: 71–84.
- Turner CG II (1986) The first Americans: the dental evidence. *Natl Geogr Res* 2: 37–46.
- Rozkvcová E, Marková M, Lánik J, Zvárová J (2004) Development of third molar in the Czech population. *Prague Med Rep* 105: 391–422.
- Rozkvcová E, Marková M, Lánik J, Zvárová J (2004) Agenesis of third molars in young Czech population. *Prague Med Rep* 105: 35–52.
- Turner CG II (1987) Late Pleistocene and Holocene population history of East Asia based on dental variation. *Am J Phys Anthropol* 73: 305–321.
- Sutter R (2005) The prehistoric peopling of the South America as inferred from epigenetic dental traits. *Andean Past*, Vol. 7. Ithaca: Cornell University.
- Gómez-Valdés J (2005) Antropología dental en poblaciones del Occidente de Mesoamérica. Tesis Licenciatura en Antropología Física. México: Escuela Nacional de Antropología e Historia.
- Turner CG II, Bird J (1981) Dentition of Chilean paleo-Indians and peopling of the Americas. *Science* 212: 1053–1055.
- Hydenblit R (1996) Dental variation among four prehispanic Mexican populations. *Am J Phys Anthropol* 100: 225–246.
- Wu DD, Zhang YP (2010) Positive selection drives population differentiation in the skeletal genes in modern humans. *Hum Mol Genet* 19: 2341–2346.
- Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L (2008) Natural selection has driven population differentiation in modern humans. *Nat Genet* 40: 340–345.
- Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. *PLoS Biol* 4: e72.
- Bamshad M, Wooding SP (2003) Signatures of natural selection in the human genome. *Nat Rev Genet* 4: 99–111.
- Harris EE, Meyer D (2006) The molecular signature of selection underlying human adaptations. *Yearb Phys Anthropol* 49: 89–130.
- Hughes DA, Tang K, Strotmann R, Schöneberg T, Preen J, et al. (2008) Parallel selection on *TRPV6* in human populations. *PLoS One* 3: e1686.
- Acuña-Alonso V, Flores-Duarte T, Kruit J, Villareal-Molina T, Arellano-Campos O, et al. (2010) A functional ABCA1 gene variant is associated with low HDL-cholesterol levels and shows evidence of positive selection in Native Americans. *Hum Mol Genet* 19: 2877–2885.

Table S2 P values of Tajima's D, and Fu and Lis D and F statistics calculated under several demographic scenarios (XLS)

Table S3 List of dental development genes. (XLS)

Table S4 Traits associated to 102 genes (XLS)

Table S5 Integrated haplotype score (iHS) values for the top score SNPs of *PAX9* (DOC)

Table S6 Sample sizes, geographic location, and linguistic information on fifteen South Amerindian, one Eskimo and one composite African populations investigated in this study (DOC)

Acknowledgments

Thanks are due to the people who, after appropriate informed consent, donated saliva and blood for this investigation. The logistic support of the National Indian Foundation (Fundação Nacional do Índio-FUNAI) in the Brazilian field work is gratefully acknowledged. We also thank S. L. Bonatto for the Eskimo sample and V. A. Alonso, M.E. Delgado-Burbano, and J. Gómez-Valdés for sources of information about third molar agenesis.

Author Contributions

Conceived and designed the experiments: MCB VRP-C. Performed the experiments: VRP-C. Analyzed the data: MCB VRP-C FMS. Contributed reagents/materials/analysis tools: DM TVP SM WASJ MAZ JE RK. Wrote the paper: MCB VRP-C FMS.

33. Yang Z, Nielsen R, Goldman N, Pedersen AM (2000) Codon-substitution models for heterogeneous selection pressure at amino acid sites. *Genetics* 155: 431–449.
34. Hughes AL (2007) Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity* 99: 364–373.
35. 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 USA* 98: 864–869.
36. Santos-Lopes SS, Pereira RW, Wilson IJ, Pena SD (2007) A worldwide phylogeography for the human X chromosome. *PLoS One* 2: e557.
37. 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: 450–455.
38. Peters H, Neubüser A, Balling R (1998) *PAX* genes and organogenesis: *PAX9* meets tooth development. *Eur J Oral Sci* 106: 38–43.
39. Hetzer-Egger C, Schorpp M, Boehm T (2000) Evolutionary conservation of gene structures of the *PAX1/9* gene family. *Biochim Biophys Acta* 1492: 517–521.
40. Santagati F, Abe K, Schmidt V, Schmitt-John T, Suzuki M, et al. (2003) Identification of cis-regulatory elements in the mouse *Pax9/Nkx2-9* genomic region: implication for evolutionary conserved synteny. *Genetics* 165: 235–242.
41. Pigliucci M (2008) Is evolvability evolved? *Nature* 9: 75–82.
42. Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–595.
43. Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics* 133: 693–709.
44. Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24: 1596–9.
45. Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* 19: 2496–2497.
46. Rogers J, Comuzzie AG (1995) When is ancient polymorphism a potential problem for molecular phylogenetics? *Am J Phys Anthropol* 98: 216–218.
47. Wooding S, Ostler C, Prasad BV, Watkins WS, Sung S, et al. (2004) Directional migration in the Hindu castes: inferences from mitochondrial, autosomal and Y-chromosomal data. *Huma Genet* 115: 221–229.
48. Liu H, Prugnolle F, Manica A, Balloux F (2006) A geographically explicit genetic model of worldwide human-settlement history. *Am J Hum Genet* 79: 230–237.
49. Fagundes NJ, Ray N, Beaumont M, Neuenschwander S, Salzano FM, et al. (2007) Statistical evaluation of alternative models of human evolution. *Proc Natl Acad Sci USA* 104: 17614–17619.
50. Cavalli-Sforza LL, Menozzi P, Piazza A (1994) *The history and geography of human genes*. Princeton: Princeton University Press.
51. Fagundes NJ, Kanitz R, Eckert R, Valls AC, Bogo MR, et al. (2008) Mitochondrial population genomics supports a single pre-Clovis origin with a coastal route for the peopling of the Americas. *Am J Hum Genet* 82: 585–592.
52. Hey J (2005) On the number of New World founders: a population genetic portrait of the peopling of the Americas. *PLoS Biol* 3: e193.
53. González-José R, Bortolini MC, Santos FR, Bonatto SL (2008) The peopling of America: craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. *Am J Phys Anthropol* 137: 175–187.
54. The Gene Ontology Consortium Gene ontology: tool for the unification of biology. *Nat Genet* May 2000;25(1): 25–9.
55. Stelzer G, Inger A, Olender T, Iny Stein T, Dalah I, et al. (2009) GeneDecks: paralog hunting and gene-set distillation with GeneCards annotation. *OMICS*, 2009 13(6).
56. Cheng Feng, Chen Wei, Richards Elliott, Deng Libin, Zeng Changqing (2009) SNP@Evolution: a hierarchical database of positive selection on the human genome. *BMC Evolutionary Biology* 2009 9: 221.

SUPPLEMENTARY INFORMATION

Table S1. Additional information's about mutations and nucleotide sequence diversity estimates of *PAT9* exon 3 and non-exon 3 bordering regions in several human populations.

Population	Region	Number of sequences	Number of sites	Number of mutations	Number of non-synonymous mutations	\pm SE
Aché	Exon 3	40	138	0	0	0
	Non-coding		452	0	0	0
	Total		590	0	0	0
Apalai	Exon 3	18	138	3	3	0.00313 \pm 0.00330
	Non-coding		452	1	0	0.00025 \pm 0.00047
	Total		590	4	3	0.00092 \pm 0.00089
Arara	Exon 3	42	138	2	2	0.00162 \pm 0.00217
	Non-coding		452	0	0	0
	Total		590	2	2	0.00038 \pm 0.00051
Bari	Exon 3	8	138	1	1	0.00311 \pm 0.00354
	Non-coding		452	0	0	0
	Total		590	1	1	0.00073 \pm 0.00083
Cayapo	Exon 3	2	138	0	0	0
	Non-coding		452	0	0	0
	Total		590	0	0	0
Galibi	Exon 3	14	138	1	1	0.00191 \pm 0.00250
	Non-coding		452	1	0	0.00058 \pm 0.00076
	Total		590	2	1	0.00089 \pm 0.00089
Guarani	Exon 3	10	138	1	1	0.00258 \pm 0.00307
	Non-coding		452	0	0	0
	Total		590	1	1	0.00060 \pm 0.00072
Jamamadi	Exon 3	10	138	1	1	0.00386 \pm 0.00398
	Non-coding		452	0	0	0
	Total		590	1	1	0.00090 \pm 0.00092
Kaingang	Exon 3	40	138	2	1	0.00199 \pm 0.00244
	Non-coding		452	0	0	0
	Total		590	2	1	0.00046 \pm 0.00057
Melkanoti	Exon 3	10	138	0	0	0
	Non-coding		452	2	0	0.00088 \pm 0.00101
	Total		590	2	0	0.00068 \pm 0.00077
Mundurucu	Exon 3	16	138	2	2	0.00175 \pm 0.00296
	Non-coding		452	0	0	0
	Total		590	2	2	0.00041 \pm 0.00055
Tenharim	Exon 3	16	138	2	2	0.00326 \pm 0.00341
	Non-coding		452	0	0	0
	Total		590	2	2	0.00076 \pm 0.00080
Wano	Exon 3	8	138	0	0	0
	Non-coding		452	0	0	0
	Total		590	0	0	0
Xikrin	Exon 3	34	138	2	2	0.00084 \pm 0.00152
	Non-coding		452	0	0	0
	Total		590	2	2	0.00020 \pm 0.00036
Yupua	Exon 3	8	138	0	0	0
	Non-coding		452	0	0	0
	Total		590	0	0	0

Table S1. Cont.

Population	Region	Number of sequences	Number of sites	Number of mutations	Number of non-synonymous mutations	$\pi \pm SE$
South Americans (Average)		276				
	Exon 3		138	9	8	0.00164 ± 0.00215
	Non-coding		452	4		0.00008 ± 0.00025
	Total		590	13		0.00045 ± 0.00055
Eskimo		88				
	Exon 3		138	2	1	0.00372 ± 0.00351
	Non-coding		220	0	0	0
	Total		590	2		0.00087 ± 0.00082
Asians (Japanese)		28				
	Exon 3		138	1	1	0.00184 ± 0.00236
	Non-coding		452	0	0	0
	Total		590	1		0.00043 ± 0.00055
Europeans (Polish/Spaniard)		30				
	Exon 3		138	1	1	0.00561 ± 0.00469
	Non-coding		452	1		0.00014 ± 0.00034
	Total		590	2		0.00143 ± 0.00116
Africans		94				
	Exon 3		138	2	1	0.00281 ± 0.000042
	Non-coding		452	2		0.00044 ± 0.000012
	Total		590	4		0.00100 ± 0.00013
All		516				
	Exon 3		138	17	13	0.00263 ± 0.000020
	Non-coding		452	7		0.00014 ± 0.00003
	Total		590	24		0.00072 ± 0.00005

Table S2. *P* values of Tajima's D, and Fu and Li's D* and F* statistics calculated under several demographic scenarios^a.

Neutrality tests		Populations and demographic expansion											
		All less South Amerindians		Africans		Europeans		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
First scenario - 6,000 years ago No=10,000	Non-coding	0.966	0.143	0.921	0.339	0.791	0.588	0.436	0.245	0.443	0.121	0.999	0.606
	Tajima's D	0.972	0.031	0.884	0.205	0.844	0.668	0.629	0.293	0.670	0.081	0.996	0.673
	Li's Test	0.987	0.038	0.887	0.199	0.826	0.595	n/c	n/c	n/c	n/c	0.999	0.748
	Fu and Li's Test	0.175	0.020	0.398	0.12	0.225	0.112	0.437	0.283	0.314	0.123	0.999	0.867
	Tajima's D	0.728	0.106	0.025	0.007	0.013	0.006	0.120	0.062	0.043	0.012	0.991	0.577
	Li's Test and Fu	0.725	0.083	0.347	0.057	0.203	0.011	0.437	0.267	0.263	0.091	0.991	0.574
Second scenario - 6,000 years ago No=1000	Non-coding	0.999	0.197	0.998	0.271	0.733	0.190	0.424	0.080	0.418	0.023	0.999	0.599
	Tajima's D	0.998	0.544	0.913	0.375	0.746	0.238	0.737	0.156	0.790	0.114	0.993	0.657
	Li's Test	0.998	0.506	0.998	0.490	0.746	0.226	n/c	n/c	n/c	n/c	0.999	0.734
	Fu and Li's Test	0.283	0.007	0.427	0.021	0.234	0.035	0.436	0.078	0.322	0.021	0.999	0.969
	Tajima's D	0.826	0.373	0.001	0.005	<0.001	<0.001	0.016	0.004	<0.001	<0.001	0.999	0.851
	Li's Test and Fu	0.827	0.284	0.349	0.007	0.003	0.001	0.440	0.085	0.231	0.017	0.999	0.851
Third scenario - 4,000 years ago No=10,000	Non-coding												
	Tajima's D				0.765	0.489	0.444	0.242	0.438	0.124	0.999	0.562	
	Li's Test				0.798	0.553	0.662	0.323	0.740	0.128	0.989	0.328	
	Fu and Li's Test				0.794	0.534	n/c	n/c	n/c	n/c	n/c	0.999	0.451
	Tajima's D				0.182	0.072	0.445	0.232	0.299	0.049	0.999	0.459	
	Li's Test and Fu				0.055	0.025	0.208	0.079	0.092	0.011	0.986	0.200	
				0.069	0.031	0.451	0.166	0.248	0.024	0.986	0.234		

Table S2. Cont.

Neutrality tests		Populations and demographic expansion											
		All less South Amerindians		Africans		Europeans		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
Fourth scenario - 37,000 years ago No=3,000	Non-coding												
	Tajima's D			0.411	0.193								
	D*			0.692	0.284								
	F*			n/c	n/c								
	Exon 3			0.439	0.223								
	Tajima's D			0.036	0.024								
Fifth scenario - 18,000 years ago No=500	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			0.340	0.049								0.999
	Tajima's D			<0.001	<0.001								0.907
Sixth scenario - 18,000 years ago No=80	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			0.343	0.016								0.999
	Tajima's D			<0.001	0.001								0.982
Fu and L's			0.295	0.015									0.982

^aStatistically significant values are shown in boldface; n/c: non computed since F* is dependent on the number of segregating sites. No.: size of the founding population. Intermediate values between 1 and 100-fold expansions were calculated but are not shown.

Table S3. List of dental development genes*.

Gene_Symbol	Aliases	Chro	Gene_start	Gene_end	Gene_size	Description
ACVR2A	ACVR2 ;ACTR-1IA ;ACTR1I ;ACTR1IA ;ameloblastin	2	148319040	148404863	85823	Activin receptor type IIA ;activin A receptor, type II ; activin A receptor, type IIA
ACVR2B	ACTR1IB ;ACTR-1IB ;ActR-1IB ;EC 2.7.11.30	3	38470794	38524948	54154	activin A receptor, type IIB ;Activin receptor type IIB ;
AMBN	ameloblastin	4	71492590	71507594	15004	Ameloblastin (enamel matrix protein) ;ameloblastin, enamel matrix protein
AMTN	RST1689 ;MGCL48133 ;UNQ689 ;MGCL48132	4	71418887	71433048	14161	Amelotin
ANKRD6	Diversin ;XIAA0957 ;diversin	6	90142897	90400124	257227	Ankyrin repeat domain 6
AQP1	AQP-CHIP ;CO ;CHIP28 ;Aquaporin-CHIP ;MGCL6324 ;AQP-1	7	30917993	30931656	13663	Channel-like integral membrane protein, 28-kDa;Colton blood group ; aquaporin 1 (channel-forming integral protein, 28kDa)
AQP3	AQP-3 ;SIL	9	33431158	33437590	6432	Gill blood group;aquaporin 3
AQP5	AQP-5 ;Aquaporin-5	12	48641546	48645227	3681	Aquaporin 5
AQP6	AQP-6 ;HKID ;Aquaporin-2-like ;AQP2L ;KID	12	48652887	48657189	4302	Aquaporin 6;Kidney-specific aquaporin ;aquaporin 2-like, kidney specific;aquaporin-6, kidney specific
AXIN2	Axin ;Conductin ;MGCL0366 ;conductin	17	60955143	60988227	33084	Axin-like protein ; axin 2 ;Axin inhibition protein 2
BARX1	BARX	9	95757390	95757429	3699	BARX homeobox 1
BGLAP	osteocalcin ;OC ;PMF1 ;BGP	1	154478629	154479736	1107	Gamma-carboxyglutamic acid-containing protein;bone gamma-carboxyglutamate (gla) protein (osteocalcin)
BMP2	BMP2A ;BMP-2 ;BMP-2A	20	6696745	6708910	12165	Bone morphogenetic protein 2
BMP4	DVR4 ;BMP-2B ;BMP2B1 ;MCOP56 ;	14	53486704	53493304	7100	Bone morphogenetic protein 4 ;bone morphogenetic protein 2B
BMPRIA	SKRS ;ACVRLK3 ;IQG23del ;CD292 ;ALK3 ; ALK3 ;	10	88506376	88684945	178569	Activin receptor-like kinase 3 ; activin A receptor, type II-like kinase 3;CD292 antigen; Serine/threonine-protein kinase receptor RS
CENPH	ICEN35 ;CENP-H ;NMF1 ;PMF1	5	68521131	68541940	20809	Interphase centromere complex protein 35; NMF1, MIND kinetochore complex component homolog
COL1A2	OIA	7	93861809	93898480	36671	Collagen I, alpha-2 polypeptide;collagen of skin, tendon and bone, alpha-2 chain; osteogenesis imperfecta type IV
CORIN	Corin ;Lrp4 ;corin ;TMPRSS10 ;PRSC ;MGCL15742 ;CRN ;	4	47290775	47534816	244041	Heart specific serine proteinase;Pro-ANP-converting enzyme ; corin, serine peptidase
CSF1	M-CSF ;M-CSF ;CSF-1 ;Laminostim ;	1	110254980	110273878	18898	Macrophage colony stimulating factor
CTNNB1	CTNNB ;FLJ25606 ;Beta-catenin ;beta-catenin ;	3	41215946	41301587	85641	Catenin (cadherin-associated protein), beta 1 (88kD)
CXCL1	PCC1 ;CFP1 ;	18	46062711	46068690	5979	DNA-binding protein with PHD finger and COXC domain
DKK1	Dkk-1 ;Dickkopf-1 ;DKK-1 ;SK ;DKK-1	10	53744047	53747423	3376	Dickkopf homolog 1 ;dickkopf related protein-1;dickkopf-1 like
DLX1	OTTHUMP0000082497 ;	2	172658454	172662648	4194	Distal-less homeo box 1
DLX2	TES1 ;TES-1	2	172672412	172675724	3312	Distal-less homeo box 2
DLX3	AH4 ;TDO	17	45427368	45427587	5219	Distal-less homeobox 3
DLX5		7	96487638	96492079	4441	Distal-less homeo box 5
DMP1	ARHR ;ARHP ;DMP-1	4	88790483	88804534	14051	Dentin matrix protein 1 ;dentin matrix acidic phosphoprotein
DMTF1	DMP1 ;DMTF ;hDMP1 ;hDMTF1 ;FLJ25188	7	86619860	86663584	43724	Cyclin-D-interacting Myb-like protein 1 ; cyclin D binding myb-like transcription factor 1 ;cyclin D-binding Myb-like protein 1
DMW2	DMT11 ;DYNII ;CMTD1B ;DYN2 ;DI-CMTB ;EC 3.6.5.5	19	10689755	10803579	113824	Dynamitin 2 ;dynamitin II ;
DSPP	DG11 ;DPP ;DTDP2 ;DFNA39 ;DSP ;DMP3	4	88748705	88757049	8344	Dentin sialoprotein;dentin phosphoryn; dentin sialophosphoprotein ;dentin phosphoryn;dentin phosphoprotein
EDAR	DL ;ED1R ;HRM1 ;ED3 ;EDA3 ;Edar ;EDA-A1R ;EDA1R	2	108877361	108972260	94899	Anhidrotic ectodysplasin receptor 1 ;ectodermal dysplasia receptor
EDARADD	ED3 ;EDA3	1	234624303	234714631	90328	ectodysplasin 1, anhidrotic receptor
EGF	Beta-urogastrone ;HOMIG4 ;URG	4	111053499	111152868	99369	Ectodysplasin A receptor associated death domain;ectodysplasin A receptor associated adapter protein;Crinkled homolog ; EDAR-associated death domain
						Epidermal growth factor (beta-urogastrone); epidermal growth factor

Table S3. List of dental development genes*.

