Natural selection and molecular evolution in primate *PAX9* gene, a major determinant of tooth development

Tiago V. Pereira[†], Francisco M. Salzano[†‡], Adrianna Mostowska[§], Wieslaw H. Trzeciak[§], Andrés Ruiz-Linares[¶], José A. B. Chies[†], Carmen Saavedra[†], Cleusa Nagamachi[∥], Ana M. Hurtado[††], Kim Hill[††], Dinorah Castro-de-Guerra[‡‡], Wilson A. Silva-Júnior[§§], and Maria-Cátira Bortolini[†]

[†]Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, Brazil; [§]Department of Biochemistry and Molecular Biology, University of Medical Sciences, 6 Swiecickiego, 60-781, Poznan, Poland; [¶]The Galton Laboratory, University College London, 4 Stephenson Way, London NW1 2HE, United Kingdom; [∥]Departamento de Genética, Centro de Ciências Biológicas, Universidade Federal do Pará, Campus Universitário do Guamá, 66075-970 Belém, Brazil; [††]Department of Anthropology, University of New Mexico, Albuquerque, NM 87131; [‡‡]Laboratório de Genética Humana, Instituto Venezolano de Investigaciones Científicas, Apartado 21827, Caracas 1020A, Venezuela; and [§§]Departamento de Genética, Faculdade de Medicina, Universidade de São Paulo, Campus Universitário, 14049-900 Ribeirão Preto, Brazil

Large differences in relation to dental size, number, and morphology among and within modern human populations and between modern humans and other primate species have been observed. Molecular studies have demonstrated that tooth development is under strict genetic control, but, the genetic basis of primate tooth variation remains unknown. The *PAX9* gene, which codes for a paired domain-containing transcription factor that plays an essential role in the development of mammal dentition, has been associated with selective tooth agenesis in humans and mice, which mainly involves the posterior teeth. To determine whether this gene is polymorphic in humans, we sequenced ≈2.1 kb of the entire four-exon region (exons 1, 2, 3 and 4; 1,026 bp) and exon-intron (1.1 kb) boundaries of 86 individuals sampled from Asian, European, and Native American populations. We provided evidence that human *PAX9* polymorphisms are limited to exon 3 only and furnished details about the distribution of a mutation there in 350 Polish subjects. To investigate the pattern of selective pressure on exon 3, we sequenced ortholog regions of this exon in four species of New World monkeys and one gorilla. In addition, orthologous sequences of *PAX9* available in public databases were also analyzed. Although several differences were identified between humans and other species, our findings support the view that strong purifying selection is acting on *PAX9*. New World and Old World primate lineages may, however, have different degrees of restriction for changes in this DNA region.

human evolution | tooth agenesis | evolutionary window

One of the several characteristics that distinguish *Homo sapiens* from other primate species is numerical and morphological tooth variation (1). In addition, although several studies have demonstrated an appreciable reduction of tooth size in modern humans when compared with earlier hominids (2–4), variations in the number and size of teeth are currently found within our species. For example, third molar agenesis, whose etiology is largely unknown, is by far the most common form of variation in number of teeth in humans (5). Based on large samples, agenesis of third molar ranged from 0.2% in Bantu speakers from Angola (6) to nearly 100% in Mexican Indians (7). Evidence supporting a genetic etiology for tooth agenesis and their shape is well established (8–10). However, despite the fact that much work has been done over the past 60 years to explore morphological and numerical diversity in the primate dentition (2–5, 11, 12), important questions remain to be answered. Do the differences in morphology and the number of teeth among humans and between humans and other primate species reflect variations of the genes involved in tooth development?

In this respect, several genes that are pivotal in initiating the development of teeth have been subjected to intense study in the past decade (10, 13). Mutations in a number of genes were found to interrupt tooth development in mice (10, 13). However, to date there are only three genes associated with the nonsyndromic form of human tooth agenesis: *AXIN2*, *MSX1*, and *PAX9* (14, 15). Among them, *PAX9* was more intensively studied. Recently, the general structure of the PAX paired domain was described (16) and the phylogenetics and relation between the several members of the *PAX* family were established (17). In addition, both gene expression (18–21) and molecular pathogenesis (22, 23) of *PAX9* have been relatively well characterized, making it a special candidate to explain at least part of primate tooth variation.

The *PAX* gene family codes for transcription factors, which play a major role in the early development of a variety of multicellular organisms (24). PAX proteins are defined by the presence of a 128-aa DNA-binding domain, the paired domain, which makes sequence-specific contacts with DNA (16, 24). In mammals, nine *PAX* genes (*PAX1–PAX9*) have been found (13, 24). *PAX9* is expressed in the pharyngeal pouches, developing vertebral column and neural-crest-derived mesenchyme of maxillary and mandibular arches, contributing to palate and tooth formation (18, 19, 25, 26). The importance of *PAX9* in tooth development is supported by the phenotype of an arrest at the bud stage that is observed in *Pax9* knockout mice (27). In humans, mutations in exons 2 and 4 of *PAX9* are associated with both nonsyndromic hypodontia and oligodontia (23, 28, 29). These mutations or the deletion of one *PAX9* allele (29) are reported to cause selective tooth agenesis, especially agenesis of molars. All mutations in *PAX9* associated with selective tooth agenesis were only found in heterozygosis (22, 29), suggesting that the mutant phenotypes are caused by haploinsufficiency and reinforcing the pivotal importance of this gene in tooth development.

Although the role and molecular pathogenesis of the paired box domain has been evaluated extensively, little attention has been given to other *PAX9* regions. It was recently suggested that some variation at exon 3 could be common in humans (28). However, because these authors primarily considered the role of the paired domain in molar oligodontia, the extent of human diversity on this gene remained unknown.