Gene_Symbol	Gene_start	Gene_end	Gene_size	Description
ENAM	71713325	71731400	18075	Enamelin; amelogenesis imperfecta 2; hypocalcification (autosomal dominant)
ENAM	71713325	71731400	18075	Enamelin; amelogenesis imperfecta 2; hypocalcification (autosomal dominant)
ENO1	8843650	8861367	17717	Plasminogen-binding protein ,non-neural enolase
ENO3	4795131	4801148	6017	Enolase 3 ,2-phospho-D-glycerate hydro-lyase ,enolase 3, (beta, muscle)
ERBB3	54760159	54783395	23236	V-erb-b3 erythroblastic leukemia viral oncogene homolog 3 (avian); Tyrosine kinase-type cell surface receptor HER3
ERBB4	211948687	213403565	1454878	V-erb-a avian erythroblastic leukemia viral oncogene homolog-like 4
ESR1	151977826	152466099	488273	Estrogen receptor alpha;Estradiol receptor ,estrogen receptor alpha delta
EVC	5712924	5866932	154008	Ellis van Creveld syndrome ,Ellis van Creveld syndrome protein
EVC2	5544499	5761195	216696	Ellis van Creveld syndrome 2
FADD	69726917	69731144	4227	Mediator of receptor induced toxicity ,growth-inhibiting gene 3 protein,Fas-associating protein with death domain
FGF3	4347654	4359155	11501	Tumor-derived hypophosphatemia-inducing factor ; fibroblast growth factor 23 ;tumor-derived hypophosphatemia inducing factor
FGF3	69333917	69343129	9212	Fibroblast growth factor 3
FGF4	69296978	69299352	2374	Fibroblast growth factor 4 splice isoform,HST oncogene
FGF8	103519877	103525817	5940	Heparin-binding growth factor 8 ,fibroblast growth factor 8;androgen-induced growth factor
FGF9	21143875	21174187	30312	Fibroblast growth factor 9 (gila-activating factor)
FGFR1	38387813	38445509	57696	Heparin-binding growth factor receptor;soluble FGFR1 variant 2; fibroblast growth factor receptor 1
FGFR2	123227845	123357972	130127	Protein tyrosine kinase, receptor like 14; fibroblast growth factor receptor 2 ;craniofacial dysostosis 1
FST	52812352	52817680	5328	Follistatin ,Activin-binding protein ,folliculin isoform FST317
GARS	30600706	30673649	72943	Glycine tRNA ligase;Charcot-Marie-Tooth neuropathy 2D; glycyl-tRNA synthetase
GAS1	88749098	88751924	2826	Growth arrest-specific gene-1
GJA1	121796444	121812572	14128	Gap junction protein, alpha 1, 43kDa
GLI	56140201	56152312	12111	GU family zinc finger 1
GLI2	121266327	121466321	199994	GU family zinc finger 2
IBSP	88939726	88952098	12372	Bone sialoprotein II;Cell-binding sialoprotein ; integrin-binding sialoprotein ;Bone sialoprotein II
IFT88	20039208	20163576	124368	Recessive polycystic kidney disease protein Tg737 homolog
IRF6	208027885	208046102	18217	Interferon regulatory factor 6
JARID1B	199428189	199509206	81,018	Jumonj, AT rich interactive domain 1B (RBP2-like) (Lysine (K)-specific demethylase 5B)
KLK4	56101420	56105806	4386	Serine protease 17 ,kallikrein 4 (protease, enamel matrix, prostate) ,enamel matrix serine protease 1
LAMB1	107351499	107431040	79541	Laminin, beta 1; cutis laxa with marfanoid phenotype; Laminin B1 chain
LEF1	108968701	109089578	120877	Lymphoid enhancer-binding factor 1
LRP4	46834994	46940173	105179	Multiple epidermal growth factor-like domains 7; low density lipoprotein receptor-related protein 4
MGP	14926094	14930095	4001	Cell growth-inhibiting gene 36 protein ; matrix Gla protein

Table S3. List of dental development genes*.

Gene Symbol	Aliases	Chro	Gene_start	Gene_end	Gene_size	Description
MMP13	MMP-13 ,CLG3 ,MADP1 ;	11	102318934	102331672	12738	Collagenase 3,matrix metalloproteinase 13, matrix metalloproteinase 13
MMP14	MTMMP1 ,MMP-14 ,MT1MMP ,MT1-MMP ,MMP-X1 ;	14	22375633	22386643	11010	Membrane-type matrix metalloproteinase 1
MMP20	Enamelysin ,MMP-20 ,A2A2 ;	11	101952776	102001273	48497	Enamel metalloproteinase,matrix metalloproteinase 20
MMP25	MMP11 ,MMP20A ,Leukolysin ,MT6-MMP ;	16	3036683	3050728	14045	MT-MMP 6 ; matrix metalloproteinase 25; membrane-type-6 matrix metalloproteinase
MSX1	Hox-7; Hox7; STHAG1; HYD1; OFCS	4	4911293	4916564	4271	Msh homeo box 1; Msh homeobox 1-like protein; Msh homeobox homolog 1; homeobox 7
MSX2	FPP; CR52; PFM1; Hox-8; MSH; HOXB; PFM	5	174084181	174090508	6327	Msh homeobox homolog 2 (Drosophila) ; msh homeo box 2
NOG	SYNS1 ,SYM1	17	52026274	52027546	1272	Noggin ,synostoses (multiple) syndrome 1; symphalangism 1 (proximal)
NOTCH2	HN2 ,AGS2	1	120255699	120413799	158100	Notch (Drosophila) homolog 2 ; Notch homolog 2 (Drosophila)
OSR1	OOD	2	19414727	19421853	7126	Odd-skipped (Drosophila) homolog ,odd-skipped related 1
OSR2	FL90037	8	100025862	100033504	7642	Odd-skipped related 2 (Drosophila) ;
OTX2	MCOF55 ;	14	56337178	56346937	9759	Orthodenticle homeobox 2; homeobox protein OTX2,orthodenticle homolog 2 (Drosophila)
OXSR1	OSR1 ,K0AA1101 ,EC 2.7.1.11.1	3	38182030	38296979	114949	Oxidative stress-responsive 1 protein
PAX1	HUP48	20	21634297	21644620	10323	Paired domain gene HUP48; paired box gene 1
PAX9	STHAG3	14	36196333	36216763	20230	Paired domain gene 9; paired box gene 9 ; paired box 9
PITX2	solurshin; IGDS; RIEG; IGDS2; PTX2; Box1; ARP1; RIEG1; IHG2; IRID2; RGS; Otx2;	4	111758028	111777957	19929	RIEG blood-related homeobox transcription factor ,Paired-like homeodomain transcription factor 2 ; pituitary homeo box 2
PMF1	PMF-1	1	154449408	154476457	27049	Est1p-like protein B (EST1B); polyamine-modulated factor 1
PVR1	PVR1 ,PVRR1 ,SK-12 ,nectin-1;	11	119014018	119104645	90627	Herpesvirus ig-like receptor; ectodermal dysplasia 4 (Margarita island type)
RUNX2	CBF-alpha-1; PEBP2A2; CBFA1; AML3; OSF2; PEBP2A1; OSF-2; CCD1; CCD;	6	45295894	45626797	330903	SL3-3 enhancer factor 1 alpha A; Osteoblast-specific transcription factor 2
SDCI	syndecan ,SDC ,CD138 ,SYND1	2	20264039	20288675	24636	Heparan sulfate proteoglycan fibroblast growth factor receptor:syndecan proteoglycan 1; syndecan 1 ;CD138 antigen [SP,EG]
SGPPI	hSPase1 ,SPPASE1 ,Spp1 ,hSPPI ;sphingosine-1-phosphatase ,EC 3.1.3.3.	14	63220687	63264509	43822	Sphingosine-1-phosphatase 1 ; sphingosine-1-phosphate phosphatase 1
SHH	MCOPCB5 ,TPTPS ,HHG1 ;	7	155288319	155297728	9409	Sonic hedgehog homolog (Drosophila)
SLC7A2	MST153 ,DTDST ,EDM4	5	149320493	149366968	46475	Solute carrier family 26 member 2; diastrophic dysplasia sulfate transporter
SMAD1	MADRI; HMNS;	4	146622401	146699778	77377	Transforming growth factor-beta signaling protein 1; Mothers against DPP homolog 1
SMAD2	hMAD-2 ,MADH2 ;	18	43613464	43711510	98046	MAD, mothers against decapentaplegic homolog 2,Mad-related protein 2
SMAD5	Smads ,VS-1 ,hSmads;	5	135468536	135546321	77785	MAD, mothers against decapentaplegic homolog 5
SOSTDC1	USAG-1 ; Ectodin ,USAG1 ,ECTODIN ;	7	16467631	16471999	4368	Ectodermal BMP inhibitor,uterine sensitization-associated protein-1
SPP1	BNSP ;Uropontin ,Nephropontin; osteopontin ;	4	89115826	89123587	7761	Secreted phosphoprotein 1 (osteopontin, bone sialoprotein I, early T-lymphocyte activation 1)
TBX3	UMS ,XHL ,TBX3-ISO	12	113592442	113606352	13910	Ulнар mammary syndrome ;T-box transcription factor TBX3,bladder cancer related protein XHL; T-box 3
TCOF1	MF01 ,Treachle	5	149717428	149779870	62442	Treachler Collins syndrome protein [SP,EG]; Treacher Collins-Franceschetti syndrome 1 ,nucleolar trafficking phosphoprotein
TGFBI	LAP ,CED ,TGF-beta-1 ,TGFB ,DPD1 ,TGFbeta	19	46528491	46551656	23165	TGF-beta 1 protein,latency-associated peptide; transforming growth factor, beta 1

Table S3. List of dental development genes*.

Gene_Symbol	Aliases	Chro	Gene_start	Gene_end	Gene_size	Description
TGFB2	Polyargin ;G-TSF ;TGF-beta-2;TGF-beta2 ;Cetermin	1	216586491	216681596	95105	BSC-1 cell growth inhibitor ; transforming growth factor, beta 2; Glioblastoma-derived T-cell suppressor factor
TGFBI	ARVD ;TGF-beta-3 ;TGF-beta3	14	75494195	75517242	23047	Transforming growth factor, beta 3
TNFSF11	RANKL ;RANKL2 ;ODF ;OPGL ;CD254 ;	13	42043795	42080148	36353	Osteoprotegerin ligand ;receptor activator of nuclear factor kappa B ligand
TUFT1	tuftelin 1 ;tuftelin	1	151512781	151556059	43278	Tuftelin 1
VDR	Nuclear receptor subfamily 1 group I member 1 ; vitamin D3 receptor	12	48235320	48298814	63494	Vitamin D (1,25- dihydroxyvitamin D3) receptor
WDR72	WD repeat-containing protein 72 ; WD repeat domain 72 ;	15	53805938	54055075	249137	WD repeat domain 72
WNT10A	wingless-type MMTV integration site family, member 10A	2	219745085	219764303	19218	Wingless-type MMTV integration site family, member 10A
ZNF22	HR-T1 ; KROX26 ; zinc finger protein KROX15 ;	10	45495923	45500777	4854	Zinc finger protein 22 (KROX 15)

*Chro = Chromosome; Searched using the Gene Ontology database [browser AmiGO - <http://amigo.geneontology.org/cgi-bin/amigo/browser.cgi>].

Table S5. Integrated haplotype score (iHS)³ values for the top score SNPs of *PAX9*.

GENE	European (CEU)						African (YRU)						Asian (ASN)					
	SNP	Physical location	Derived allele frequency	iHS	iHS/ASN	iHS/YRU	SNP	Physical location	Derived allele frequency	iHS	iHS/ASN	iHS/CEU	SNP	Physical location	Derived allele frequency	iHS	iHS/YRU	iHS/CEU
<i>PAX9</i>	rs723204	35441383	0.083	3.531	1.0880	2.0703	rs7146383	31382048	0.300	4.252	1.000	1.000	rs1349124	40409503	0.094	-3.529	ND	-0.4104
	rs766056	35437846	0.083	3.505	1.0642	2.064	rs8005763	31296963	0.692	-4.154	ND	ND	rs7156770	34614051	0.950	3.383	1.0050	0.4925
	rs1169007	35431912	0.092	3.355	1.0893	1.6048	rs2022738	35163591	0.167	-4.147	ND	0.4155	rs8019741	34593017	0.950	3.359	ND	0.6437
	rs12883235	32753206	0.133	-3.260	ND	ND	rs7143532	31355122	0.275	4.093	ND	ND	rs8011060	34616124	0.950	3.305	0.9911	0.5050
	rs17556373	32753590	0.133	-3.254	ND	ND	rs11625294	35143319	0.183	-3.999	ND	0.1066	rs7111276	40518850	0.161	-3.221	ND	ND
	rs10483431	32751215	0.133	-3.228	ND	ND	rs3985198	40557969	0.625	3.921	2.6532	0.6225	rs10130324	39188660	0.400	-3.189	ND	ND
	rs7155975	32751670	0.133	-3.227	ND	ND	rs8020397	31265993	0.267	3.809	0.4679	1.0533	rs17111338	40564605	0.161	-3.134	ND	-0.6263
	rs17471435	32752876	0.133	-3.219	ND	ND	rs10141958	31992333	0.150	3.800	ND	ND	rs7160889	34605736	0.050	-3.076	-0.3874	ND
	rs7161508	32752268	0.133	-3.219	ND	ND	rs919521	31277641	0.233	3.705	0.6781	1.0647	rs8007575	34612510	0.050	-3.073	-0.3605	0.1132
	rs7161423	32752050	0.133	-3.219	ND	ND	rs7149151	31298293	0.175	3.701	0.6483	1.3398	rs8012747	39200076	0.400	-3.049	ND	-0.3240
	rs1028799	32754712	0.133	-3.206	ND	ND	rs4981120	31270378	0.092	3.625	0.4355	-0.0812	rs1176970	40505514	0.100	-3.027	ND	-0.7572
	rs12435308	34567833	0.200	-3.163	-1.8134	ND	rs10129541	31293590	0.158	3.596	0.5587	1.285	rs4982053	32668619	0.067	-3.009	ND	ND
	rs12886242	32750225	0.133	-3.162	ND	ND	rs1755778	35763742	0.167	3.556	1.4941	0.9993	rs17103033	34626231	0.050	-2.995	00.1856	-0.067
	rs12882375	38886532	0.158	-3.092	ND	ND	rs1169032	35319325	0.183	-3.525	ND	1.9952	rs8011549	34624884	0.050	-2.995	-0.3927	-0.0670
	rs10129933	39724743	0.183	3.083	1.0361	1.2410	rs10148038	31304269	0.725	-3.518	ND	ND	rs12100841	34641402	0.950	2.970	0.2700	-0.1126
	rs17102768	34354035	0.250	-3.078	-0.4351	-0.5906	rs8008432	31403319	0.167	3.501	ND	ND	rs4899845	40504963	0.089	-2.965	ND	-0.2751
	rs17471995	32765033	0.133	-3.077	ND	ND	rs8016527	31392362	0.300	3.490	ND	0.51253	rs7151008	40538564	0.089	-2.942	ND	-0.2109
	rs12882122	32737356	0.133	-3.076	ND	ND	rs1376835	40519817	0.308	-3.483	ND	ND	rs10137019	34817787	0.950	2.907	-0.1565	0.0782
	rs1958552	33138796	0.133	3.070	0.9891	2.4009	rs7158021	31301646	0.075	3.477	0.6483	0.3318	rs10149823	34818391	0.950	2.907	0.5080	ND
rs1958549	33134526	0.233	3.040	1.1647	2.8621	rs12433539	31401393	0.158	3.470	ND	ND	rs7218	34812814	0.950	2.905	-0.4428	-0.4388	
Total				14					6							13		
negative values				70					30							65		

³ iHS values recovered from the Haplotter (<http://plab-server.uchicago.edu/selection/>) and SNP@evolution (<http://highmapmap.big.ac.cn/>) databases.