Table 1. Allele and genotype distributions of the Ala240Pro polymorphism in different populations

Group and method of study	n	Ala/Ala	Ala/Pro	Pro/Pro	Pro allele frequency
		Genotype, n (%)			
Sequencing					
Native Americans	57	52 (91)	5 (9)	0 (0)	0.04
Asians	14	11 (79)	2 (14)	1 (7)	0.14
Europeans	15	10 (67)	3 (20)	2 (13)	0.23
RFLP tests					
Polish patients					
H/O	52	22 (42)	29 (56)	1 (2)	0.30
CL/P	154	77 (50)	67 (44)	10 (6)	0.28
Controls					
Healthy Polish	144	60 (42)	73 (51)	11 (7)	0.33

H/O, hypodontia/oligodontia; CL/P, cleft lip/palate.

Table 2. Population variability parameters for the coding region of human *PAX9*

Parameter	Native Americans	Europeans	Asians	Total
Chromosomes	114	30	28	172
η	4	2	1	7
η^s	2	0	0	2
π	1.3 ± 0.4	8.3 ± 0.8	2.5 ± 0.9	3.2 ± 0.5
θ_W	7.3 ± 4.0	4.9 ± 3.6	2.5 ± 2.5	6.8 ± 3.6
Tajima's D	−1.59[†]	1.40[‡]	−0.02	−0.98
Fu and Li's D*	−2.74[†]	0.80[‡]	0.60	−1.65
Fu and Li's F*	−2.79[†]	1.13[‡]	0.50	−1.69

η is the number of segregating sites; η^s is the number of singletons; π and θ_W are given in 10^4 ± standard deviations.
[†]Statistically significant under the assumption of constant population size (Tajima's D, $P = 0.02$; Fu and Li's D* $P = 0.003$; Fu and Li's F*, $P = 0.01$).
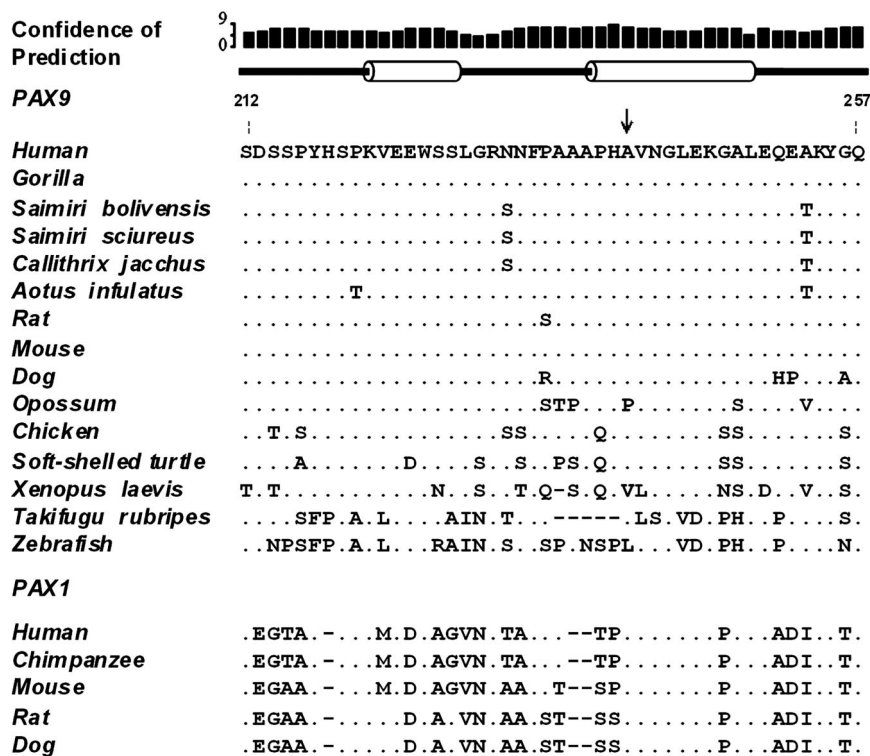[‡]Statistically significant under the assumption of 100-fold population expansion (Tajima's D, $P = 0.008$; Fu and Li's D*, $P = 0.01$; Fu and Li's F*, $P = 0.01$).

To examine the extent of *PAX9* variation, we carried out an analysis of DNA sequence of the entire coding regions (exons 1–4) and exon–intron boundaries of this gene in 86 individuals from three major human geographical groups: Asians, Europeans, and Native Americans. In addition, 144 healthy individuals and 206 patients with different facial/dental anomalies living in Poland were investigated for a mutation in exon 3. We further extended our analysis of exon 3 to four species of New World (NW) monkeys (*Callithrix jacchus*, *Saimiri boliviensis*, *Saimiri sciureus*, and *Aotus infulatus*), and one gorilla (*Gorilla gorilla*). Orthologous *PAX9* sequences available in public databases were also analyzed, and new insights about *PAX9* evolution and its possible contribution to the variation in the dental phenotype within our species and between humans and other primates were evaluated.