Table S6. Sample sizes, geographic location, and linguistic information on fifteen South Amerindian, one Eskimo and one composite African populations investigated in this study

Populations	No. of individuals	Geographic location	Country/Region	Language classification ^a	References for additional information
South Amerindians					
Aché	20	23°30' - 56°30'W	Paraguay	Tupian	[3-6]
Apalaí	9	1° 20'N - 54° 40'W	Brazil/North	Cariban	[7, 8]
Arara	21	3° 30' S - 54° 10'W	Brazil/North	Arara	[9]
Bari	4	12° 50' S - 73° 5'W	Venezuela	Chibchan	[6]
Cayapo	1	10° 20' S - 53° 5'W	Brazil/North	Jéan	[10]
Guarani	5	23° 6' S - 55° 12' W	Brazil/South	Tupian	[6, 11, 12]
Galibi	7	3° 45' N - 51° 15' W	Brazil/North	Cariban	[13]
Jamamadi	5	7° 15' S - 66° 41' W	Brazil/North	Arahuan	[9]
Kaingang	20	27° 20' S - 52° 45' W	Brazil/South	Jéan	[6, 12, 14, 15]
Mekranoti	5	8° 40' S - 54° W	Brazil/North	Jéan	[10]
Mundurucu	8	6° 23' S - 59° 9' W	Brazil/North	Tupian	[9]
Tenharim	8	8° 20' S - 54° W	Brazil/North	Tupian	[13]
Warao	4	11° 50' S - 74° 5' W	Venezuela	Warao	[6]
Xikrin	17	5° 55' S - 51° W	Brazil/North	Jéan	[10]
Yucpa	4	12° 50' S - 71° 5' W	Venezuela	Chibchan	[6]
Northern Arctic					
Eskimo	44	64° N - 175° W	Russia	Eskimo-Aleut	[16-18]
Africa					
African ^b	43		Democratic Republic of Congo, Cameroon, Ivory Coast	Niger-Congo	[19, 20]

^a According to Greenberg [1] and Campbell [2].

^b A composite sample from Bantu and non-Bantu speakers living in the Democratic Republic of Congo, Cameroon, and Ivory Coast, while four additional individuals were ascertained in Paris, France.

References Supplementary Material

1. Greenberg JH (1963) *The language of Africa*. Bloomington: Indiana University.
2. Campbell L (1997) *American Indian languages: the historical linguistics of Native America*, New York: Oxford University Press.
3. Brown SM, Gajdusek DC, Leyshon WC, Steinberg AG, Brown KS, et al. (1974) Genetic studies in Paraguay: blood group, red cell, and serum genetic patterns of the Guayaki and Ayore Indians, Mennonite settlers, and seven other Indian tribes of the Paraguayan Chaco. *Am J Phys Anthropol* 41: 317–343.
5. Hill K, Hurtado M (1996) *Ache life history: the ecology and demography of a foraging people*. New York: Aldine de Gruyter.
6. Bortolini MC, Salzano FM, Thomas MG, Stuart S, Nasanen SP, et al. (2003) Y-chromosome evidence for differing ancient demographic histories in the Americas. *Am J Hum Genet* 73: 524–539.
7. Salzano FM, Black FL, Callegari-Jacques SM, Santos SE, Weimer TA, et al. (1988) Genetic variation within a linguistic group: Apalai-Wayana and other Carib tribes. *Am J Phys Anthropol* 75: 347–356.
8. Mazières S, Guitard E, Crubézy E, Dugoujon JM, Bortolini MC, et al (2008) Uniparental. mtDNA, Y-chromosome. polymorphisms in French Guiana and two related populations—implications for the region's colonization. *Ann Hum Genet* 72: 145–156.
9. Salzano FM, Black FL, Callegari-Jacques SM, Santos SE, Weimer TA, et al. (1991) Blood genetic systems in four Amazonian tribes. *Am J Phys Anthropol*. 85: 51–60.
10. Salzano FM, Neel JV, Weitkamp LR, Woodall JP (1972) Serum proteins, hemoglobins and erythrocyte enzymes of Brazilian Cayapo Indians. *Hum Biol* 44: 443–458.
11. Salzano, FM, Callegari-Jacques SM, Weimer TA, Franco MHL, Hutz MH, et al (1997) Electrophoretic protein polymorphisms in Kaingang and Guarani Indians of southern Brazil. *Am J Hum Biol* 9: 505–512.
12. Marrero AR, Silva-Junior WA, Bravi CM, Hutz MH, Petzl-Erler ML, et al. (2007) Demographic and evolutionary trajectories of the Guarani and Kaingang natives of Brazil. *Am J Phys Anthropol* 132: 301–310.
13. Santos SEB, Ribeiro-Dos-Santos AK, Guerreiro JF, Santos EJ, Weimer TA, et al. (1998) New protein genetic studies in six Amazonian Indian populations. *Ann Hum Biol* 25: 505–522.
14. Salzano FM, Callegari Jacques SM, Franco MH, Hutz MH, Weimer TA, et al. 1980. The Caingang revisited: blood genetics and anthropometry. *Am J Phys Anthropol* 53: 513–524.
15. Belich MP, Madrigal JA, Hildebrand WH, Zemmour J, Williams RC et al. (1992) Unusual HLA-B alleles in two tribes of Brazilian Indians. *Nature* 357: 326–329.
16. Erdesz S, Shubin SV, Shoch BP, Krylov My, Mylov NM et al. (1994) Spondyloarthropathies in circumpolar populations of Chukotka. Eskimos and Chukchi): epidemiology and clinical characteristics. *J Rheumatol* 21: 1101–1104.
17. Krylov M, Erdesz S, Alexeeva L, Benevolenskaya L, Arnett FC, et al. (1995) HLA class II and HLA-B27 oligotyping in two Siberian native population groups. *Tissue Antigens* 46: 382–386.
18. Krylov M, Erdesz S, Alekseeva LI, Benevolenskaia LI (1995) DNA typing of HLA class II genes among the aboriginal inhabitants of Chukotka. *Genetika* 31: 852–858.
19. Silva WA Jr, Bortolini MC, Meyer D, Salzano FM, Elion J, et al. (1999) Genetic diversity of two African and sixteen South American populations determined on the basis of six hypervariable loci. *Am J Phys Anthropol* 109: 425–437.
20. Silva WA Jr, Bortolini MC, Schneider MPC, Marrero AR, Elion J, et al. (2006) mtDNA haplogroups analysis of Black Brazilian and sub-Saharan populations: implications for the Atlantic slave trade. *Hum Biol* 78: 29–41.

***CAPÍTULO 5 - RESULTADOS- “EVOLUTIONARY HISTORY OF
VERTEBRATE PAX GENES –DYNAMICS OF CHANGE IN A
COMPLEX GENE FAMILY”***

**EVOLUTIONARY HISTORY OF VERTEBRATE *PAX* GENES—DYNAMICS OF CHANGE IN
A COMPLEX GENE FAMILY**

Manuscrito em preparação

BMC Evolutionary Biology

Vanessa Rodrigues Paixão-Côrtes, Francisco Mauro Salzano & Maria Cátira Bortolini*

Programa de Pós-Graduação em Genética e Biologia Molecular, Instituto de
Biociências, Universidade Federal do Rio Grande do Sul, Av. Bento Gonçalves 9500,
Porto Alegre, RS 91501-970, Brazil

Running head: Molecular evolution of *PAX genes*

***Corresponding author:**

Maria Cátira Bortolini, PhD

Universidade Federal do Rio Grande do Sul

Departamento de Genética - PPGBM

Av. Bento Gonçalves 9500 - Prédio 43.323

Porto Alegre, RS - 91501-970, Brazil

E-mail: maria.bortolini@ufrgs.br

Phone: +55 (51) 3308.6746

ABSTRACT

Background

Paired box (*PAX*) genes are transcription factors which play important roles in embryonic development. The *PAX* gene family occurs in animals only, but within them shows a wide distribution. Among vertebrates, its nine genes seem to be the product of a whole duplication of an original set of four, followed by an additional partial duplication. Despite some previous studies, no comprehensive survey of these genes in the entire taxonomic unit has been attempted.

Results

We studied the sequences of 51 jawed vertebrates, one lamprey and one tunicate. A detailed comparison was made among them which showed restricted variation. Notable was the absence of *PAX4* and *PAX8* among reptiles and birds, but the nine genes are present in all 39 mammals genomes investigated. A search for selection implicated the all genes are being subjected to purifying selection, with a possible constraint relaxation in *PAX4*, *PAX7* and *PAX8*.

Conclusions

The results are in accordance with what is expected from a transcription factor. The family duplications made possible, through the maintenance of ancestral functions, the two gene deletions after the split of the mammal from the reptiles/bird lineages, which possibly occurred 312 to 330 million years ago.

Keywords: *PAX*, molecular evolution, selection, vertebrates, birds

BACKGROUND

Paired box (*PAX*) genes are transcription factors which play key roles in many aspects of embryonic development and organogenesis, working as genetic switches for cell lineage decisions [1–3]. The *PAX* family is specific to the animal lineage, but the evolutionary history of their genes continues uncertain. An unique *PAX* gene (*TriPaxB*) has been isolated from *Trichoplax adhaerens* (Placozoa), the morphologically simplest of all non-parasitic multicellular metazoan animals. The *TriPaxB* protein contains the characteristic DNA-binding domain of the *PAX* family, the paired domain (PD) with 128 amino acids, an octapeptide motif (OC) and a paired-type homeobox DNA-binding domain (HD)[1], [4].

Four *PAX* genes are found in the lancelet or amphioxus and tunicates, basal chordates [5–7], these genes apparently corresponding to progenitor types of the four subfamilies, as inferred by sequence similarity and structural domains. Afterwards a duplication event took place (leading to eight genes) followed by another partial duplication, giving a total of nine genes. Nine paralogous genes are found in humans, mice and fishes (*PAX1* & *PAX9*; *PAX2*, *PAX5* & *PAX8*; *PAX3* & *PAX7*; *PAX4* & *PAX6* - [5], [8]). *PAX* genes in vertebrates, as well as in other chordates, are expressed during development with significant stage/tissue-specific variations (see Supplementary Table 1). Remarkably, in both adult tissues (in vertebrates) and the corresponding differential stage (in lancelet), each daughter gene in the vertebrate keep partial roles of their parental gene in the ancestor *PAX*.

Various novel functions, however, were developed through the emergence of new regulatory networks by duplication, divergence and subsequent gain of function, during the evolutionary history of vertebrates [5], [9–11]. A variety of organisms have been analyzed for *PAX* occurrence ([6–8], [12], [13], but a wide-ranging survey of the vertebrate *PAX* genes has not yet been done.

The aim of the present study was to address questions regarding the occurrence and evolution of the *PAX* family in vertebrates. We used genomic sequences to evaluate: (1) the presence/absence; and (2) the evolutionary rates and properties of the *PAX* genes in 53 publicly available chordate genomes. In this way we are contributing to unravel mechanisms of change in complex gene families.

RESULTS AND DISCUSSION

Search and Identification of vertebrate *PAX* genes

We characterized vertebrate *PAX* genes by searching for available sequences in the Ensembl animal genome databases, including 51 jawed vertebrates, 1 lamprey and 1 tunicate; see details in Figure 1 and Table 1. In the sea squirt (*Ciona Intestinalis*), the most basal animal in our study, we recovered 4 *PAX* genes, which can be considered equivalent to the 4 subfamily ancestor vertebrate *PAX* types [5–7].

Pretromyson marinus (lamprey), on the other hand, is a jawless vertebrate, belonging to the Agnatha, the sister group of gnathostome vertebrates. Its phylogenetic position makes the species a model to study vertebrate ancestral characteristics. We only recovered 2 *PAX* in the lamprey genome (reported as Pax1/9 and Pax1/9b) in the first search. In a second search, however, using BLAT/BLAST, we found 2 segments of DNA of 158 and 144 bp, with 88 % identity with the *PAX3/PAX7* genes. Interestingly, *in vitro* studies identified in this species 2 members of the subfamily *PAX1/9* [14], *PAX7* and *PAX2* [15], *PAX6* [16], and *PAX3/PAX7* genes [17]. This *in vitro* information, as well as our genome data suggested that the lamprey has the four ancestor *PAX* subtype genes, plus a second copy of the *PAX1/9* group.

The difference between the number of *PAX* genes between the sea squirt (4 *PAX*) and lamprey (5 *PAX*) basal chordates and jawed vertebrates (9 *PAX*) can be associated to the emergence of adaptive novelties, at the tunicate/vertebrate and agnathan/gnathostome transition.

For the jawed vertebrates, we found all nine *PAX* genes in 5 of 6 ray-finned bony fishes (Actinopterygii; Table 2). Additionally, some *PAX* genes were found duplicated in all 5 species (Figure 2), probably as a consequence to the whole genome duplication in the early evolution of teleost fishes, which occurred approximately 320-350 million years ago [18–20]. The only exception, the Medaka fish, *Oryzias latipes*, has only eight *PAX* genes. We could not identify a *PAX9* orthologous in both BIOMART and BLAT/BLAST searches. However, this result should be considered with caution, since despite the fact that the Medaka genome has an excellent coverage (9x; Table 1), this

PAX9 loss was not observed in any other fish species analyzed. The coelacanth, *Latimeria chalumnae*, the lobe-finned fish, presents all nine *PAX* genes, confirming the idea that the lineage that gave rise to the tetrapods has all members of the *PAX* family. The frog, *Xenopus tropicalis*, has also nine *PAX* genes.

The analysis of the sauropsid genomes (reptiles and birds[21]) showed a surprising finding: this taxon seems to have only 7 *PAX* members (Table 3). The 2 missing genes (*PAX4* and 8) were detected in multiple species, and 3 of the 4 genomes considered here present high coverage (higher than 6x). This fact implies that *PAX4* and *PAX8*, which were a part of the ancestral vertebrate genome, were lost after the split of mammals (Synapsida) and reptiles/bird (Sauropsida) which occurred between 312 to 330 million years ago[22].

In a genome-wide comparison of gene family sizes between the chicken and five mammalian genomes, several chicken gene families had significantly fewer numbers of genes than their mammalian equivalents [23]. The primary searches failed to find, identify or predict, orthologous to Chicken *PAX1* in the chicken genome (WASHUC2), contrary to the abundant literature, which described patterns of expression of this gene in chicken embryos [24–28]. But in a second exploration in the Pre!Ensembl (release 66 - April 2012), Galgal4 assembly, with the BLAT tool, it was possible to determine an unannotated ~2.8kb long *PAX1* ortholog sequence (Table 3), including 4 probable exons, from which a 352 amino acid-long protein sequence could be predicted (97% identical to the turkey *PAX1*).The nucleotide and translated amino acid

sequences are in Supplementary Material 1. This result reinforces the general bird pattern, where only *PAX 4* and *PAX8* are missing.

Other results illustrate this particular evolutionary molecular pattern found in birds. A classic example is that birds have lost several MHC and immunoglobulin genes in contrast to most other vertebrates [29]. Thus, bird evolution involved substantial loss of ancestral duplicates. The loss of PAX genes observed here is an additional evidence that support the hypothesis that the reduction of the genome size in birds when compared to other tetrapods is an important adaptation and that it can be correlated with the evolution of flight [23].

Overall, the searches showed that all nine *PAX* genes seem to be present in the 39 mammal genomes. Some exceptions to this general pattern were found, but seem to be a consequence of the low coverage of the genome in question (i.e. sloth 2x coverage), rather than a real gene loss.

Comparative analysis of the vertebrate *PAX* proteins

Pairwise amino acid distances for each *PAX* subfamily member with its probable ancestor (proto-*PAX* type from *Ciona Intestinalis*) were calculated (Table 4 and Figure 3). The distances of each gene with its probable ancestral were also compared between subfamilies. For instance, the distance values of the human *PAX1* and *PAX9* genes to proto-*PAX 1/9* type from *Ciona Intestinalis* are 0.447 and 0.455, respectively, which are not statistically significant. Considering all comparisons (52 Chordate species and the nine *PAX* genes) some specific and general patterns emerged: no significant

average values were observed when the *PAX1* and *PAX9*, as well as within *PAX3* and *PAX7*, were compared ($p= 0.069$ and $p= 0.704$ respectively). On the other hand, when *PAX2*, *PAX5* and *PAX8* were compared with the proto-*PAX 2/5/8* type from *Ciona Intestinalis*, some instigating patterns emerged. The distance of *PAX8* with its ancestral was significantly different ($p=0.001$), when compared to the other subfamily members, while for *PAX2* and *PAX5* the differences are not significant ($p=0.091$). This situation raises the hypothesis that *PAX8* could be the latest gene to appear by duplication in this group. Similarly, *PAX4* is more dissimilar to the ancestral (*PAX4/6*) than *PAX6* ($p=0.001$). An alternative suggestion would be that since *PAX4* and *PAX8* would have higher rates of non-synonymous substitution, these genes may have evolved more rapidly than the other members of these transcription factor family. We consider this idea below.

Selection on *PAX* genes

Our molecular evolution analyses (Table 5) showed that purifying selection has acted in the majority of vertebrate *PAX* evolutionary history (ω s always less than one), consistent with the idea that development genes are conserved in the metazoa [30]. The clade model D test [31] used to infer differences in the dN/dS ($\omega = \omega$) ratio between the four *PAX* subfamilies performed significantly better ($P<0.001$ for all comparisons; Table 5) than the neutral model M1a. This result suggests that the *PAX* genes present distinct molecular evolutionary patterns. Although neither ω estimate values is greater than 1, a possible relaxation of the selective constraint can be assumed when the subfamilies are compared: In 38 % of the sites the *PAX4* ω value is

sixteen times greater than the *PAX6* ω number. Additionally, in 37% and 31% of them, the *PAX8* ω and *PAX7* ω are about two and four times greater than the *PAX2/PAX5* and *PAX3* ω s, respectively. The possible inference is that these genes (*PAX4*, *PAX8* and *PAX7*) have evolved more rapidly than the other members of this transcription factor family.

Conclusion

The *PAX* master gene family, as assessed by the ω values, shows little variation at the sequence level when 53 species of chordates are compared. However, our data indicates potential group-specific changes in functional constraint. These changes could be the result of the duplication and/or subsequent relaxation of selection events. The main finding seems to be that in some vertebrate lineages (reptiles and birds) a loss of two *PAX* genes (*PAX4* and *PAX8*) took place. Since generally *PAX* genes preserved most of their ancestral functions, the preserved genes could substitute those lost by deletion.