Results and Discussion

Sequence Variation of Human *PAX9* Coding Regions. For each of the 86 individuals, we sequenced ≈2.1 kb, including the entire four-exon region (1,026 bp) and ≈1.1 kb of the 5′ UTR region (390 bp), as well as the and the exon–intron boundaries (700 bp). A schematic view of the human *PAX9* is presented in Fig. 2, which is published as supporting information on the PNAS web site. The overall genetic variability of the coding regions was extremely low, as was expected for a pivotal gene in embryogenesis. Of the 172 chromosomes sequenced, 12% showed variations. Curiously, however, these variations are represented by only four distinct mutations, all located in exon 3. Three of these variations were associated with amino acid replacements. We observed a G→T transversion mutation at nucleotide position 677, which results in the substitution of a Ser for Ile at residue 226 (A*G*C→A*T*C, Ser226Ile); a C→A transversion at nucleotide position 751, which results in a substitution of Gln for Lys at amino acid 251 (*C*AG→*A*AG, Gln251Lys); and a G→C transversion at nucleotide 718, which causes a substitution of Ala for Pro at residue 240 (*G*CG→*C*CG, Ala240Pro). In addition, a C→T transition at position 717, resulting in a silent change at residue 239 (CA*C*→CA*T*, His239His), was observed. Both Ser226Ile and Gln251Lys variants are singletons found in Native Americans (Bari and Guarani tribes), whereas His239His and Ala240Pro are previously described polymorphisms (28) but with genotype and allele distributions in different human populations hitherto not available. Larger samples will be necessary to determine whether the two singletons found in Native Americans are isolated mutations or specific population polymorphisms and whether both variants came with the first inhabitants of South America.

Although the His239His polymorphism was observed in one individual (a Kaingang) of 57 Native Americans investigated,

this variant was not detected among the 14 analyzed Asians. By contrast, among the 15 Europeans, 5 were heterozygous and 3 were homozygous for this polymorphism ($\chi^2 = 36.01$; df = 4; $P < 0.001$). The Ala240Pro genotype and allele distributions observed in Native Americans, Asians, and Europeans listed in Table 1 are different among the groups studied ($\chi^2 = 41.6$, df = 4, $P < 0.001$), being more frequent in Europeans as compared to Native Americans and Asians. Patients did not differ significantly from controls. Except for the hypodontia/oligodontia group ($P = 0.02$), the genotype distribution in all samples did not differ significantly from the expected Hardy–Weinberg equilibrium.

Low Genetic Diversity in Coding Regions Indicates Conservation. Human population variability parameters for the entire *PAX9* coding region are summarized in Table 2. Estimates of π and θ_W for the total sample (3.2 and 6.8, respectively) are low in comparison with those reported for other nuclear genes (30). When the combined sample was analyzed under the assumption of constant population size, Tajima's D statistics (-0.98, $P = 0.16$) failed to detect a deviation from the standard neutral model, whereas Fu and Li's D* and F* tests yielded marginal values of departure from it (-1.65, $P = 0.03$; -1.69, $P = 0.08$, respectively). As demonstrated in ref. 31, these three statistics are highly sensitive to assumptions about population growth. Thus, we iteratively tested these indices considering population histories with magnitudes of population growth from 1- to 100-fold and dates of population expansion from 0 to 100,000 years ago (32, 33). Except for Tajima's D, the hypothesis of neutrality was rejected under all population histories assuming 25- to 100-fold expansions, starting 50,000–100,000 years ago. As shown in Table 3, which is published as supporting information on the PNAS web site, all three statistics rejected the hypothesis of neutrality when a 100-fold population expansion starting ≈100,000 years ago was assumed. Although the significantly negative values obtained by these three statistics can reveal a strong bottleneck followed by population expansion, they can also indicate the action of purifying selection.

The division of the total sample into three continental groups reveals that the Native American sample shows significantly negative values of Tajima's D and Fu and Li's D* and F* only under the assumption of constant population size; when population growth was taken into account, the significance disappeared. The opposite was observed in Europeans. For Asians, no deviations were observed under the two models, but it should be remembered that only the Ala240Pro mutation was found in this

Confidence of Prediction

```
9
0
```

PAX9   212   257

Human   SDSSPYHSPKVEEWSSLGRNNFPAAAPHAVNGLEKGALEQEAKYGQ
Gorilla   ....
Saimiri bolivensis ....................S....................T....
Saimiri sciureus   ....................S....................T....
Callithrix jacchus ....................S....................T....
Aotus infulatus   .......T.................................T....
Rat   ....S...................
Mouse   ....
Dog   ..........................R..............HP...A.
Opossum   ......................STP...P.......S....V....
Chicken   ..T.S.............SS.....Q.......SS......S.
Soft-shelled turtle ....A.......D....S...S..PS.Q.......SS......S.
Xenopus laevis   T.T..........N..S..T.Q-S.Q.VL....NS.D..V..S.
Takifugu rubripes ....SFP.A.L....AIN.T..------.LS.VD.PH..P....S.
Zebrafish   ..NPSFP.A.L...RAIN.S..SP.NSPL...VD.PH..P....N.

PAX1

Human   .EGTA.-....M.D.AGVN.TA...--TP.......P...ADI..T.
Chimpanzee   .EGTA.-....M.D.AGVN.TA...--TP.......P...ADI..T.
Mouse   .EGAA.-....M.D.AGVN.AA..T--SP.......P...ADI..T.
Rat   .EGAA.-.....D.A.VN.AA.ST--SS.......P...ADI..T.
Dog   .EGAA.-.....D.A.VN.AA.ST--SS.......P...ADI..T.
```

**Fig. 1.** Structure predictions and multiple alignment of the deduced amino acid sequences of PAX9 exon 3 from several species. Secondary protein structure prediction of the wild-type human exon 3 with PSIPRED (42) is shown above the alignment. Cylinders represent predicted α-helices and lines predicted coiled regions. Confidences indicating the reliability of the prediction are also shown (9 = high and 0 = poor, respectively). The arrow indicates the position where the Ala240Pro polymorphism is located. The alignment of PAX1 homolog regions is also shown.

sample, corroborating the view that strong purifying selection is acting on human *PAX9*.

**Strong Conservation Is also Evident at the *PAX9* Noncoding Regions.** Surprisingly, we found a higher genetic diversity in the human *PAX9* coding regions than in its noncoding regions ($\pi$ = 3.2 and 0.0, respectively). Indeed, only one individual (a Pole) of the 86 sequenced for ≈1.1 kb of noncoding regions presented a nucleotide variation. This G→C transversion was present in heterozygous state and was located 62 bp upstream from exon 3. To investigate the extent of interspecific variation in *PAX9* noncoding regions, we used the human 5′ UTR of exon 1 as a query in a BLAST search. The results revealed 99.5% and 85.9% identity between human and *PAX9* ortholog regions of chimpanzee and mouse, respectively. In addition, comparisons made on the available human genome and the working draft of the chimpanzee genome with the larger human intron 3, which encompasses ≈10 kb, resulted in 99.3% of identity over >8.5 kb. Several recent reports using robust comparative genomic approaches have identified highly conserved noncoding regions among different organisms (34, 35). The role of these conserved sequences is not exactly clear, but functional constraints are likely.

**Interspecific Variation in the Ortholog Regions of Human *PAX9* Exon 3.** Whereas the *PAX9* paired domain (located in exon 2) has been highly conserved in evolution, being 100% identical at the amino acid level in mice and humans (17), no data concerning the variability of other *PAX9* regions are available. Because a protein may be subjected to different pressures at diverse regions of its sequence, adaptative evolution likely affects only a specific region or a few specific sites, whereas most parts of a protein are not allowed to vary (36).

Because we had found variants in only the small human *PAX9*

exon 3 (140 bp, 46 codons), it was of interest to examine this exon in non-human primates and other species. We PCR-amplified this region in four species of NW monkeys and one gorilla. Fig. 1 shows an alignment of the *PAX9* homolog sequences derived from this study compared with those obtained from public databases. A striking conservation is observed. Our data disclosed that humans and gorilla are identical at the nucleotide level, and both are identical with the mouse at the amino acid level. During the roughly 64–74 million years that separated the human and mouse lineages (37), no amino acid replacement occurred at this exon. In contrast, in the last 35- to 48-million-year-old lineage that separates apes/humans and the NW monkeys (38, 39), five differences occurred, three causing amino acid replacements (Fig. 1) and two being silent substitutions. Two of these mutations are fixed in all species of NW monkeys studied here: a substitution of an Ala for Thr at residue 253 (*G*CC→*A*CC, Ala253Thr), and a T→C transition at position 730, which results in a silent change at residue 244 (*T*TG→*C*TG, Leu244Leu). Two other changes are present in three of the NW monkey species investigated (*C. jacchus*, *S. boliviensis*, and *S. sciureus*): a variation at residue 231 (A*A*C→A*G*C, Asn231Ser), and a C→A transversion at position 690 (CG*C*→CG*A*, Arg230Arg). Finally, we detected one private nonsynonymous amino acid change at residue 220 (*C*CC→*A*CC, Pro220Thr) in *A. infulatus*.

In addition, our results indicated that NW monkeys also have several fixed differences in exon/intron boundaries of exon 3. Sixteen variants were identified, including substitutions and insertions in the first 150 nt upstream, and four substitutions in the first 100 nt downstream of this exon (data not shown). Except for the night monkey, all other NW monkeys shared the same mutations. The gorilla sequence differs from the human sequence by only a C→A transversion at nucleotide position 18

downstream of exon 3. This mutation is also fixed in all other NW monkeys. Among the three squirrel monkeys individuals analyzed, no polymorphism was detected.