METHODS

Genomic Search

Nucleotide and amino acid sequences, for all *PAX* genes on available vertebrate data were obtained using Biomart (Ensembl v66- <http://www.ensembl.org/biomart/martview/> [32], [33]) with the Protein Domains/ Limit filter to genes of this family or domain [InterPro (ID): IPR001523 for filtering sequences having the paired box domain]. One second approach was the one-to-one ortholog gene maps, also obtained from Biomart. The third was BLAST/BLAT searches in the Ensembl (<http://www.ensembl.org/Multi/blastview>) and UCSC databases (<http://genome.ucsc.edu/>), to find possible unannotated orthologues. Additional search was made in the Pre!Ensembl database (<http://pre.ensembl.org/index.html>) for the new release of the chicken genome (Galgal4). For genes encoding multiple transcripts, the transcript with the longest genomic transcribed length was used. Possible gene losses were accepted only when: (1) were observed in multiple species; and (2) high coverage genome assemblies were available.

Variation in the *PAX* family

Mega version 5.0 [34] was used to: align the amino acid sequences, for higher accuracy using the MUSCLE program [35]; The same software was employed to evaluate the variation in the *PAX* groups using the Pairwise Distance which compared the members of each subfamily with its sea squirt probable ancestor. The SPSS16

software was used to calculate the significance of the differences between the *PAX1/9*, *PAX2/5/8*, *PAX3/7* and *PAX4/6* paralogous sequences using t-test pairs.

Tests for selection

Patterns of selection and rates of evolutionary changes in the *PAX* family were evaluated by performing standard tests[36–38]. We used the phylogeny-based maximum likelihood analysis of ω (dN/dS) as implemented in the CODEML program of the PAML 4.4 package to statistically test for positive selection and/or relaxation of functional constraints. The heterogeneity of evolutionary rates among paralogous groups was tested by using the clade models [31] of the CODEML in PAML4.4. Branches on the phylogeny were *a priori* divided into two clades, and a likelihood ratio test (LRT) was used to evaluate divergences in selective pressures between them, indicated by different ω ratios. We employed clade model type D that assumes 2 site classes, which was compared with the neutral model M1a by a LRT test with 2 degrees of freedom. Such model is primarily used to detect positive selection, but our goal here was also to evaluate acceleration along the evolutionary history through direct inferences of dN/dS differences.

Empirical Bayes approaches implemented in CODEML were used to infer which sites of the *PAX* sequences may have evolved under positive selection. To determine sites under selection, the naive-empirical Bayes (NEB) test was employed. Positive selection was inferred if the posterior probability of $\omega > 1$ for a site was higher than

0.95. The unrooted tree input file for the PAML4.4 analyses was a phylogenetic tree provided by Ensembl, and edited with PhyloWidget for the 53 species of this study.

AUTHORS' CONTRIBUTIONS

MCB and VRP-C conceived and designed the experiments. VRP-C carried out bioinformatics work. MCB, FMS and VRP-C analysed the data. All authors were involved in the experimental design and approved the final manuscript.

ACKNOWLEDGEMENTS

This research was financially supported by the Coordenadoria de Aperfeiçoamento de Pessoal de Nível Superior(CAPES), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) and Fundação à Amparo a Pesquisa do Estado do Rio Grande do Sul (FAPERGS). VRP-C received a doctoral fellowship from CNPq.

REFERENCES

1. Chi N , Epstein JA: **Getting your PAX straight: Pax proteins in development and disease.** *Trends Genet* 2002, **1**:41–47.
2. Dressler GR: **Patterning and early cell lineage decisions in the developing kidney: the role of PAX genes.** *Pediatr Nephrol* 2011, **26**: 1387-1394.
3. Lang D, Powell SK, Plummer RS, Young KP, Ruggeri BA: **PAX genes: roles in development, pathophysiology, and cancer.** *Biochem Pharmacol* 2007, **73**:1–14.
4. Hadrys T, DeSalle R, Sagasser S, Fischer N, Schierwater B: **The Trichoplax PAXB gene: a putative Proto-PaxA/B/C gene predating the origin of nerve and sensory cells.** *Mol Bio Evol* 2005. **22**:1569–1578.
5. Chen L, Zhang Q, Wang W, Wang Y: **Spatiotemporal expression of PAX genes in amphioxus: insights into PAX-related organogenesis and evolution.** *Sci China Life Sci* 2010, **53**:1031–1040.
6. Mazet F, Hutt JA, Millard J, Shimeld SA: **PAX gene expression in the developing central nervous system of *Ciona intestinalis*.** *Gene Expr Patt* 2003, **3**: 743–745.
7. Vorobyov E, Horst J: **Getting the proto-PAX by the tail.** *J Mol Evol* 2006,**63**: 153–164.
8. Suga H, Tschopp P, Graziussi DF, Stierwald M, Schmid V, Gehring WJ: **Flexibly deployed PAX genes in eye development at the early evolution of animals demonstrated by studies on a hydrozoan jellyfish.** *Proc Natl Acad Sci USA* 2010, **107**:14263–14268.
9. Boyle AP, Song L, Lee B-K, London D, Keefe D, Birney D, Lyer VR, Crawford GE, Furey TS: **High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells.** *Genome Res* 2010, **21**: 456-464
10. Lynch M, Conery JS: **The evolutionary fate and consequences of duplicate genes.** *Science* 2000, **290**: 1151–1155.
11. Lynch VJ, Wagner GP: **Resurrecting the role of transcription factor change in developmental evolution.** *Evolution* 2008, **62**: 2131–2154.
12. Balczarek KA, Lai ZC, Kumar S: **Evolution of functional diversification of the paired box (PAX) DNA-binding domains.** *Mol Bio Evol* 1997, **14**: 829–842.
13. Catmull J, Hayward DC, McIntyre NE, Reece-Hoyes JS, Mastro R, Callaerts P, Ball EE, Miller DJ: **Pax-6 origins—implications from the structure of two coral PAX genes.** *Dev Genes Evol* 1998, **208**: 352–356.
14. Ogasawara M, Shigetani S, Hirano S, Satoh N, Kuratani S: **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* 2000, **223**: 399–410.**
15. McCauley DW , Bronner-Fraser M: **Conservation of PAX gene expression in ectodermal placodes of the lamprey.** *Gene* 2002, **287**: 129-139.
 16. Derobert Y, Baratte B, Lepage M, Mazan S:**PAX6 expression patterns in Lampetra fluviatilis and Scyliorhinus canicula embryos suggest highly conserved roles in the early regionalization of the vertebrate brain.** *Brain Res Bull* 2002, **57**: 277-280.
 17. Osorio J, Mazan S, Rétaux S: **Organisation of the lamprey (Lampetra fluviatilis) embryonic brain: Insights from LIM-homeodomain, PAX and hedgehog genes,** *Dev Biol* 2005, **288**: 100-112.
 18. Braasch I, Volff JN, Schartl M: **The evolution of teleost pigmentation and the fish-specific genome duplication.** *J Fish Biol* 2008, **73**: 1891-1918.
 19. Laisney JA, Braasch I, Walter RB, Meierjohann S, Schartl M: **Lineage-specific co-evolution of the Egf receptor/ligand signaling system.** *BMC Evol Biol* 2010, **10**: 27.
 20. Taylor JS, Raes J: **Duplication and divergence: the evolution of new genes and old.** *Annu Rev Genet* 2004, **38**: 615–643.
 21. Modesto SP, Anderson JS, **The phylogenetic definition of Reptilia.** *Syst Biol* 2004, **53**: 815 -821.
 22. Benton MJ, Donoghue PCJ: **Paleontological evidence to date the tree of life.** *Mol Biol Evol* 2007, **24**: 26 -53.
 23. Hughes AL, Friedman R: **Genome size reduction in the chicken has involved massive loss of ancestral protein-coding genes.** *Mol Biol Evol* 2008, **25**: 2681–2688.
 24. Balling R, Neubuser A, Christ B: **PAX genes and sclerotome development.** *Semin Cell Dev Biol* 1996, **7**: 129–136.
 25. Christ B, Huang R, Scaal M: **Formation and differentiation of the avian sclerotome.** *Anat Embryol* 2004, **208**: 333–350.
 26. Malashichev Y, Christ B, Pröls F: **Avian pelvis originates from lateral plate mesoderm and its development requires signals from both ectoderm and paraxial mesoderm.** *Cell Tissue Res* 2008, **331**: 595–604.
 27. Peters H, Doll U, Niessing J: **Differential expression of the chicken PAX-1 and PAX-9 gene: in situ hybridization and immunohistochemical analysis.** *Dev Dyn* 1995, **203**: 1–16.
 28. Wang B, He L, Eehalt F, Geetha-Loganathan P, Nimmagadda S, Christ B, Scaal M, and Huang R: **The formation of the avian scapula blade takes place in the hypaxial domain of the somites and requires somatopleure-derived BMP signals.** *Dev Biol* 2005, **287**: 11–18.
 29. Hughes CR, Miles S, Walbroehl JM: **Support for the minimal essential MHC hypothesis: a parrot with a single, highly polymorphic MHC class II B gene.**

- Immunogenetics* 2008, **60**: 219–231.
30. Bates MD, Wells JM, Venkatesh B: **Comparative genomics of the *Hlx* homeobox gene and protein: conservation of structure and expression from fish to mammals.** *Gene* 2005, **352**: 45–56.
 31. Bielawski JP, Yang Z: **A maximum likelihood method for detecting functional divergence at individual codon sites, with application to gene family evolution.** *J Mol Evol* 2004, **59**: 121–132.
 32. Durinck S, Huber W: **The biomaRt user's guide.** *Database* 2011, **1**: 1–22.
 33. Haider S, Ballester B, Smedley D, Zhang J, Rice P, Kasprzyk A: **BioMart Central Portal—unified access to biological data,** *Nucleic Acids Res* 2009, **37**:23 -27.
 34. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S: **MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.** *Mol Bio Evol* 2011, **28**: 2731 -2739.
 35. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32**: 1792 -1797.
 36. Nielsen R, Yang Z: **Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene.** *Genetics* 1998, **148**: 929–936.
 37. Yang Z, Bielawski J: **Statistical methods for detecting molecular adaptation.** *Trends Ecol Evol* 2000, **15**: 496–503.
 38. Yang Z, Nielsen R: **Mutation-selection models of codon substitution and their use to estimate selective strengths on codon usage.** *Mol Bio Evol* 2008, **25**: 568–579.

FIGURE LEGENDS

Figure 1. Phylogenetic tree of the 53 chordate species considered in this study, provided by Ensembl, and edited with PhyloWidget.

Figure 2. Schematic view for the presence of the *PAX* family members in vertebrates.

Figure 3. Pairwise *PAX* changes, observed in different vertebrate taxa, when compared to the sea squirt ancestor.