We measured the nonsynonymous divergence (dN)/synonymous divergence (dS) ratios considering the exon 3 ortholog regions for 14 species of the 15 presented in Fig. 1 (the sequence of *Takifugu rubripes* was not included in this analysis because of the presence of many gaps). In the absence of selective pressures, the dN/dS ratio ($\omega$) is expected to equal 1. Any significant deviation from this prediction can indicate that the studied gene or region is under selective pressure and, thus, is likely to be functional. We also estimated the transition/transversion ratio (R) by using several methods and obtained an R value of $\approx 2$. Using this value, we computed the numbers of dN and dS between pairs of species and the overall mean for all species using the modified Nei–Gojobori method with Jukes–Cantor correction. The dN/dS ratio is expected to be $\ll 1$ if there is selective pressure against nonsynonymous substitutions, because synonymous substitutions are generally assumed to be neutral or nearly so. The overall $\omega$ estimated, 0.18, was significantly different from 1 ($P < 0.001$, Z test), leading to the conclusion that nonsynonymous substitutions in exon 3 are under strong negative selection. To examine the selective pressure at individual sites, the likelihood method implemented in PAML (Phylogenetic Analysis by Maximum Likelihood) with several site-specific models was used. No evidence for a better fit of the M1a (nearly neutral) model, which has two classes of $\omega$ ($\omega_0 < 1$, which is estimated from the data, and $\omega_1 = 1$), as compared with M0 (one-ratio, no variation) was found (likelihood ratio test, df = 1, P = 0.99). Comparison between M1a and M2 ($\omega_0 < 1$, $\omega_1 = 1$, and $\omega_2 > 1$) furnished in all cases codons with $\omega < 1$ (average, $\omega = 0.029$) with posterior probabilities of $>99\%$. The use of more complex models as M3, M7, and M8 gave similar results. Taken together, these data demonstrate that the driving force behind the evolution of *PAX9* exon 3 ortholog regions apparently is intensely purifying selection.

**The Ala240Pro Substitution Likely Affects the PAX9 Protein Structure.** It is not surprising that PAX9, a transcription factor playing a key role in the development of vertebral column (26), limbs, and teeth (27), is well conserved. Such lack of variation in human genes that have pivotal roles in several pathways are well documented (40). Thus, two intriguing questions arise: (*i*) Why is human *PAX9* polymorphic in exon 3 at codon 240? and (*ii*) Why is all variation located in exon 3?

From an evolutionary perspective, it is interesting to note that the Ala at codon 240 is highly conserved among several species (Fig. 1). Furthermore, this residue is also highly conserved in *PAX1*, a paralog that with *PAX9* arose after the amphioxus–vertebrate split $\approx 500$ millions years ago (17).

Although the analysis of selective pressure at individual sites from the 14 species indicates that ortholog regions of residue 240 of human PAX9 has a $\omega < 1$ with a Bayesian posterior probability of $>99\%$, this finding does not completely rule out a pattern of selective pressure other than purifying selection in the human lineage only. In fact, a strong purifying selection at this codon in most lineages might be masking a selective pressure specific for the human lineage. Hence, an important question is whether the human Ala240Pro polymorphism is a neutral or nearly neutral mutation or whether it might be maintained by the action of some form of positive selection.

The Ala240Pro polymorphism is located in a region homologous to one in the zebrafish, where a putatively potent trans-activating domain was reported in ref. 41. Because no PAX9 three-dimensional structure is available, we predicted its secondary protein structure using the PSIFRED (42), SAM-T99 (43) and PSA (44) programs. These programs indicated a high probability for the existence of an $\alpha$-helix in the region of amino acids

238–250. Interestingly, Pro is rarely found in $\alpha$-helix structures (45) because it seems to be a potent $\alpha$-helix breaker in soluble (globular) proteins, its chemical properties therefore being detrimental for $\alpha$-helix conformation (46).

In contrast, residue 240 could be localized in or near the first turn of the $\alpha$-helix, the N-capping residue. The latter is preceded by two amino acid residues termed N'' and N', respectively, which act as a stop signal to helix propagation (46). It was reported that Pro is only common at the N'' position (47) but may eventually play a role in N-capping (47, 48). Consequently, the Pro replacement might shorten the $\alpha$-helix length. Because $\alpha$-helices become less stable as chain lengths decrease (48), homozygotes for the Pro allele might present a PAX9 protein with a slightly reduced DNA-binding capacity. Inadequate nuclear transport of a protein with an abnormal $\alpha$-helix conformation can also be considered (49).

*Pax9* was identified as a marker for tooth development in mice, where it was found to be a marker for prospective tooth mesenchyme before the first morphological manifestation of odontogenesis, marking the areas where tooth will develop. *Pax9* expression has not just an earlier onset but, in addition, is expressed in higher levels in the prospective molars compared with prospective incisors. At this stage, not only *Pax9* but also another transcription factor, *Msx1*, is expressed. At the bud stage, *Msx1* and *Pax9* are coexpressed in an overlapping manner, contributing to the mesenchyme odontogenic potential. Both genes are required for the expression of *BMP4*, which is involved in the induction of the enamel knot, a transient signaling center of the epithelium that directs the next phase of tooth development (18). It was verified that the larger nuclear protein PLU-1 and PAX9 interact *in vivo* through a conserved Ala-X-Ala-Ala-X-Val-Pro-X$_4$-Val-Pro-X$_8$-Pro sequence motif. The replacement of a Pro for Ala via site-directed mutagenesis of this motif completely abolished the PLU-1-PAX9 interaction (50). A recent study also reported that Pax9 and Msx1 interact physically *in vivo*, leading to the possibility of physical PAX9 interactions with several others factors in tooth development (51). These findings suggest other possibility, that any subtle modification in PAX9 structure might modify PAX9 protein–protein interactions.