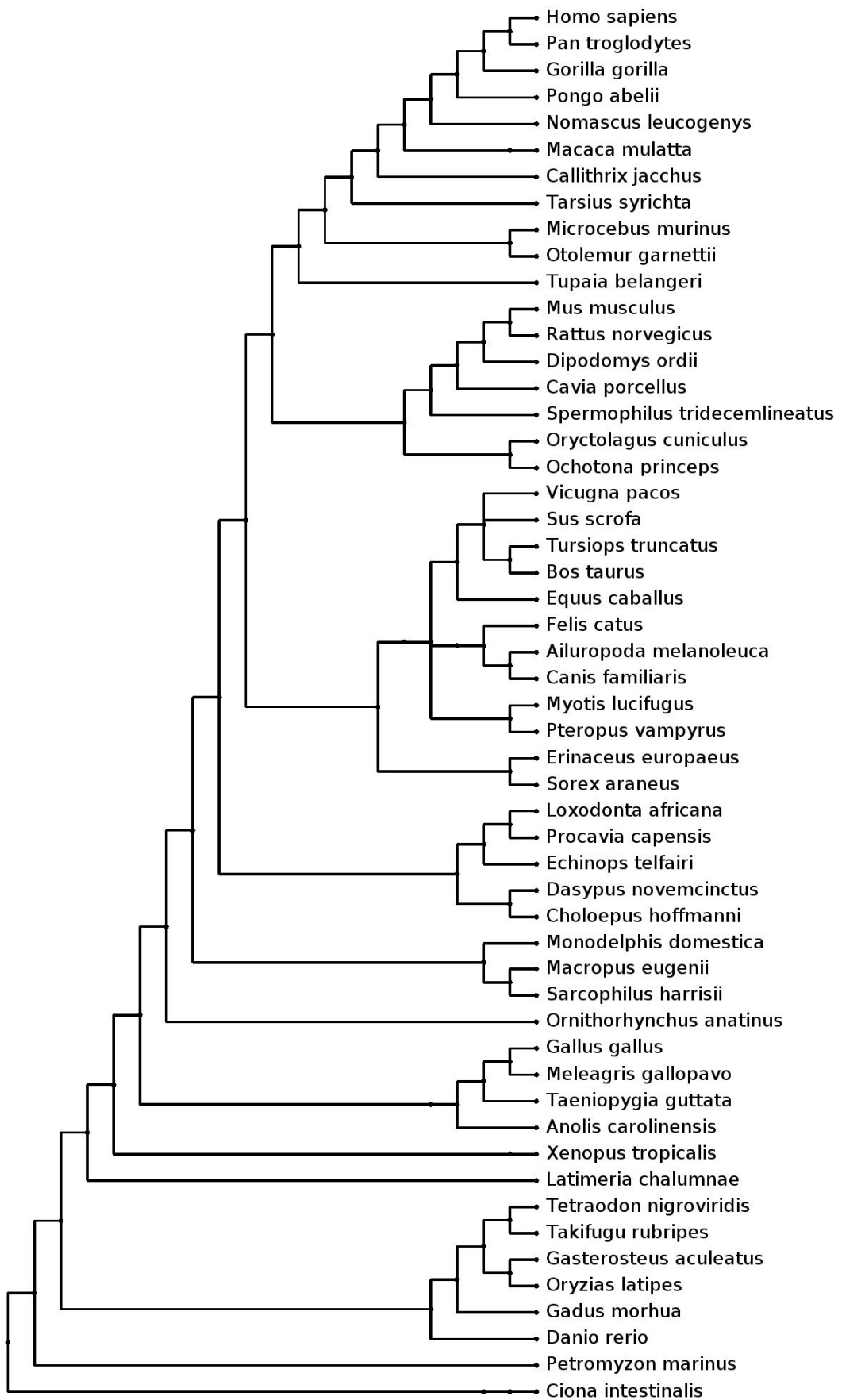


Figure 1

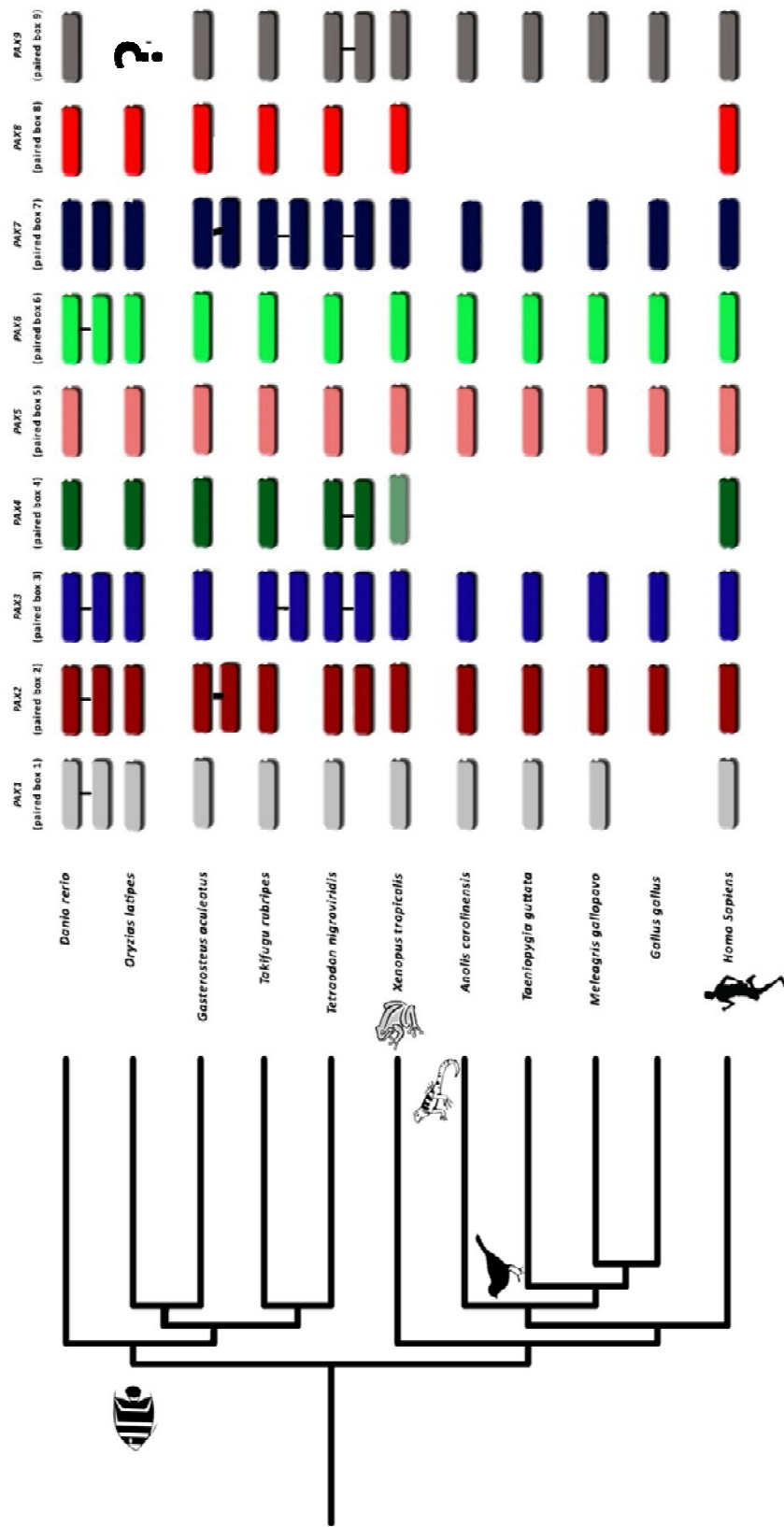


Figure 2

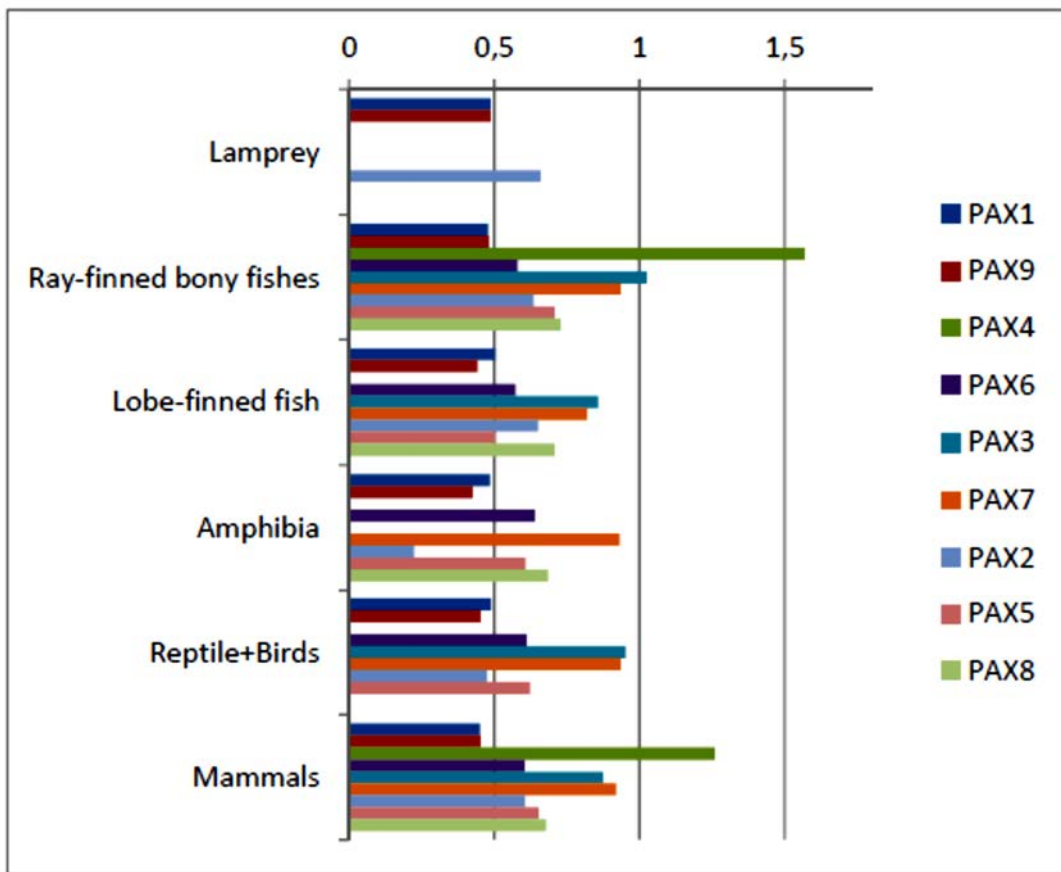


Figure 3

Table 1. Species analyzed in this study

SPECIES	COMMON NAME	GENOME ASSEMBLY	CLASS	ORDER	FAMILY	SIZE (GB) ¹	COVERAGE
<i>Ailuropoda melanoleuca</i>	Panda	ailMel1	Mammalia	Carnivora	Ursidae	2.9	1X
<i>Anolis carolinensis</i>	Lizard	AnoCar2.0	Reptilia	Squamata	Iguanidae	~2.2	7x
<i>Bos taurus</i>	Cow	UMD3.1	Mammalia	Cetartiodactyla	Bovidae	~3.5	7x
<i>Callithrix jacchus</i>	Marmoset	C_jacchus3.2.1	Mammalia	Primates	Cebidae	3.4	6X
<i>Canis familiaris</i>	Dog	BROADD2	Mammalia	Carnivora	Canidae	2.8	7x
<i>Cavia porcellus</i>	Guinea Pig	cavPor3	Mammalia	Rodentia	Caviidae	~3.8	6.79X
<i>Choloepus hoffmanni</i>	Sloth	choHof1	Mammalia	Xenarthra	Megalonychidae	4.23*	2x
<i>Ciona intestinalis</i>	SeaSquirt*	KH	Ascidiacea	Phlebobranchia	Cionidae	0.20	8.2X
<i>Danio rerio</i>	Zebrafish	Zv9	Actinopterygii	Cypriniformes	Cyprinidae	1.8	Deep
<i>Dasyopus novemcinctus</i>	Armadillo	dasNov2	Mammalia	Xenarthra	Dasyopodidae	5.3	2x
<i>Dipodomys ordii</i>	Kangaroo rat	dipOrd1	Mammalia	Rodentia	Heteromyidae	~4.4	2x
<i>Echinops telfairi</i>	Tenrec	TENREC	Mammalia	Afrosoricida	Tenrecidae	~5.03*	2x
<i>Equus caballus</i>	Horse	EquCab2	Mammalia	Perissodactyla	Equidae	~3.2	7x
<i>Erinaceus europaeus</i>	Hedgehog	HEDGEHOG	Mammalia	Eulipotyphla	Erinaceidae	3.6	2x
<i>Felis catus</i>	Cat	CAT	Mammalia	Carnivora	Felidae	~3.0	2x
<i>Gadus morhua</i>	Cod	gadMor1	Actinopterygii	Gadiformes	Gadidae	~0.6	x25
<i>Gallus gallus</i>	Chicken	WASHUC2 & Galgal4	Aves	Galliformes	Phasianidae	1.3	High
<i>Gasterosteus aculeatus</i>	Stickleback	BROADS1	Actinopterygii	Gasterosteiformes	Gasterosteidae	~0.64	9x
<i>Gorilla gorilla</i>	Gorilla	gorGor3.1	Mammalia	Primates	Hominidae	~3.5	7x
<i>Homo sapiens</i>	Human	GRCh37	Mammalia	Primates	Hominidae	3.5	Deep
<i>Latimeria chalumnae</i>	Coelacanth	LatCha1	Sarcopterygii	Coelacanthiformes	Coelacanthidae	2.7-6.4*	2x
<i>Loxodonta africana</i>	Elephant	loxAfr3	Mammalia	Proboscidea	Elephantidae	~4.2	7x
<i>Macaca mulatta</i>	Macaque	MMUL_1	Mammalia	Primates	Cercopithecidae	~3.3	2x
<i>Macropus eugenii</i>	Wallaby	Meug_1.0	Mammalia	Diprotodontia	Macropodidae	~4.2*	2x
<i>Meleagris gallopavo</i>	Turkey	UMD2	Aves	Galliformes	Phasianidae	~1.4	30x
<i>Microcebus murinus</i>	Mouse Lemur	micMur1	Mammalia	Primates	Cheirogaleidae	3.1	2x
<i>Monodelphis domestica</i>	Opossum	BROADO5	Mammalia	Didelphimorphia	Didelphidae	~4.6*	7,33x
<i>Mus musculus</i>	Mouse	NCBIM37	Mammalia	Rodentia	Muridae	~3.2	2x
<i>Myotis lucifugus</i>	Microbat	Myoluc2.0	Mammalia	Chiroptera	Vespertilionidae	1.9	2x
<i>Nomascus leucogenys</i>	Gibbon	Nleu1.0	Mammalia	Primates	Hylobatidae	~3.5*	5.6x
<i>Ochotona princeps</i>	Pika	pika	Mammalia	Lagomorpha	Ochotonidae	~3.2*	2x
<i>Ornithorhynchus anatinus</i>	Platypus	OANA5	Mammalia	Monotremata	Ornithorhynchidae	3	6x

Table 1. Cont.

SPECIES	COMMON NAME	GENOME ASSEMBLY	CLASS	ORDER	FAMILY	SIZE (GB) ¹	COVERAGE
<i>Oryctolagus cuniculus</i>	Rabbit	oryCun2	Mammalia	Lagomorpha	Leporidae	~3.1	7x
<i>Oryzias latipes</i>	Medaka	MEDAKA1	Actinopterygii	Beloniformes	Adrianichthyidae	1	9x
<i>Otolemur garnettii</i>	Bushbaby	OtoGar3	Mammalia	Primates	Galagidae	~3.5*	2x
<i>Pan troglodytes</i>	Chimpanzee	CHIMP2.1.4	Mammalia	Primates	Hominidae	~3.5	7x
<i>Petromyzon marinus</i>	Lamprey	Pmarinus_7.0	Cephalaspodomorphi	Petromyzontiformes	Petromyzontidae	2,1	5x
<i>Pongo abelii</i>	Orangutan	PPYG2	Mammalia	Primates	Hominidae	~3.8	6x
<i>Procapia capensis</i>	Hyrax	proCap1	Mammalia	Hyracoidea	Procaviidae	4	2x
<i>Pteropus vampyrus</i>	Megabat	pteVam1	Mammalia	Chiroptera	Pteropodidae	2,3	2x
<i>Rattus norvegicus</i>	Rat	RGSC3.4	Mammalia	Rodentia	Muridae	~3.3	7x
<i>Sarcophilus harrisi</i>	Tasmanian devil	DEVIL7.0	Mammalia	Dasyuromorphia	Dasyuridae	3,6	2x
<i>Sorex araneus</i>	Shrew	COMMON_SHREW1	Mammalia	Eulipotyphla	Soricidae	2,8	2x
<i>Spermophilus tridecemlineatus</i>	Squirrel	SQUIRREL	Mammalia	Rodentia	Sciuridae	~4,17*	2x
<i>Sus scrofa</i>	Pig	Sscrofa9	Mammalia	Cetartiodactyla	Suidae	~3.1	15x
<i>Taeniopygia guttata</i>	Zebra Finch	taeGut3.2.4	Aves	Passeriformes	Estrilidae	1,25	6x
<i>Takifugu rubripes</i>	Fugu	FUGU 4.0	Actinopterygii	Tetraodontiformes	Tetraodontidae	0,4	8.7x
<i>Tarsius syrichta</i>	Tarsier	tarSyr1	Mammalia	Primates	Tarsiidae	5,14	6x
<i>Tetraodon nigroviridis</i>	Tetraodon	TETRAODON8	Actinopterygii	Tetraodontiformes	Tetraodontidae	0,42	2x
<i>Tupaia belangeri</i>	Tree Shrew	TREESHREW	Mammalia	Scandentia	Tupaiaidae	~3.4*	2x
<i>Tursiops truncatus</i>	Dolphin	turTru1	Mammalia	Cetartiodactyla	Delphinidae	~3.1	2x
<i>Vicugna pacos</i>	Alpaca	vicPac1	Mammalia	Cetartiodactyla	Camelidae	3.2*	2.51X
<i>Xenopus tropicalis</i>	Xenopus tropicalis	JGI_4.2	Amphibia	Anura	Pipidae	~1.7	7.65x

¹ GENOME SIZE: ~ = average estimate; * = estimate based on a congeneric species; Data from: the Animal Genome Size Database (<http://www.genomesize.com/>), Ensembl Database, Medaka Genome Project, Fugu Genome Project and Suine Genomic Database.

Table 2. Presence of vertebrate *PAX* genes ¹

SPECIES	NAME	PAX1	PAX2	PAX3	PAX4	PAX5	PAX6	PAX7	PAX8	PAX9
<i>Ailuropoda melanoleuca</i>	Panda	1	1	1	1	1	1	1	1	1
<i>Anolis carolinensis</i>	Lizard	1	1	1	0	1	1	0	0	1
<i>Bos taurus</i>	Cow	1	1	1	1	1	1	1	1	1
<i>Callithrix jacchus</i>	Marmoset	1	1	1	1	1	1	1	1	1
<i>Canis familiaris</i>	Dog	1	1	1	1	1	0	1	1	1
<i>Cavia porcellus</i>	Guinea Pig	1	1	1	1	1	1	1	1	1
<i>Choloepus hoffmanni</i>	Sloth	1	1	1	1	1	3	1	1	3
<i>Ciona intestinalis</i>	SeaSquirt	*	*	*	*	*	*	*	*	*
<i>Danio rerio</i>	Zebrafish	2	2	2	1	1	2	2	1	1
<i>Dasybus novemcinctus</i>	Armadillo	1	1	1	3	1	1	1	1	1
<i>Dipodomys ordii</i>	Kangaroo rat	1	1	3	1	1	1	1	1	3
<i>Echinops telfairi</i>	Tenrec	1	1	1	1	1	1	1	1	1
<i>Equus caballus</i>	Horse	1	1	1	1	1	1	1	1	1
<i>Erinaceus europaeus</i>	Hedgehog	1	1	1	1	1	3	3	1	1
<i>Felis catus</i>	Cat	1	1	1	1	1	1	3	1	1
<i>Gadus morhua</i>	Cod	2	1	2	1	1	2	2	1	1
<i>Gallus gallus</i>	Chicken	1	1	1	0	1	1	1	0	1
<i>Gasterosteus aculeatus</i>	Stickleback	1	2	1	1	1	1	2	1	1
<i>Gorilla gorilla</i>	Gorilla	1	1	1	1	1	1	1	1	1
<i>Homo sapiens</i>	Human	1	1	1	1	1	1	1	1	1
<i>Latimeria chalumnae</i>	Coelacanth	1	1	1	1	1	1	1	1	1
<i>Loxodonta africana</i>	Elephant	1	1	1	1	1	1	1	1	1
<i>Macaca mulatta</i>	Macaque	1	1	1	1	1	1	1	1	1
<i>Macropus eugenii</i>	Wallaby	1	1	1	3	1	1	1	1	1
<i>Meleagris gallopavo</i>	Turkey	1	1	1	0	1	1	1	0	1
<i>Microcebus murinus</i>	Mouse Lemur	1	1	1	1	1	1	1	1	1
<i>Monodelphis domestica</i>	Opossum	1	1	1	1	1	1	3	1	1
<i>Mus musculus</i>	Mouse	1	1	1	1	1	1	1	1	1
<i>Myotis lucifugus</i>	Microbat	1	1	1	1	1	1	3	1	1
<i>Nomascus leucogenys</i>	Gibbon	1	1	1	1	1	1	1	1	1
<i>Ochotona princeps</i>	Pika	1	1	1	1	1	1	1	1	1
<i>Ornithorhynchus anatinus</i>	Platypus	1	1	1	1	1	1	1	1	1
<i>Oryctolagus cuniculus</i>	Rabbit	1	1	1	1	1	1	1	1	1
<i>Oryzias latipes</i>	Medaka	1	1	1	1	1	1	1	1	3
<i>Otolemur garnettii</i>	Bushbaby	1	1	1	1	1	1	1	1	1
<i>Pan troglodytes</i>	Chimpanzee	1	1	1	1	1	1	1	1	1
<i>Petromyzon marinus</i>	Lamprey	*	*	*	*	*	*	*	*	*
<i>Pongo abelii</i>	Orangutan	1	1	1	1	1	1	1	1	1
<i>Procavia capensis</i>	Hyrax	1	1	1	1	1	1	1	1	1
<i>Pteropus vampyrus</i>	Megabat	1	1	1	1	1	1	1	1	1
<i>Rattus norvegicus</i>	Rat	1	1	1	1	1	1	1	1	1
<i>Sarcophilus harrisii</i>	Tasmanian devil	1	1	1	1	1	1	1	1	1
<i>Sorex araneus</i>	Shrew	3	3	1	1	1	1	1	1	1
<i>Spermophilus tridecemlineatus</i>	Squirrel	1	1	3	1	1	3	1	1	1
<i>Sus scrofa</i>	Pig	1	1	1	1	1	1	1	1	1
<i>Taeniopygia guttata</i>	Zebra Finch	1	1	1	0	1	1	1	0	1
<i>Takifugu rubripes</i>	Fugu	1	1	2	1	1	1	2	1	1
<i>Tarsius syrichta</i>	Tarsier	3	1	1	1	1	1	3	1	3
<i>Tetraodon nigroviridis</i>	Tetraodon	1	1	2	2	1	1	2	1	2
<i>Tupaia belangeri</i>	Tree Shrew	3	1	1	0	1	1	1	1	1
<i>Tursiops truncatus</i>	Dolphin	1	1	1	1	1	1	1	1	0
<i>Vicugna pacos</i>	Alpaca	3	3	1	3	3	3	1	1	3
<i>Xenopus tropicalis</i>	Frog	1	1	1	1	1	1	1	1	1

¹ 0= confirmed loss ; 1 = presence; * presence of the *PAX* ancestor ; 2= duplication; 3= missing data

Table 3. Characterization and location of PAX genes in Saurropsida (Reptiles and Birds)¹

Gene	Human Position	Chicken Ensembl Gene ID	Chicken Position	Turkey Ensembl Gene ID	Turkey Position	Zebra Finch Ensembl Gene ID	Zebra Finch Position	Lizard Ensembl Gene ID	Lizard Position
PAX1	chr20:21686297-21696620	new ortholog^a	Chr3:3419144-3422033	ENSMGAG000000003186	chr2:35050533-3507914	ENSTGUG0000000007526	chr3:29163929-29168275	ENSACAG000000010563	Chr4:81638985-81649093
PAX2	chr10:102495360-102589698	ENSGALG000000005689	Chr6:18179496-18258588	ENSMGAG000000007277	chr6:16208765-16290461	ENSTGUG0000000007620	chr6:16079464-16161804	ENSACAG000000011033	GL343388.1:918083:926311
PAX3	chr2:223064607-223163715	ENSGALG000000005246	Chr9:8937722-9001288	ENSMGAG000000004645	chr11:75117133-7539848	ENSTGUG0000000007761	chr9:8803902-8879407	ENSACAG0000000004981	Chr3:25697633-25825544
PAX4	chr7:127250346-127255982								
PAX5	chr9:36833272-37034476	ENSGALG000000016431	Chr7:74379713-74488259	possible ortholog^b	ChrZ:80746851-80747029	ENSTGUG0000000007215	chrZ:72604672-72742717	ENSACAG000000011646	GL343201.1:3094781-3339110
PAX6	chr11:31806340-31839509	ENSGALG000000012123	Chr5:5255551-5264599	ENSMGAG00000001533	chr5:1179661-1183246	ENSTGUG0000000004824	chr5:2920118-2931028	ENSACAG0000000002252	Chr1:60707886-60734555
PAX7	chr1:18957500-19075360	ENSGALG000000003782	Chr21:4432735-4518485	ENSMGAG000000005870	chr23:4623202-4712664	ENSTGUG000000016719^c	Chr21_random:284161:284358	possible ortholog^d	chrUn_AAW:202036617:1-24388
PAX8	chr2:113973574-114036577								
PAX9	chr14:37126773-37148920	ENSGALG000000010114	Chr5:39103416-39121700	ENSMGAG000000012530	chr5:38167072-38184166	ENSTGUG000000012015	chr5:36753571-36771385	ENSACAG0000000001481	Chr1:24821681-24822340

¹chr= chromosome; ^a = Chicken PAX1 identified in PreEnsemble Galgal4; ^b = turkey PAX5 possible ortholog to a segment of 60 amino acids with 95% identity with Chicken PAX5; ^c = Zebra Finch PAX7 possible ortholog to a segment of 66 amino acids with 100% identity with Chicken PAX7; ^d = Lizard PAX7 possible ortholog due to its similarity of the chicken PAX7.