**Ala240Pro Is Not Associated with Hypodondotia/Oligodontia or Cleft Lip/Palate but May Cause Congenitally Missing Third Molars.** As mentioned above, *PAX9* plays a critical role in craniofacial development. Mice homozygous for a *Pax9* deletion die shortly after birth because of breathing problems and exhibit a wide range of developmental defects (27). These animals have secondary cleft palate, facial abnormalities, and complete anodontia with an arrest of tooth development at the bud stage. There is also strong evidence for a direct correlation of *Pax9* expression with palatal morphogenesis and fusion (52). However, no association was found in the present study with a variant in this region and hypodontia/oligodontia or cleft lip/palate in a Polish population (Table 1). A recent family study also showed that the Ala240Pro mutation did not seem to produce an entirely consistent pattern of tooth agenesis but verified that all homozygotes for this mutation had missing third molars (49).

**Could the Ala240Pro Mutation Be Advantageous in Humans?** A detrimental mutation at the protein level may not necessarily be associated with a disadvantageous phenotype at the organismal level. A good example is the recently reported human sarcomeric myosin gene (*MYH16*) expressed in the masticatory muscles of most primates (53). A frameshift deletion at codon 660 of the human *MYH16* gene truncates the predicted 224-kDa myosin heavy chain. As a result, a 76-kDa fragment containing an unstable portion of the myosin head domain is formed, leading to hypotrophy of $>80\%$ of the human type II fibers of the masticatory muscles as compared

with macaque. Apparently, such mutation could be extremely deleterious, because powerful masticatory muscles are essential for survival in most primates. However, because the appearance and fixation of this inactivating mutation in a hominid ancestor ≈2.4 million years ago, a diminished contractile force was likely translated into a reduction in the stress across patent sutures, sites of the dura-mater-patterned growth in the immature neurocranium, presumably giving a rise to a process of enhanced encephalization, which is of obvious importance for humans. Indeed, the *MYH16* frameshift deletion at codon 660 is fixed and was only found in our species (53).

In the cases for which a reduced number of teeth might be beneficial (54), the replacement of an Ala for a Pro in *PAX9* codon 240, slightly decreasing the DNA-binding activity/ stability/mobility, might be advantageous if such a mutation had a specific effect in reducing the number of third molars. It is tempting to speculate that Pro carriers, harboring a slightly altered structure of the transcription factor PAX9, might have normal first and second molars dentition because of the putative overlapping expression of *MSX1* and *PAX9* during tooth development (18). Nevertheless, they might fail to form third molars beyond the bud stage, because the maintenance of *BMP4* expression (or another signal essential to the next round of tooth formation) in prospective third molars may depend predominantly (or exclusively) on *PAX9* expression and/or physical interaction with this transcription factor (55).

A reduced number of molars may be advantageous from a human evolutionary perspective. Because of the dramatic lifestyle and diet shift experienced since the discovery of fire and the development of cooking utensils, third molars, which could have been essential for the survival of earlier hominids, became not only functionless but also an important cause of morbidity for modern humans (54). Dental arches have been reduced over hominid evolution (2–4, 56). As a result, third molars became frequently impacted or malpositioned, preventing the teeth from attaining a functional position. Furthermore, because of the difficulty of cleaning them and keeping them free of disease, impacted or malpositioned third molars lead to a higher susceptibility to periodontal disease, such as infections, carious lesions, cysts, tumors, and destruction of adjacent teeth and bone (57). These problems led some authors to raise the possibility of a complete elimination of these teeth in humans (54), and studies for intentionally stopping third molar growth for possible clinical application are being carried out (58).

### Conclusions

In summary, our data show a remarkable conservation of the *PAX9* coding sequence and a complete conservation of ≈1.1 kb of noncoding regions, including the 5′ UTR of exon 1 and exon–intron boundaries in three major human geographical groups. Our observations also indicate that the Ala240Pro polymorphism in humans is unlikely to be neutral, suggesting the intriguing possibility that the Pro variant might represent an advantageous transient polymorphism. Finally, it is notable that all DNA variations were located in exon 3, which can represent an evolutionary window (in an extremely conserved gene) to functional adaptations in the primate dentition.