Table 4. Pairwise distances between the *PAX* subfamilies and the Sea Squirt *PAX*⁴

Species	<i>PAX1</i>	<i>PAX9</i>	<i>PAX4</i>	<i>PAX6</i>	<i>PAX3</i>	<i>PAX7</i>	<i>PAX2</i>	<i>PAX5</i>	<i>PAX8</i>
Lamprey	0.398	0.398	*	*	*	*	0.660	*	*
Zebrafish	0.438	0.468	1.569	0.578	0.946	0.945	0.634	0.646	0.737
Cod	0.502	0.461	MD	0.584	1.249	0.928	MD	MD	0.732
Medaka	0.463	?	MD	0.581	1.005	0.904	0.634	0.705	0.717
Stickleback	0.507	0.503	ND	0.581	1.002	0.944	MD	0.676	MD
Fugu	0.471	0.488	MD	0.585	1.024	0.932	MD	0.796	MD
Tetraodon	0.489	0.485	MD	0.571	0.927	0.959	MD	0.717	MD
Coelacanth	0.502	0.442	MD	0.573	0.857	0.819	0.651	0.502	0.707
Frog	0.486	0.425	MD	0.639	MD	0.932	0.222	0.608	0.685
Lizard	0.490	0.467	L	0.579	0.859	MD	0.210	0.605	L
Zebra Finch	0.522	0.470	L	0.675	0.855	MD	0.607	0.627	L
Turkey	0.458	0.436	L	MD	1.142	0.939	MD	MD	L
Chicken	0.486	0.435	L	0.581	MD	0.932	0.606	0.639	L
Platypus	0.445	0.422	1.018	0.833	0.862	MD	MD	0.606	0.690
Tasmanian devil	0.429	0.425	1.155	0.576	MD	0.924	0.607	0.623	0.664
Wallaby	0.429	0.428	MD	0.669	0.864	MD	0.610	MD	MD
Opossum	0.429	0.428	MD	MD	0.876	MD	0.610	0.612	0.685
Armadillo	MD	0.443	MD	MD	MD	MD	MD	MD	MD
Tenrec	MD	0.460	MD	MD	MD	MD	0.615	0.610	MD
Hyrax	0.457	MD	1.347	0.579	0.871	MD	0.606	0.611	MD
Elephant	0.462	0.441	1.440	0.578	0.869	0.907	0.610	0.629	0.670
Shrew	MD	0.451	1.351	MD	MD	MD	MD	MD	MD
Hedgehog	0.476	MD	1.237	MD	MD	MD	0.625	0.626	MD
Megabat	0.477	MD	1.268	MD	MD	MD	0.610	MD	MD
Microbat	0.465	0.455	1.504	0.578	0.878	MD	0.646	0.620	0.677
Dog	0.471	0.455	1.212	MD	0.866	0.909	0.610	0.620	MD
Panda	0.461	0.455	1.235	0.578	0.881	0.904	0.610	0.620	0.677
Horse	MD	0.453	1.197	0.641	0.884	MD	0.610	0.631	MD
Cow	0.470	0.455	1.270	0.578	0.876	0.911	0.610	0.620	0.667
Dolphin	0.470	MD	1.373	MD	MD	0.923	0.610		0.696
Pig	0.470	0.455	1.191	0.578	MD	MD	0.599	0.620	MD
Pika	MD	MD	1.318	0.581	MD	0.952	0.610	0.620	MD
Rabbit	0.460	0.508	1.207	0.578	0.876	0.916	0.620	0.626	0.676
Squirrel	MD	MD	1.299	MD	MD	MD	MD	MD	MD
Guinea Pig	0.448	0.449	1.160	0.577	0.875	MD	0.607	0.620	0.667
Kangaroo rat	0.334	MD	1.265	MD	MD	0.923	0.594	MD	MD
Rat	0.453	0.465	1.429	0.578	0.878	0.912	0.626	MD	0.674
Mouse	0.454	0.465	1.277	0.578	0.872	0.914	0.626	0.620	0.672
Tree Shrew	MD	0.455	MD	MD	0.873	MD	0.459	MD	MD
Bushbaby	0.459	0.437	1.240	0.578	0.874	MD	0.610	0.620	0.677
Mouse Lemur	0.452	0.455	1.324	0.578	0.882	0.930	0.588	0.620	MD
Marmoset	0.448	0.456	1.218	0.578	0.883	0.905	0.610	0.620	MD
Macaque	0.447	0.460	1.261	0.578	0.876	0.932	0.610	1.490	MD
Gibbon	0.447	0.455	1.231	0.578	0.876	0.919	MD	0.620	0.679
Orangutan	0.447	0.455	1.204	MD	0.876	0.910	0.610	0.620	0.683
Gorilla	0.447	0.455	1.219	0.578	0.876	0.937	0.610	0.620	0.679
Chimpanzee	0.447	0.455	1.172	0.772	0.876	0.921	0.610	0.620	MD
Human	0.447	0.455	1.180	0.578	0.876	0.921	0.610	0.620	0.679

⁴MD = missing data - the sequences are too short or too divergent to be aligned; * presence of the *PAX* ancestor; L= genes losses.

Table 5. Branch-site clade model D values for ω (dN/dS ratio) estimated for two site classes.

PAX subfamily	Proportion (p)	Clade Model D		Branch type 1 (ω)	M1a - Parameters estimates		P* (LRT) M1a X Clade Model
		Branch type 0 (ω)	Branch type 1 (ω)		PAX1/PAX9	PAX2/PAX5/PAX8	
PAX1 and PAX9	0.87268	0.00186	PAX1	0.00186	p: 0.98611	0.01389	<0.001
	0.12732	0.08140	PAX9	0.09185	ω : 0.01023	1.00000	
PAX2, PAX5 and PAX8	0.62462	0.01192	PAX2/PAX5	PAX8	PAX2/PAX5/PAX8		<0.001
	0.37538	0.00027	PAX3	0.01192	p: 0.99999	0.00001	
PAX3 and PAX7	0.68577	0.01016	PAX3	PAX7	PAX3/PAX7		<0.001
	0.31423	0.09976	PAX4	0.01016	p: 0.99360	0.00640	
PAX4 and PAX6	0.61493	0.02335	PAX4	PAX6	PAX4/PAX6		<0.001
	0.38507	0.30208	PAX5	0.02335	p: 0.83385	0.16615	
				0.01784	ω : 0.06400	1.00000	

*Degrees of freedom: 2; LRT: $2\Delta l = 2(l_1 - l_0)$;

Supplementary Table 1. Description and functions of *PAX* genes

Gene	Description	Summary	Function	Human Disorders
<i>PAX1</i>	paired box 1	The <i>PAX</i> genes, including <i>PAX1</i> , are a highly conserved family of developmental control genes that encode transcription factors and have been shown to play a role in pattern formation during embryogenesis in vertebrates.	This protein is a transcriptional activator. Play a role in the formation of segmented structures of the embryo and have a important role in the normal development of the vertebral column.	Vertebral anomalies.
<i>PAX2</i>	paired box 2	<i>PAX2</i> encodes paired box gene 2, one of many human homologues of the <i>Drosophila</i> melanogaster gene <i>prd¹</i> . The central feature of this transcription factor gene family is the conserved DNA-binding paired box domain. <i>PAX2</i> is believed to be a target of transcriptional suppression by the tumor suppressor gene <i>WT1</i> . Alternative splicing of this gene results in multiple transcript variants.	Probable transcription factor that may have a role in kidney cell differentiation. Has a critical role in the development of the urogenital tract, the eyes, and the CNS.	Defects in <i>PAX2</i> are the cause of renal-coloboma syndrome (RCS); also known as papillorenal syndrome or optic nerve coloboma with renal disease. RCS is an autosomal dominant disease characterized by the association of renal hypoplasia, vesicoureteral reflux and dysplasia of the retina and optic disk.
<i>PAX3</i>	paired box 3	This gene is a member of the paired box (<i>PAX</i>) family of transcription factors. These genes play critical roles during fetal development. Alternative splicing results in transcripts encoding isoforms with different C-termini.	Transcription factor have a role in myogenesis, and in CNS.	Mutations in paired box gene 3 are associated with Waardenburg syndrome, craniofacial-deafness-hand syndrome, and alveolar rhabdomyosarcoma. The translocation t(2;13)(q35;q14), which represents a fusion between <i>PAX3</i> and the forkhead gene, is a frequent finding in alveolar rhabdomyosarcoma.
<i>PAX4</i>	paired box 4	This gene is a member of the paired box (<i>PAX</i>) family of transcription factors. These genes play critical roles during fetal development and cancer growth. The paired box 4 gene is involved in pancreatic islet development and mouse studies have demonstrated a role for this gene in differentiation of insulin-producing beta cells.	Plays an important role in the differentiation and development of pancreatic islet beta cells. Transcriptional repressor that binds to a common element in the glucagon, insulin and somatostatin promoters. Competes with <i>PAX6</i> for this same promoter binding site.	Defects in <i>PAX4</i> are a cause of noninsulin-dependent diabetes mellitus (NIDDM) also known as diabetes mellitus type 2 or maturity-onset diabetes.
<i>PAX5</i>	paired box 5	This gene encodes a member of the paired box (<i>PAX</i>) family of transcription factors. The central feature of this gene family is a novel, highly conserved DNA-binding motif, known as the paired box. <i>PAX</i> proteins are important regulators in early development, and alterations in the expression of their genes are thought to contribute to neoplastic transformation. This gene encodes the B-cell lineage specific activator protein that is expressed at early, but not late stages of B-cell differentiation. Its expression has also been detected in developing CNS and testis and so the encoded protein may also play a role in neural development and spermatogenesis.	May play an important role in B-cell differentiation as well as neural development and spermatogenesis. Involved in the regulation of the CD19 gene, a B-lymphoid-specific target gene	A chromosomal aberration involving <i>PAX5</i> is a cause of acute lymphoblastic leukemia. Translocation t(9;18)(p13;q11.2) with ZNF521. Translocation t(9;3)(p13;p14.1) with <i>FOXP1</i> . Translocation t(9;12)(p13;p13) with <i>ETV6</i>

Supplementary Table 1. Cont.

Gene	Description	Summary	Function	Human Disorders
PAX6	paired box 6	This gene encodes paired box gene 6, one of many human homologs of the <i>Drosophila melanogaster</i> gene <i>prd</i> . In addition to the hallmark feature of this gene family, a conserved paired box domain, the encoded protein also contains a homeo box domain. Both domains are known to bind DNA, and function as regulators of gene transcription. This gene is expressed in the developing nervous system, and in developing eyes. Alternatively spliced transcript variants encoding either the same or different isoform have been found for this gene.	Transcription factor with important functions in the development of the eye, nose, central nervous system and pancreas. Required for the differentiation of pancreatic islet alpha cells . Competes with PAX4 in binding to a common element in the glucagon, insulin and somatostatin promoters. Regulates specification of the ventral neuron subtypes by establishing the correct progenitor domains .	Mutations in this gene are known to cause ocular disorders such as aniridia and Peter's anomaly.
PAX7	paired box 7	This gene is a member of the paired box (PAX) family of transcription factors. Three transcript variants encoding different isoforms have been found for this gene.	Play a role in myogenesis, satellite cells an CNS.	Defects in PAX7 are a cause of rhabdomyosarcoma type 2 (RMS2). It is a form of rhabdomyosarcoma, a highly malignant tumor of striated muscle derived from primitive mesenchymal cells and exhibiting differentiation along rhabdomyoblastic lines. Rhabdomyosarcoma is one of the most frequently occurring soft tissue sarcomas and the most common in children.
PAX8	paired box 8	This gene encodes a member of the paired box (PAX) family of transcription factors. This nuclear protein is involved in thyroid follicular cell development and expression of thyroid-specific genes. Alternatively spliced transcript variants encoding different isoforms have been described	Transcription factor for the thyroid-specific expression of the genes exclusively expressed in the thyroid cell type, maintaining the functional differentiation of such cells	Mutations in this gene have been associated with thyroid dysgenesis, thyroid follicular carcinomas and atypical follicular thyroid adenomas.
PAX9	paired box 9	This gene is a member of the paired box (PAX) family of transcription factors. The specific function of the paired box 9 gene is unknown but it involve development of stratified squamous epithelia as well as various organs and skeletal elements.	Transcription factor required for normal development of thymus, parathyroid glands, ultimobranchial bodies, teeth, skeletal elements of skull and larynx as well as distal limbs.	Defects in PAX9 are the cause of tooth agenesis selective type 3 (STHAG3) . A form of selective tooth agenesis, a common anomaly characterized by the congenital absence of one or more teeth.

Data collected from Ensemble (<http://www.ensembl.org/>); OMIM (<http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim>); Uniprot (<http://beta.uniprot.org/>); GeneCards (<http://www.genecards.org/index.shtml>);

Supplementary Material

>PAX1_chick - Coding sequence

```
GAGCACACGTACGGGGAGGTGAACCAGCTGGGTGGTGTCTTCGTGAACGGGCGTCCGCTG
CCCAACGCCATCCGCTGCGGATCGTGGAGCTGGCCCAGCTCGGCATTCGCCCGTGCAC
ATCAGCCGCCAGCTGCGGGTCTCGCACGGCTGCGTCAGCAAGATCCTGGCCCGCTACAAC
GAGACCGGCTCCATCCTGCCCGGAGCCATCGGAGGCAGCAAGCCGCGGGTCACCACCCCC
AACGTGGTCAAGCACATCCGCGACTACAAGCAGGGAGACCCCGGCATCTTCGCCTGGGAG
ATCCGCGACCGGCTGCTGGCCGACGGCGTCTGCGACAAGTACAACGTGCCGTGGTTCAGC
TCCATCAGCAGGATCCTGCGGAACAAAATCGGCAGCCTCTCGCAGCCCGCGGCCCTTAC
GACGGCGGCAAGCAGCCCCCCCCGAGCCCGCCCTGCCCTACAACCCCTTTACAGTAC
CCGTACCCCGGCCCATGGCCCCGCCGCGAGCCAAGATGAGCGCTCATCCCGGGCCCGTC
CCCGCGGCCACGTGCGGGTCCCCGCTCCTGGCCCTCGGCTCACTCTGTACCAACATC
CTGGGCATCCGCACCTTCGTGGAGCAGACGGGGGCTTTGGGTGGCACCGAGGGCTCTGCT
TACCCCCCAAATGGAAGACTGGCCCGCGTGAACAGGACTGCCTTCCCTTCCGCCAG
GCCGACAAAGCCGCCGTGGAAGGGGACATCAAATACCCGCAGCCCGCCCCGGGACTCTCC
TCCGTGGGCAGCTTCTCCCGCCTGCGCCTATCCCCCGGCAACCAGCACGGCGTCTAC
GGCGCGCCCGGCGGCTACATCCCCCGGGGCACCCCTGGCAGCCGCAGGGCCACCACGGC
CCCGGTGTGACGGTGCACGGAGGCGACCTGGCCGGCGCCATGGCCTTCAAGCAGCCCGGC
AGGGAAGTCACGGACAGAAAACCCGGCAGCCCGTGGGGAAGTCCCGGATCCTCTCAGC
CCCATCCACGGACTCTCGATCCCAGCCTCCTCCTCC
```

>PAX1_chick translated amino acid

```
EHTYGEVNQLGGVFNVRPLPNAIRLRIVELAQLGIRPCDISRQLRVSHGCVSKILARYN
ETGSILPGAIGGSKPRVTPNVVKHIRDYKQGDPGIFAWEIRDLLADGVCDKYNVPSVS
SISRILRNKIGSLSQPGGYPYDGGKQPPPQALPYNPLYQYYPGPMAPPAKMSAHPGPV
PAAHVGLPRSWPSAHSVTNILGIRTFVEQTGALGGTEGSAYPPKMEDWPGVNRRTAFPSAQ
ADKAAVEGDIKYPQAPGLSSVGSFLPACAYPPGNQHGVYGAPGGYIPPGHPWQPQGHHG
PGVTVHGGDLAGAMAFKQPGREVTDRKPGSPVGKSPDPLSPIHGLSIPASSS
```