### Materials and Methods

**Populations Examined and Sequence Data.** DNA sequences from the entire *PAX9* coding region (1,026 bp, 341 aa) were obtained from 86 unrelated, healthy individuals from three major human geographical groups: Europeans (9 Spanish and 6 Polish), Asians (14 Japanese), and 57 Native Americans from the Guarani ($n = 5$), Kaingang ($n = 20$), Yucpa ($n = 4$), Bari ($n = 4$), Warao ($n = 4$), and Aché ($n = 20$) tribes. Appropriate informed consent was obtained from all of them. Additional information about these human samples can be obtained in previous reports (59–62). Six

samples from four species of NW primates, including three *S. sciureus* (squirrel monkey), one *S. boliviensis* (Bolivian squirrel monkey), one *A. infulatus* (night monkey), and one *C. jacchus* (common marmoset) as well as one sample of the Great Apes (*G. gorilla*), were also analyzed for exon 3. The NW monkey samples were collected at the National Primates Center in Ananindeua, Brazil, whereas the gorilla sample was kindly provided by the Belo Horizonte Zoo (Belo Horizonte, Brazil). The following published *PAX9* nucleotide sequences (with accession numbers in parentheses) were retrieved from GenBank: zebrafish (U40931), mouse (X84000), chicken (X92652), and *Pelodiscus sinensis*, Chinese soft-shelled turtle (AB181136). Additional predicted ortholog regions of human exon 3 were extracted using the Ensembl database (www.ensenbl.org) and TBLASTN ($E < 10^{-10}$). Sequence orthology was further investigated by BLAST using the protein sequence of human exon 3 as a query.

To determine the putative involvement of the common Ala240Pro polymorphism in the etiology of selective tooth agenesis and cleft lip with or without cleft palate in the Polish population, we collected peripheral blood samples from 52 unrelated patients with hypodontia and oligodontia, 154 patients with cleft lip with or without cleft palate, and 144 healthy individuals. All affected individuals were patients from the Department of Orthodontics, University of Medical Sciences in Poznan and the Department of Plastic Surgery, University of Medical Sciences in Wroclaw, Poland. The inclusion criterion for patients with the lack of permanent teeth was agenesis of at least one of the permanent teeth, excluding third molars, as verified by panoramic x-ray analysis and dental history. The control group consisted of healthy individuals with normal permanent dentition and that did not exhibit any craniofacial abnormalities. All subjects signed informed consent forms agreeing with the investigation.

**PCR and DNA Sequencing.** Full details about primers and PCR conditions are available in *Supporting Materials and Methods* and Tables 4 and 5, which are published as supporting information on the PNAS web site. Briefly, we designed a set of specific PCR pair primers to amplify the entire four-exon human *PAX9* coding regions, including exon 1 untranslated region (390 bp) and ≈100 bp of intronic boundaries. The same primers were used for DNA amplification of the NW monkeys and gorilla. PCR products were purified with ExoSAP-IT (United States Biochemical) and were direct-sequenced on both strands using a BigDye Terminator kit (Applied Biosystems) on an ABI377 automatic sequencer (Applied Biosystems). For the Polish patients and controls, the genotypes were determined by restriction fragment length polymorphism analysis (MspI restriction enzyme). Digestion products were analyzed by PAGE.

**Data Analysis.** Two diversity indices were calculated: Watterson's $\theta_W$, based on the number of segregating sites in the sample (63), and $\pi$, the average number of pairwise differences in the sample (64). To detect departures from a standard neutral model, we performed Tajima's D (65) and Fu and Li's D* and F* tests (66). *P* values for these statistics were estimated from $10^4$ coalescent simulations. Size population changes were incorporated with Rogers' algorithm (67) as described in ref. 31. For the interspecific data, we used the modified Nei–Gojobori method with Jukes–Cantor correction to estimate the dN/dS ratio ($\omega$), where dN is the number of nonsynonymous substitutions ($n$) per nonsynonymous site ($N$) and dS is the number of synonymous substitutions ($s$) per synonymous site ($S$). To detect selective pressure at individual sites, we used the likelihood methods implemented in the PAML package (68), version 3.14a, with several site-specific models (36, 69).

1. Fleagle, J. G. (1999) *Primate Adaptation and Evolution* (Academic, New York).
2. Wolpoff, M. H. (1975) *Am. J. Phys. Anthropol.* **43,** 307–308.
3. Calcagno, J. M. (1986) *Am. J. Phys. Anthropol.* **70,** 349–363.
4. Calcagno, J. M. & Gibson, K. R. (1988) *Am. J. Phys. Anthropol.* **77,** 505–517.
5. Dahlberg, A. (1945) *J. Am. Dent. Assoc.* **32,** 676–690.
6. Almeida, R. (1949). *Contribuição para o Estudo de Alguns Caracteres Dentários dos Indígenas da Luanda* (Publicações Culturais, Lisbon).
7. Rozkovcova, E., Markova, M. & Dolejsi, J. (1999) *Sbornik Lekarsky* **100,** 71–84.
8. Lundstrom, A. (1963) *Am. J. Hum. Genet.* **15,** 34–43.
9. Salazar-Ciudad, I. & Jernvall, J. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 8116–8120.
10. Peters, H. & Balling, R. (1999) *Trends Genet.* **15,** 59–65.
11. Garn, S. M., Lewis A. B. & Vicinus, J. H. (1962) *J. Dent. Res.* **41,** 717.
12. McKee, J. K. (1984) *Am. J. Phys. Anthropol.* **65,** 231–241.
13. Peters, H., Neubuser, A. & Balling, R. (1998) *Eur. J. Oral Sci.* **106,** Suppl. 1, 38–43.
14. Lammi, L., Arte, S., Somer, M., Jarvinen, H., Lahermo, P., Thesleff, I., Pirinen, S. & Nieminen, P. (2004) *Am. J. Hum. Genet.* **74,** 1043–1050.
15. Mostowska, A., Kobielak, A. & Trzeciak, W. H. (2003) *Eur. J. Oral Sci.* **111,** 365–370.
16. Xu, W., Rould, M. A., Jun, S., Desplan, C. & Pabo, C. O. (1995) *Cell* **80,** 639–650.
17. Balczarek, K. A., Lai, Z. C. & Kumar, S. (1997) *Mol. Biol. Evol.* **14,** 829–842.
18. Neubuser, A., Peters, H., Balling, R. & Martin, G. R. (1997) *Cell* **90,** 247–255.
19. Neubuser, A., Koseki, H. & Balling, R. (1995) *Dev. Biol.* **170,** 701–716.