>chromosome:Galgal4:3:3419144:3422033:-1

GAGCACACGTACGGGGAGGTGAACCAGCTGGGTGGTGTCTTCGTGAACGGGCGTCCGCTG
CCCAACGCCATCCGCTGCGGATCGTGGAGCTGGCCAGCTCGGCATTCGCCGTGCGAC
ATCAGCCGCCAGCTGCGGGTCTCGCACGGCTGCGTCAGCAAGATCCTGGCCCGCTACAAC
GAGACCGGCTCCATCTGCCCGGAGCCATCGGAGGCAGCAAGCCGCGGGTCAACACCCCC
AACGTGGTCAAGCACATCCGCGACTACAAGCAGGGAGACCCCGGCATCTTCGCCTGGGAG
ATCCGCGACCGGCTGCTGGCCGACGGGTCTGCGACAAGTACAACGTGCCGTGGTCAGC
TCCATCAGCAGGATCTGCGGAACAAAATCGGCAGCCTCTCGCAGCCCGCGGGCCCTTAC
GACGGCGCAAGCAGCCCCCGCAGCCCGCCCTGCCCTACAACCCCTTTACCAGTAC
CCGTACCCCGGCCCATGGCCCCGCCGCGCAGCAAGATGAGCGCTCATCCCGGGCCCGTC
CCCGCGGCCACGTCGGGCTGCCCGCTCCTGGCCCTCGGCTCACTCTGTACCAACATC
CTGGGCATCCGCACCTTCGTGGAGCAGACGGGTGGGTACCGAAAGTCACTGGGACAGCGC
TTCTGCCCTCGGGTTTTTCCGTGCGGAAAAGCCCCGACGGCTGAGAGTTGATCCTGACT
GCTCCGCGGCTCTGTCTGCCAGGGGCTTTGGGTGGCACCGAGGGCTCTGCTTACCCCC
CCAAAATGGAAGACTGGCCCGCGTGAACAGGACTGCCTTCCCTTCGCCCCAGGCCGACA
AAGCCGCGTGGAAGGGGACATCAAATACCCGAGGTAAGGGGCGGAGGGGCCGGGGCCG
GGGGCCCGCTCAGGGCTGCCACGCGCCTCCGACACGAGTGTCTGTGCCAGCCCGCCCC
GGGACTCTCTCCGTGGGAGCTTCTCCCGCCTGCGCCTATCCCCCGGCAACCAGCA
CGGCGTCTACGGCGCGCCCGCGGCTACATCCCCCGGGGCACCCCTGGCAGCCGAGGG
CCACCAGGCCCCGGTGTGACGGTGACGGAGGCGACCTGGCCGGCGCCATGGCCTTCAA
GCAGCCCGGAGGGAAGGTGAGCGGGGGCCGGGGTTCGGGGGGCGAAGGGGGGCGAGAGGG
ACGCGACTGGACGCGACGGGGCGGTGGGACCCCCCTACCTCGCTGTGCCCGATGGGCGC
AAATCCCTCCCGCACTACCGGCTAACGCGGCCCTGGGGCTGGAAACCATTTCCCGGA
GCCGTCCCGATCTTAAAGTTCGAAGGCAGTCCCAGAAAAGGCAGAGCGATAAATCGGGGA
TCGTTGCATCTTATTCCGAATCCCGACTGCCAGGGCCGCGCTCTCCGCCCTGGGCAGGGT
TTAAAGATAAACTCCGTCTACGGAACGCGGCTTTGATTTGAACGCGTTCACCGCTAATTA
GGACGATTCTTTTATCTGCCGTGAGTGATTCGGCTCAGTTCTGAAAGATTTATTTCCCC

GATAGGTGGTTTCGTTAAGAAATCCTGACTCATCTGTTCTGGATATAAAATCCCCAGTC
GATATCTTCTCCTTCTTCGTGGTATTGGACTTTCTAATATTTTTACATTTGCAATCCA
TTTCCCACTTATTAGTCCTTCTTTATTCTATTTCTTTTTAATATATATATGCATACAT
ATATATATATATACGACCTTTTACCGTTTCCATCCAACCTCCAGCAACCTGCTGCGAGTTA
AAACATCGGCCGTCTCGGGGACGAGTTATTGCGAATATCAGTGGGTGAAAACGGGTCGTT
GTCAACGTATTTTCTTTAAAAAGAATAAGAATAGCGAAAGAGAAAAGCAGGGAAAAGAA
AGGAAAGGAACAAAGAAAAAGAGAGAAAGGAGAGAGGGAAAGGAAAAAGAAAAAGAGGAAT
GGAGAGAAGGAATGAAGAAAGGGGGAGAAAAGTGAAAGGGAAAGGGGAAGGAAACGAACG
AAGCGGGCTTTTAATATATTTTTCGGATTATTCTCCCCACATTTCAAGCGACTTGAGAGT
TGAATGTTCTGACTTGTCTGCTTAAGCTGTCCCTGCCGAAAGGCCACGAACTTTCTT
TCATTTGTTATGGTTGTTTTCTATATTCTCCCTTTTCCCGTCGCTGAGCTCGGCTCT
CTGGCCCAAGTCGGTACCGCGGTGCTCCGGGCCGCTCTGCTCCGGGCTCGGCGGCTCCC
GCTGCGGGTGCGGGGATCTGCTCGGTGCTGCGGGCGGACGGGGGGTGGGGATTTGGGGAG
AGGGCAGGGAAAAAGAAGGAGAAAAGGAGAGAAAAATCAGTCGTGCTTTTTGTCTGTAG
GGTTCAGACCTTTTCCGACGTCGGGGCTGTTGTTTCGCTAGAAATTCTCCCCGTTTCTT
TGCTCGGCACAGCAGCGCTGGGACGTTTCGCACCCCTGCCCGTGCGGGGGATACGCGG
GGGATGCGCGGGTCCGACCCGCACCCGGGGCTGTGCCACAGCTGTACCACAGCCGGCAC
GGGCTGCAGAACACCCCCCTCCCCCGGTTCCCTGCCTTTCCCTTCTCTTCTT
TCCCTTCTTTGGCAAAGGGGATTCGTATCATCACGGTCCCTACTGCCTGGGGGACTCG
ACGGGATCTCCCTTCCCTGTGTCCCGGGAGGGTTGGAGGTTGCCACCCACCGCTCTGC
TGCTGTCTTTCTTTTTGTCCCTCTCCTCCTCCTGCAGTCACGGACAGAAAACCCGGCAG
CCCCGTGGGGAAAGTCCCCGGATCCTCTCAGCCCCATCCACGGACTCTCGATCCCAGCCTC
CTCCTCCTAG

CAPÍTULO 6 – CONCLUSÕES

Como as discussões específicas relativas a cada um dos temas deste trabalho já foram abordados nos manuscritos apresentados, este item será reservado somente para a apresentação de um enfoque mais geral sobre os assuntos estudados.

O primeiro passo, para que um gene seja indicado, como um bom candidato para estudos que visam desvendar a relação deste com um fenótipo específico, é avaliar se há alguma evidência funcional de que o mesmo estaria envolvido com o traço em consideração. Desta forma, é possível especular quais genes envolvidos em situações patológicas em humanos, como é o caso do *PAX9*, por exemplo, são bons candidatos para estudos que visam identificar as causas moleculares que levam as variações normais encontradas dentro e entre espécies. Outra especulação possível é a extensão da homologia de uma espécie para outra, ou seja, assumir num primeiro momento que uma alteração numa sequência codificadora ou regulatória do DNA de uma espécie, responsável por um determinado fenótipo, poderia também estar relacionada a um fenótipo correspondente em outra espécie.

Pereira e cols. (2006) sugeriram a hipótese que uma mutação no gene *PAX9*, Ala240Pro, estaria associada ao fenótipo de agenesia dos terceiros molares, a mais frequente malformação craniofacial não-sindrômica humana. Para testar esta hipótese e também para avaliar como as mutações no *PAX9* afetam outros fenótipos envolvidos no desenvolvimento dentário, foi realizado o estudo piloto com pacientes (agenesias congênitas, não sindrômicas, identificadas através de radiografias) e controles

(presença de todos os 32 dentes permanentes), descrito no item Resultados-Capítulo 3. Foi possível observar que homozigotos para o alelo *240Pro* apresentavam sempre agenesia de pelo menos um terceiro molar, independente do quadrante. Já os heterozigotos mostram variação no padrão da agenesia de terceiro-molares, o que corrobora a ideia de uma possível haploinsuficiência (Pereira e cols., 2006; Wang e cols., 2009).

Com o padrão de herança definido pelo estudo piloto (herança autossômica recessiva com expressividade variada), foi possível partir para o entendimento dos padrões de agenesia de terceiro(s)-molar(es) em nível populacional, já que tem sido descritas variações marcantes com relação a este traço ao longo dos grupos continentais humanos (Resultados-Capítulo 4). Para tanto, foi investigada a diversidade no éxon 3 e em regiões adjacentes, considerando 125 indivíduos nativos americanos e 77 africanos do sul do Saara. Esses dados foram integrados com a informação disponível para a mesma região de outros 115 indivíduos de diferentes origens geográficas obtidos por Pereira e cols. (2006). O alelo *240Pro* revela uma variada distribuição nos Sul-ameríndios (frequência média = 7 %), estando presente também em esquimós (30%), asiáticos (44%), europeus (30%), e africanos (20%). Os dados parecem indicar que uma sucessão de eventos, provavelmente ligados à saída dos humanos modernos da África para os outros continentes. Os resultados preliminares em testes de neutralidade com simulações demográficas permitiram observar que os valores negativos e significantes encontrados, considerando as amostras dos ameríndios Sul-americanos, seriam uma consequência de processos demográficos, em especial decorrentes da deriva genética que tem marcado o perfil

genético das populações ameríndias. Por outro lado, para europeus e esquimós, a introdução de cenários demográficos envolvendo expansão populacional, destaca que os achados envolvendo o polimorfismo *Ala240Pro* do éxon 3 do gene *PAX9* não podem ser só explicados pelo modelo neutro. Interessantemente, a frequência da agenesia do terceiro molar aumenta a partir da África, no mesmo sentido que aumenta a frequência do alelo variante. Igualmente instigante foi o fato de que a distribuição do alelo *240Pro*, e conseqüentemente do genótipo homocigoto *240Pro/Pro* observado (2%), poderia explicar toda a incidência de agenesia de terceiros molares encontrada na África (~2%, excluindo Khoi-San e Pigmeus). Para explicar a presença alta de agenesia dos terceiros molares fora do continente africano, por outro lado, o polimorfismo *Ala240Pro* não é suficiente. Buscou-se então avaliar a variação em outros 101 genes (aproximadamente 6.001 SNPs), além do *PAX9*, relacionados com o desenvolvimento craniofacial, incluindo evidentemente o desenvolvimento dental. Observou-se um padrão marcante, onde os africanos apresentavam um excesso de alelos ancestrais relacionadas aos 102 genes do desenvolvimento craniofacial, quando comparados aos europeus e asiáticos. Além deste excesso de alelos ancestrais em genes do desenvolvimento dentário em africanos quando comparados aos outros grupos continentais, foi encontrado ainda uma proporção relativamente alta (1.3%) de valores extremos de F_{ST} (0.6) quanto esta proporção é comparada com aquela encontrada para um conjunto de SNPs escolhidos ao acaso ao longo do genoma (Wu & Zhang, 2010). O cenário que emerge estaria relacionado com a pressão seletiva (inicialmente relaxamento da seleção purificadora, com posterior ação de seleção positiva num conjunto relativamente grande de genes/variantes), que poderia estar

atuando tanto na morfologia dentária, bem como em outras características relacionadas (tamanho da mandíbula, por exemplo), e que teria levado à adaptação a diversos e novos ambientes quando os humanos modernos deixaram a África.

Quando os primeiros migrantes entram na América (último continente a ser colonizado), outro cenário parece dominar a evolução do éxon 3 do *PAX9*. Há um excesso de variantes raras neste gene considerando este grupo nativo continental, o que ilustra claramente o papel dos fatores casuísticos na história evolutiva dos ameríndios, o que pode ter sobrepujado a ação de outras forças microevolutivas, tais como a seleção natural, cujo sinal de atuação foi detectado em outros continentes (Resultados-Capítulo 4). Vários estudos de nosso grupo de pesquisa têm atestado as particularidades genéticas dos nativos americanos quando estes são comparados com os demais grupos geográficos humanos. Em particular, as investigações com marcadores seletivamente neutros, tem indicado baixos níveis de variabilidade intrapopulacional, que contrasta com altos níveis de diversidade interpopulacional quando populações ameríndias são comparadas com africanos, europeus, e asiáticos. O particular processo de colonização e dispersão ao longo do continente Americano (Gonzalez-José cols., 2010), com especial destaque para a ação da deriva genética e efeito do fundador, resultou neste inconfundível padrão de variabilidade, onde a presença de variantes autóctones e exclusivas é marcante (ver como exemplos variantes do cromossomo Y; apêndices 7 e 8 desta tese). Os resultados apresentados acima revelaram que mesmo tratando de genes com forte implicação funcional, como é o caso do gene *PAX9*, o padrão parece ser mantido, já que a deriva pode facilmente levar a perda de variantes vantajosas, ou mesmo mascarar a ação da seleção natural,

que pode não ser detectada pelos métodos de análise aqui empregados. Deste modo, os resultados apresentados no presente estudo reforçam a ideia de que não se pode pretender entender o padrão da evolução molecular de genes de importante implicação no desenvolvimento sem se levar em conta aspectos particulares da história evolutiva das populações envolvidas.

Os estudos *in silico*, por sua vez (Resultados-Capítulo 5), revelaram também resultados instigantes. É sabido que o gene *PAX9* interage diretamente com o gene *MSX1* no desenvolvimento dentário. Entretanto, o *MSX1* também interage com outros genes da família *PAX*, e este fato introduz diversas possibilidades nos processos relativos à evolução molecular deste conjunto de genes e de suas implicações em nível fenotípico. Normalmente, eventos de duplicação genômica que resultam em famílias gênicas fornecem material para inovações funcionais, desde que após a duplicação de um dado gene, uma de suas cópias preserve sua função original enquanto outra possa ser “liberada” (relaxamento de seleção purificadora) para acumular mutações que sob determinadas e novas circunstâncias possam ser selecionadas e fixadas (Wang & Gu, 2001). Mais ainda, nossas investigações mostraram conjuntamente que determinados éxons de um gene regulador específico também poderiam estar “livres” para acumular mutações, desde que os outros éxons do mesmo gene ou ainda algum parálogo com função similar preservassem algumas das funções primordiais do gene regulador com função pleitrópica (Pereira e cols., 2006; Resultados-Capítulo 5).

Considerando o cenário discutido acima, através do estudo da família *PAX* em um contexto de 53 genomas, foi possível desvendar alguns padrões específicos nos

cordados. O padrão de evolução molecular mais claramente observado foi aquele onde a seleção purificadora é preponderante. Entretanto, há variação neste padrão geral ao longo dos membros das subfamílias *PAX* (Resultados-Capítulo 5), bem como nos diferentes táxons investigados. Os resultados indicaram o possível papel da subfuncionalização, onde membros específicos de uma subfamília embora compartilhem os mesmos domínios temporal/espacialmente de expressão podem adquirir funções tecido ou estágio específica. O mais instigante é o fato de uma linhagem de vertebrados (Sauropsida) composta por aves e répteis apresentar a perda de dois genes da família *PAX* (genes *PAX4* e *PAX8*). Pode-se especular que, funcionalmente este fato seria possível na medida em que os demais genes *PAX*, embora com funções especializadas também preservassem as funções originais, e, portanto, os membros restantes das subfamílias poderiam assumir as funções daqueles faltantes. Uma questão que emerge desses resultados é: porque houve essa perda gênica? Uma hipótese é a de que este fenômeno acompanha a redução, como um todo, do genoma das aves. Esta redução aconteceu por perda de regiões intergênicas, diminuição do tamanho dos introns e perda de vários genes e componentes de famílias gênicas (ex: MHC - (Hughes & Friedman, 2008; Hughes e cols., 2008). De acordo com a hipótese de Hughes & Friedman (2008) esta diminuição geral estaria vinculada com adaptações ao voo, embora esta sugestão seja alvo de controvérsias.

Sendo assim e tomando como um todo o conjunto de dados apresentados aqui, é possível especular que os genes mestres do desenvolvimento organizados em famílias gênicas, têm permitido que mudanças acontecessem em vários níveis: (1) A

variabilidade no éxon 3 do gene *PAX9* pode constituir o que Pereira e cols. (2006) chamou de “janela evolutiva do gene *PAX9*”;(2) Diferenças em genes do desenvolvimento craniofacial, bem como nos fenótipos comuns associados (agenesia dos terceiros molares, por exemplo), encontradas entre populações humanas podem estar relacionadas a adaptações na saída dos humanos modernos da África; (3) Membros de cada uma das subfamílias *PAX* parecem estar sobre diferentes pressões seletivas; (4) Houve a perda de dois genes *PAX* no táxon Sauropsida (aves e répteis), reforçando a ideia de evolução em “redes gênicas” (*gene networks*), robustas e redundantes, onde há uma integração funcional, e a implementação de novas funções ou a retomada de antigas funções pode acontecer.

A família *PAX*, por tanto, seria um exemplo notável de como se processa a evolvabilidade (Kirschner & Gerhart, 1998; Pigliucci, 2008; Resultados-Capítulo-4).

REFERÊNCIAS BIBLIOGRÁFICAS

- Abzhanov, A., Rodda, S.J., McMahon, A.P., and Tabin, C.J. (2007). Regulation of skeletogenic differentiation in cranial dermal bone. *Development (Cambridge, England)* 134, 3133–3144.
- Acuña-Alonzo, V., Flores-Dorantes, T., Kruit, J.K., Villarreal-Molina, T., Arellano-Campos, O., Hünemeier, T., Moreno-Estrada, A., Ortiz-López, M.G., Villamil-Ramírez, H., León-Mimila, P., et al. (2010). A functional ABCA1 gene variant is associated with low HDL-cholesterol levels and shows evidence of positive selection in Native Americans. *Human Molecular Genetics* 19, 2877–2885.
- Alappat, S., Zhang, Z.Y., and Chen, Y.P. (2003). Msx homeobox gene family and craniofacial development. *Cell Research* 13, 429–442.
- Balczarek, K. a, Lai, Z.C., and Kumar, S. (1997). Evolution of functional diversification of the paired box (Pax) DNA-binding domains. *Molecular Biology and Evolution* 14, 829–842.
- Bang, A.G., Papalopulu, N., Goulding, M.D., and Kintner, C. (1999). Expression of Pax-3 in the Lateral Neural Plate Is Dependent on a Wnt-Mediated Signal from Posterior Nonaxial Mesoderm. *Developmental Biology* 212, 366–380.
- Basch, M.L., Bronner-Fraser, M., and García-Castro, M.I. (2006). Specification of the neural crest occurs during gastrulation and requires Pax7. *Nature* 441, 218–222.
- Bersaglieri, T., Sabeti, P.C., Patterson, N., Vanderploeg, T., Schaffner, S.F., Drake, J. a, Rhodes, M., Reich, D.E., and Hirschhorn, J.N. (2004). Genetic signatures of strong recent positive selection at the lactase gene. *American Journal of Human Genetics* 74, 1111–1120.
- Blake, J. a, Thomas, M., Thompson, J. a, White, R., and Ziman, M. (2008). Perplexing Pax: from puzzle to paradigm. *Developmental Dynamics : an Official Publication of the American Association of Anatomists* 237, 2791–2803.
- Bowler, P.J. (1989). *Evolution: The History of an Idea* (University of California Press).
- Boyle, A.P., Song, L., Lee, B.-K., London, D., Keefe, D., Birney, E., Iyer, V.R., Crawford, G.E., and Furey, T.S. (2010). High-resolution genome-wide in vivo footprinting of diverse transcription factors in human cells. *Genome Research*.
- Brakefield, P.M. (2006). Evo-devo and constraints on selection. *Trends in Ecology & Evolution (Personal Edition)* 21, 362–368.
- Breitling, R., and Gerber, J.K. (2000). Origin of the paired domain. *Development Genes and Evolution* 210, 644–650.
- Buckingham, M., and Relaix, F. (2007). The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions. *Annual Review of Cell and Developmental Biology* 23, 645–673.

- Carroll, S.B. (2000). Endless forms: the evolution of gene regulation and morphological diversity. *Cell* 101, 577–580.
- Carroll, S.B. (2003). Genetics and the making of *Homo sapiens*. *Nature* 422, 849–857.
- Charles Darwin (2000). *A origem das espécies e a seleção natural* (São Paulo: Hemus).
- Chen, L., Zhang, Q., Wang, W., and Wang, Y. (2010). Spatiotemporal expression of Pax genes in amphioxus: insights into Pax-related organogenesis and evolution. *Science China. Life Sciences* 53, 1031–1040.
- Chen, Y., Zhang, Y., Jiang, T.X., Barlow, a J., St Amand, T.R., Hu, Y., Heaney, S., Francis-West, P., Chuong, C.M., and Maas, R. (2000). Conservation of early odontogenic signaling pathways in Aves. *Proceedings of the National Academy of Sciences of the United States of America* 97, 10044–10049.
- Chi, N., and Epstein, J. a (2002). Getting your Pax straight: Pax proteins in development and disease. *Trends in Genetics : TIG* 18, 41–47.
- Conway-Morris, S. (2003). The Cambrian “explosion” of metazoans and molecular biology: would Darwin be satisfied? *The International Journal of Developmental Biology* 47, 505–515.
- Crespi, B.J., and swanson, W.J. (2011). Beyond the phenotypic gambit: molecular behavioural ecology and the evolution of genetic architecture. *Molecular Ecology* 20, 2240–2257.
- Crombach, A., and Hogeweg, P. (2008). Evolution of evolvability in gene regulatory networks. *PLoS Computational Biology* 4, e1000112.
- Damante, G. (1998). Thyroid defects due to Pax8 gene mutations. *European Journal of Endocrinology* 139, 563 –566.
- Dressler, G.R. (2011). Patterning and early cell lineage decisions in the developing kidney: the role of Pax genes. *Pediatric Nephrology* (Berlin, Germany).
- Epstein, J., Cai, J., Glaser, T., Jepeal, L., and Maas, R. (1994). Identification of a Pax paired domain recognition sequence and evidence for DNA-dependent conformational changes. *The Journal of Biological Chemistry* 269, 8355–8361.
- Finnerty, J.R., Mazza, M.E., and Jezewski, P. a (2009). Domain duplication, divergence, and loss events in vertebrate Msx paralogs reveal phylogenomically informed disease markers. *BMC Evolutionary Biology* 9, 18.
- Fleischmannova, J., Matalova, E., Tucker, A.S., and Sharpe, P.T. (2008). Mouse models of tooth abnormalities. *European Journal of Oral Sciences* 116, 1–10.
- Foerst-Potts, L., and Sadler, T.W. (1997). Disruption of Msx-1 and Msx-2 reveals roles for these genes in craniofacial, eye, and axial development. *Developmental Dynamics* 209, 70–84.

- Frazier-Bowers, S. a, Guo, D.C., Cavender, a, Xue, L., Evans, B., King, T., Milewicz, D., and D'Souza, R.N. (2002). A novel mutation in human PAX9 causes molar oligodontia. *Journal of Dental Research* 81, 129–133.
- Gerhart, J., and Kirschner, M. (2007). The theory of facilitated variation. *Proceedings of the National Academy of Sciences of the United States of America* 104 Suppl, 8582–8589.
- Giampietro, P.F., Raggio, C.L., Reynolds, C.E., Shukla, S.K., McPherson, E., Ghebranious, N., Jacobsen, F.S., Kumar, V., Faciszewski, T., Pauli, R.M., et al. (2005). An analysis of PAX1 in the development of vertebral malformations. *Clinical Genetics* 68, 448–453.
- González-José, R., Bortolini, M. C., Santos, F. R., & Bonatto, S. L. (2008). The peopling of America: craniofacial shape variation on a continental scale and its interpretation from an interdisciplinary view. *American Journal of Physical Anthropology*, 137(2), 175-187.
- Gould, S.J., and Eldredge, N. (1993). Punctuated equilibrium comes of age. *Nature* 366, 223–227.
- Gröger, H., Callaerts, P., Gehring, W.J., and Schmid, V. (2000). Characterization and expression analysis of an ancestor-type Pax gene in the hydrozoan jellyfish *Podocoryne carnea*. *Mechanisms of Development* 94, 157–169.
- Hetzer-Egger, C., Schorpp, M., and Boehm, T. (2000). Evolutionary conservation of gene structures of the Pax1/9 gene family. *Biochimica Et Biophysica Acta* 1492, 517–521.
- Hill, A., Boll, W., Ries, C., Warner, L., Osswalt, M., Hill, M., and Noll, M. (2010). Origin of Pax and Six gene families in sponges: Single PaxB and Six1/2 orthologs in *Chalinula loosanoffi*. *Developmental Biology* 343, 106–123.
- Holland, L.Z., and Short, S. (2010). Alternative splicing in development and function of chordate endocrine systems: a focus on pax genes. *Integrative and Comparative Biology* 50, 22–34.
- Hu, G., Lee, H., Price, S.M., Shen, M.M., and Abate-Shen, C. (2001). Msx homeobox genes inhibit differentiation through upregulation of cyclin D1. *Development (Cambridge, England)* 128, 2373–2384.
- Hughes, A.L., and Friedman, R. (2008). Genome Size Reduction in the Chicken Has Involved Massive Loss of Ancestral Protein-Coding Genes. *Molecular Biology and Evolution* 25, 2681–2688.
- Hughes, C.R., Miles, S., and Walbroehl, J.M. (2008). Support for the minimal essential MHC hypothesis: a parrot with a single, highly polymorphic MHC class II B gene. *Immunogenetics* 60, 219–231.
- Hünemeier, T., Ruiz-Linares, A., Silveira, Á., Paixão-Côrtes, V.R., Salzano, F.M., and Bortolini, M.C. (2010). Brief communication: Population data support the adaptive nature of HACNS1 sapiens/neandertal-chimpanzee differences in a

- limb expression domain. *American Journal of Physical Anthropology* 143, 478–481.
- Jones, F.C., Grabherr, M.G., Chan, Y.F., Russell, P., Mauceli, E., Johnson, J., Swofford, R., Pirun, M., Zody, M.C., White, S., et al. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484, 55–61.
- Kirschner, M., and Gerhart, J. (1998). Evolvability. *Proc Natl Acad Sci U S A* 95, 8420–8427.
- Kist, R., Watson, M., Wang, X., Cairns, P., Miles, C., Reid, D.J., and Peters, H. (2005). Reduction of Pax9 gene dosage in an allelic series of mouse mutants causes hypodontia and oligodontia. *Human Molecular Genetics* 14, 3605–3617.
- Kleinjan, D. a, Seawright, A., Mella, S., Carr, C.B., Tyas, D. a, Simpson, T.I., Mason, J.O., Price, D.J., and van Heyningen, V. (2006). Long-range downstream enhancers are essential for Pax6 expression. *Developmental Biology* 299, 563–581.
- Kubic, J., Young, K., Plummer, R., Ludvik, A., and Lang, D. (2008). Pigmentation PAX-ways: the role of Pax3 in melanogenesis, melanocyte stem cell maintenance, and disease. *Pigment Cell Melanoma Res* 21, 627–645.
- Laakso, M. (2009). Not for the eyes only: PAX6 and glucose metabolism. *Diabetologia* 52, 381–384.
- Lang, D., Powell, S.K., Plummer, R.S., Young, K.P., and Ruggeri, B. a (2007). PAX genes: roles in development, pathophysiology, and cancer. *Biochemical Pharmacology* 73, 1–14.
- Lefèvre, C.M., Sharp, J. a, and Nicholas, K.R. (2010). Evolution of Lactation: Ancient Origin and Extreme Adaptations of the Lactation System. *Annual Review of Genomics and Human Genetics* 219–238.
- Lewin, R., and others (2005). Human evolution: an illustrated introduction. *Human Evolution: An Illustrated Introduction*.
- Li, C., and Eccles, M. (2012). PAX genes in cancer; friends or foes? *Frontiers* 3,.
- Line, S.R.P. (2003). Variation of tooth number in mammalian dentition: connecting genetics, development, and evolution. *Evolution & Development* 5, 295–304.
- Lynch, V.J., Tanzer, A., Wang, Y., Leung, F.C., Gellersen, B., Emera, D., and Wagner, G.P. (2008). Adaptive changes in the transcription factor HoxA-11 are essential for the evolution of pregnancy in mammals. *Proceedings of the National Academy of Sciences of the United States of America* 105, 14928–14933.
- Lynch, V.J., and Wagner, G.P. (2008). Resurrecting the role of transcription factor change in developmental evolution. *Evolution; International Journal of Organic Evolution* 62, 2131–2154.
- Miller, D.J., Hayward, D.C., Reece-Hoyes, J.S., Scholten, I., Catmull, J., Gehring, W.J., Callaerts, P., Larsen, J.E., and Ball, E.E. (2000). Pax gene diversity in the basal cnidarian *Acropora millepora* (Cnidaria, Anthozoa): implications for the

- evolution of the Pax gene family. *Proceedings of the National Academy of Sciences of the United States of America* 97, 4475–4480.
- Nei, M. (2007). The new mutation theory of phenotypic evolution. *Proceedings of the National Academy of Sciences of the United States of America* 104, 12235–12242.
- Nielsen, R. (2009). ADAPTIONISM—30 YEARS AFTER GOULD AND LEWONTIN. *Evolution* 63, 2487–2490.
- Ogasawara, M., Shigetani, Y., Hirano, S., Satoh, N., and 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. *Developmental Biology* 223, 399–410.
- Ogawa, T., Kapadia, H., Feng, J.Q., Raghow, R., Peters, H., and D'Souza, R.N. (2006). Functional consequences of interactions between Pax9 and Msx1 genes in normal and abnormal tooth development. *The Journal of Biological Chemistry* 281, 18363–18369.
- Ogawa, T., Kapadia, H., Wang, B., and D'Souza, R.N. (2005). Studies on Pax9-Msx1 protein interactions. *Archives of Oral Biology* 50, 141–145.
- Pasca di Magliano, M., Di Lauro, R., and Zannini, M. (2000). Pax8 has a key role in thyroid cell differentiation. *Proceedings of the National Academy of Sciences* 97, 13144–13149.
- Pereira, T.V., Salzano, F.M., Mostowska, A., Trzeciak, W.H., Ruiz-Linares, A., Chies, J. a B., Saavedra, C., Nagamachi, C., Hurtado, A.M., Hill, K., et al. (2006). Natural selection and molecular evolution in primate PAX9 gene, a major determinant of tooth development. *Proceedings of the National Academy of Sciences of the United States of America* 103, 5676–5681.
- Peres, R.C.R., Scarel-Caminaga, R.M., do Espírito Santo, A.R., and Line, S.R.P. (2005). Association between PAX-9 promoter polymorphisms and hypodontia in humans. *Archives of Oral Biology* 50, 861–871.
- Peters, H., Neubüser, A., and Balling, R. (1998). Pax genes and organogenesis: Pax9 meets tooth development. *European Journal of Oral Sciences* 106 Suppl 1, 38–43.
- Peters, H., Schuster, G., Neubüser, a, Richter, T., Höfler, H., and Balling, R. (1997). Isolation of the Pax9 cDNA from adult human esophagus. *Mammalian Genome : Official Journal of the International Mammalian Genome Society* 8, 62–64.
- Peters, H., Wilm, B., Sakai, N., Imai, K., Maas, R., and Balling, R. (1999). Pax1 and Pax9 synergistically regulate vertebral column development. *Development (Cambridge, England)* 126, 5399–5408.
- Pigliucci, M. (2008). Is evolvability evolvable? *Nature Reviews Genetics* 9, 75–82.

- Ritz-Laser, B., Estreicher, A., Gauthier, B., and Philippe, J. (2000). The Paired Homeodomain Transcription Factor Pax-2 Is Expressed in the Endocrine Pancreas and Transactivates the Glucagon Gene Promoter. *Journal of Biological Chemistry* 275, 32708–32715.
- Robson, E.J.D., He, S.-J., and Eccles, M.R. (2006). A PANorama of PAX genes in cancer and development. *Nature Reviews. Cancer* 6, 52–62.
- Rozkocová, E., Marková, M., Lánik, J., and Zvárová, J. (2004). Agenesis of third molars in young Czech population. *Prague Medical Report* 105, 35–52.
- Santagati, F., Gerber, J.K., Blusch, J.H., Kokubu, C., Peters, H., Adamski, J., Werner, T., Balling, R., and 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. *Mammalian Genome : Official Journal of the International Mammalian Genome Society* 12, 232–237.
- Short, S., and Holland, L.Z. (2008). The evolution of alternative splicing in the Pax family: the view from the Basal chordate amphioxus. *Journal of Molecular Evolution* 66, 605–620.
- Smith, S.B., Watada, H., Scheel, D.W., Mrejen, C., and German, M.S. (2000). Autoregulation and Maturity Onset Diabetes of the Young Transcription Factors Control the Human PAX4 Promoter. *Journal of Biological Chemistry* 275, 36910–36919.
- Suga, H., Tschopp, P., Graziussi, D.F., Stierwald, M., Schmid, V., and Gehring, W.J. (2010). Flexibly deployed Pax genes in eye development at the early evolution of animals demonstrated by studies on a hydrozoan jellyfish. *Proceedings of the National Academy of Sciences of the United States of America* 107, 14263–14268.
- Tiedge, M. (2005). Beta cell regeneration meets autoimmunity: PAX4 variants in type 1 diabetes. *Diabetologia* 48, 803–804.
- Tishkoff, S. a, Reed, F. a, Ranciaro, A., Voight, B.F., Babbitt, C.C., Silverman, J.S., Powell, K., Mortensen, H.M., Hirbo, J.B., Osman, M., et al. (2007). Convergent adaptation of human lactase persistence in Africa and Europe. *Nature Genetics* 39, 31–40.
- Torlakovic, E., Slipicevic, A., Robinson, C., DeCoteau, J.F., Alfsen, G.C., Vyberg, M., Chibbar, R., and Flørenes, V.A. (2006). PAX5 Expression in Nonhematopoietic Tissues. *American Journal of Clinical Pathology* 126, 798–804.
- Varki, A., and Altheide, T.K. (2005). Comparing the human and chimpanzee genomes: searching for needles in a haystack. *Genome Research* 15, 1746–1758.
- Vogel, C., and Chothia, C. (2006). Protein family expansions and biological complexity. *PLoS Computational Biology* 2, e48.

- Vorobyov, E., and Horst, J. (2006). Getting the proto-Pax by the tail. *Journal of Molecular Evolution* 63, 153–164.
- Wang, W., Zhong, J., and Wang, Y.-Q. (2010). Comparative genomic analysis reveals the evolutionary conservation of Pax gene family. *Genes & Genetic Systems* 85, 193–206.
- Wang, Y., Groppe, J.C., Wu, J., Ogawa, T., Mues, G., D'Souza, R.N., and Kapadia, H. (2009). Pathogenic mechanisms of tooth agenesis linked to paired domain mutations in human PAX9. *Human Molecular Genetics* 18, 2863–2874.
- Wang, Y., and Gu, X. (2001). Functional divergence in the caspase gene family and altered functional constraints: statistical analysis and prediction. *Genetics* 158, 1311–1320.
- Wilm, B. (1998). Targeted disruption of Pax1 defines its null phenotype and proves haploinsufficiency. *Proceedings of the National Academy of Sciences* 95, 8692–8697.
- Xu, P., Zhang, X., Heaney, S., Yoon, A., Michelson, A., and Maas, R. (1999). Regulation of Pax6 expression is conserved between mice and flies. *Development* 126, 383–395.
- Xu, W., Rould, M. a, Jun, S., Desplan, C., and Pabo, C.O. (1995). Crystal structure of a paired domain-DNA complex at 2.5 Å resolution reveals structural basis for Pax developmental mutations. *Cell* 80, 639–650.
- Xue, L., and Noll, M. (1996). The functional conservation of proteins in evolutionary alleles and the dominant role of enhancers in evolution. *The EMBO Journal* 15, 3722–3731.

***APÊNDICE 1 – “HOMO SAPIENS, HOMO NEANDERTHALENSIS
AND THE DENISOVA SPECIMEN: NEW INSIGHT ON THEIR
EVOLUTIONARY HISTORIES USING WHOLE-GENOME
COMPARISONS”***

Manuscrito aceito no periódico Genetics and Molecular Biology

***APÊNDICE 2 – “PREDICTING HOMO PIGMENTATION PHENOTYPE
THROUGH GENOMIC DATA: FROM NEANDERTHAL TO JAMES
WATSON”***

Manuscrito publicado no periódico American Journal of Human Biology

***APÊNDICE 3 – “LINKING DOPAMINE NEUROTRANSMISSION AND
NEUROGENESIS: THE EVOLUTIONARY HISTORY OF THE NTAD
(NCAM1-TTC12-ANKK1-DRD2) GENE CLUSTER”***

Manuscrito aceito no periódico Genetics and Molecular Biology

***APÊNDICE 4 – “EVOLUTIONARY RESPONSES TO A CONSTRUCTED
NICHE: ANCIENT MESOAMERICANS AS A MODEL OF GENE-
CULTURE COEVOLUTION”***

Manuscrito aceito no periódico PLoS One (Public Library of Science)

***APÊNDICE 5 – “POPULATION DATA SUPPORT THE ADAPTIVE
NATURE OF HACNS1 SAPIENS/NEANDERTAL-CHIMPANZEE
DIFFERENCES IN A LIMB EXPRESSION DOMAIN”***

Manuscrito publicado no periódico American Journal of Physical Anthropology

APÊNDICE 6 – “THE TP53 FERTILITY NETWORK”

Manuscrito aceito no periódico Genetics and Molecular Biology

***APÊNDICE 7 – “DISTRIBUTION OF Y-CHROMOSOME Q LINEAGES
IN NATIVE AMERICANS”***

Manuscrito publicado no periódico American Journal of Human Biology

***APÊNDICE 8 - "A NEW SUBHAPLOGROUP OF NATIVE AMERICAN
Y-CHROMOSOMES FROM THE ANDES"***

Manuscrito publicado no periódico American Journal of Physical Anthropology