20. Ogasawara, M., Shigetani, Y., Hirano, S., Satoh, N. & Kuratani, S. (2000) *Dev. Biol.* **223,** 399–410.
21. Watanabe, T., Tarttelin, E., Neubuser, A., Kimura, M. & Solter, D. (1994) *Mamm. Genome* **5,** 768–770.
22. Mensah, J. K., Ogawa, T., Kapadia, H., Cavender, A. C. & D'Souza, R. N. (2004) *J. Biol. Chem.* **279,** 5924–5933.
23. Stockton, D. W., Das, P., Goldenberg, M., D'Souza, R. N. & Patel, P. I. (2000) *Nat. Genet.* **24,** 18–19.
24. Chi, N. & Epstein, J. A. (2002) *Trends Genet.* **18,** 41–47.
25. Muller, T. S., Ebensperger, C., Neubuser, A., Koseki, H., Balling, R., Christ, B. & Wilting, J. (1996) *Dev. Biol.* **178,** 403–417.
26. Peters, H., Wilm, B., Sakai, N., Imai, K., Maas, R. & Balling, R. (1999) *Development (Cambridge, U.K.)* **126,** 5399–5408.
27. Peters, H., Neubuser, A., Kratochwil, K. & Balling, R. (1998) *Genes Dev.* **12,** 2735–2747.
28. Nieminen, P., Arte, S., Tanner, D., Paulin, L., Alaluusua, S., Thesleff, I. & Pirinen, S. (2001) *Eur. J. Hum. Genet.* **9,** 743–746.
29. Das, P., Stockton, D. W., Bauer, C., Shaffer, L. G., D'Souza, R. N., Wright, T. & Patel, P. I. (2002) *Hum. Genet.* **110,** 371–376.
30. Sachidanandam, R., Weissman, D., Schmidt, S. C., Kakol, J. M., Stein, L. D., Marth, G., Sherry, S., Mullikin, J. C., Mortimore, B. J., Willey, D. L., *et al.* (2001) *Nature* **409,** 928–933.
31. Wooding, S., Kim, U. K., Bamshad, M. J., Larsen, J., Jorde, L. B. & Drayna, D. (2004) *Am. J. Hum. Genet.* **74,** 637–646.
32. Excoffier, L., Smouse, P. E. & Quattro, J. M. (1992) *Genetics* **131,** 479–491.
33. Stiner, M. C., Munro, N. D., Surovell, T. A., Tchernov, E. & Bar-Yosef, O. (1999) *Science* **283,** 190–194.
34. Gibbs, R. A., Weinstock, G. M., Metzker, M. L., Muzny, D. M., Sodergren, E. J., Scherer, S., Scott, G., Steffen, D., Worley, K. C., Burch, P. E., *et al.* (2004) *Nature* **428,** 493–521.
35. Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W. J., Mattick, J. S. & Haussler, D. (2004) *Science* **304,** 1321–1325.
36. Yang, Z., Nielsen, R., Goldman, N. & Pedersen, A. M. (2000) *Genetics* **155,** 431–449.
37. Eizirik, E., Murphy, W. J. & O'Brien, S. J. (2001) *J. Hered.* **92,** 212–219.
38. Kumar, S. & Hedges, S. B. (1998) *Nature* **392,** 917–920.
39. Chiu, C. H., Schneider, H., Schneider, M. P., Sampaio, I., Meireles, C., Slightom, J. L., Gumucio, D. L. & Goodman, M. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 6510–6515.
40. Sale, S., Sung, R., Shen, P., Yu, K., Wang, Y., Duran, G. E., Kim, J. H., Fojo, T., Oefner, P. J. & Sikic, B. I. (2002) *Mol. Cancer Ther.* **1,** 215–225.
41. Nornes, S., Mikkola, I., Krauss, S., Delghandi, M., Perander, M. & Johansen, T. (1996) *J. Biol. Chem.* **271,** 26914–26923.
42. McGuffin, L. J., Bryson, K. & Jones, D. T. (2000) *Bioinformatics* **16,** 404–405.
43. Karplus, K., Barrett, C. & Hughey, R. (1998) *Bioinformatics* **14,** 846–856.
44. White, J. V., Stultz, C. M. & Smith, T. F. (1994) *Math. Biosci.* **119,** 35–75.
45. Chou, P. Y. & Fasman, G. D. (1974) *Biochemistry* **13,** 222–245.
46. Harper, E. T. & Rose, G. D. (1993) *Biochemistry* **32,** 7605–7609.
47. Parker, M. H. & Hefford, M. A. (1997) *Protein Eng.* **10,** 487–496.
48. Rohl, C. A., Scholtz, J. M., York, E. J., Stewart, J. M. & Baldwin, R. L. *Biochemistry* **31,** 1263–1269.
49. Trimmell, J. B. (2004) M.Sc. thesis (University of Missouri, Kansas City).
50. Tan, K., Shaw, A. L., Madsen, J., Jensen, K., Taylor-Papadimitriou, J. & Freemont, P. S. (2003) *J. Biol. Chem.* **278,** 20507–20513.
51. Ogawa, T., Kapadia, H., Wang, B. & D'Souza, R. N. (2005) *Arch. Oral Biol.* **50,** 141–145.
52. Hamachi, T., Sasaki, Y., Hidaka, K. & Nakata, M. (2003) *Arch. Oral Biol.* **48,** 581–587.
53. Stedman, H. H., Kozyak, B. W., Nelson, A., Thesier, D. M., Su, L. T., Low, D. W., Bridges, C. R., Shrager, J. B., Minugh-Purvis, N. & Mitchell, M. A. (2004) *Nature* **428,** 415–418.
54. Silvestri, A. R. & Singh, I. (2003) *J. Am. Dent. Assoc.* **134,** 450–455.
55. Kist, R., Watson, M., Wang, X., Cairns, P., Miles, C., Reid, D. J. & Peters, H. (2005) *Hum. Mol. Genet.* **14,** 3605–3617.
56. Macho, G. A. & Moggi-Cecchi, J. (1992) *Am. J. Phys. Anthropol.* **87,** 151–159.
57. Song, F., O'Meara, S., Wilson, P., Golder, S. & Kleijnen, J. (2000) *Health Technol. Assess.* **4,** 1–55.
58. Silvestri, A. R., Connolly, R. J. & Higgins, M. T. (2004) *J. Am. Dent. Assoc.* **135,** 1397–1405.
59. Mostowska, A., Kobielak, A., Biedziak, B. & Trzeciak, W. H. (2003) *Eur. J. Oral Sci.* **111,** 272–276.
60. Bortolini, M. C., Salzano, F. M., Thomas, M. G., Stuart, S., Nasanen, S. P., Bau, C. H., Hutz, M. H., Layrisse, Z., Petzl-Erler, M. L., Tsuneto, L. T., *et al.* (2003) *Am. J. Hum. Genet.* **73,** 524–539.
61. Bortolini, M. C., Thomas, M. G., Chikhi, L., Aguilar, J. A., Castro-de-Guerra, D., Salzano, F. M. & Ruiz-Linares, A. (2004) *Genet. Mol. Biol.* **27,** 1–8.
62. Silva, W. A., Jr., Bortolini, M. C., Meyer, D., Salzano, F. M., Elion, J., Krishnamoorthy, R., Schneider, M. P., Castro-de-Guerra, D., Layrisse, Z., Castellano, H., *et al.* (1999) *Am. J. Phys. Anthropol.* **109,** 425–437.
63. Watterson, G. A. (1975) *Theor. Popul. Biol.* **7,** 256–276.
64. Nei, M. & Kumar, S. (2000) *Molecular Evolution and Phylogenetics* (Oxford Univ. Press, Oxford).
65. Tajima, F. (1983) *Genetics* **105,** 437–460.
66. Fu, Y. X. & Li, W. H. (1993) *Genetics* **133,** 693–709.
67. Rogers, A. R. (1995) *Evolution* 608–615.
68. Yang, Z. (1997) *Comput. Appl. Biosci.* **13,** 555–556.
69. Yang, Z. & Nielsen, R. (2002) *Mol. Biol. Evol.* **19,** 908–917.

**ANTHROPOLOGY**