ARCHIVES OF ORAL BIOLOGY 52 (2007) 1026–1031

available at www.sciencedirect.com

ScienceDirect

journal homepage: www.intl.elsevierhealth.com/journals/arob

A study of polymorphism in human *AMELX*

Béatrice Richard [a,b], *Sidney Delgado* [c], *Philippe Gorry* [b], *Jean-Yves Sire* [c,*]

[a] *UFR d'Odontologie, 16 cours de la Marne, 33082 Bordeaux, France*
[b] *Service de Génétique, CHU Bordeaux, place Amélie Raba-Léon, 33076 Bordeaux, France*
[c] *Equipe "Evolution & Developpement du Squelette", UMR 7138-SAE, Université Pierre & Marie Curie-Paris 6, 7 quai St-Bernard, Case 05, 75005 Paris, France*

ARTICLE INFO

Article history:
Accepted 14 June 2007

Keywords:
Amelogenin
Polymorphism
Evolution
Sex chromosome

ABSTRACT

Amelogenin gene (*AMEL*) encodes for a protein that plays important roles in the organization and structure of enamel. A recent evolutionary analysis of AMELX in mammals has revealed, aside to well-conserved 5′ and 3′ regions, a variable region located in the largest exon (exon 6), which strongly suggested the possible existence of polymorphism in human *AMELX*. A detailed analysis of this region was of fundamental importance for genetic studies.

We have looked for variations in human *AMELX* exon 6 from 100 *AMELX* alleles in a randomized European population, using denaturing high-performance liquid chromatography (dHPLC). We also have looked for *AMELX* variants in databases, and compared this region in nine primates.

There were no variations in the *AMELX* sequences analysed, but two synonymous single-nucleotide polymorphisms were found in databases. Alignment of the primate exon 6 sequences revealed that *AMELX* is highly constrained, as illustrated by 100% nucleotide identity found between humans and chimpanzee, and from 99.9 to 94.8% nucleotide identity in the other species.

In contrast to what was suspected from the evolutionary analysis, we conclude that *AMELX* polymorphism should occur at low level in humans. This finding leads us to speculate that the high constraint observed in primate *AMELX* is related to its location on the X chromosome, and is due to selection at a single locus.

1. Introduction

Amelogenin (AMEL) is the major matrix protein of forming enamel, in which it represents 90% of the organic content.[1] In humans, AMEL is encoded by two genes located on the sex chromosomes, *AMELX* and *AMELY*. In males, AMELY is expressed at a low level.[2] It is not under strong functional constraint and is masked by the expression of the X-linked homolog. AMELX is, therefore, the main actor for the organization and structure of this highly mineralized tissue. These important roles are demonstrated by several mutations of *AMELX* leading to X-linked amelogenesis imperfecta (AIH1) with various hypoplastic or hypomineralized phenotypes,[3–5] and prove that AMELY does not provide protection.

A recent evolutionary analysis of AMEL sequences in representative species of the main mammalian lineages has revealed that the various regions of this protein have evolved differently during approximately 200 million years.[6–8] The N- and C-terminal regions are well conserved, while the central region (encoded by exon 6) is variable. The evolutionary-constrained regions indicate that important functions are supported by specific residues or group of residues.[9] In the

* *Corresponding author*. Tel.: +33 1 44 27 35 72; fax: +33 1 44 27 35 72.
E-mail address: sire@ccr.jussieu.fr (J.-Y. Sire).
0003–9969/$ – see front matter
doi:10.1016/j.archoralbio.2007.06.001

N-ter region, deletions and/or substitutions are scarce, and several of the 15 *AMEL* mutations reported so far to lead to AIH1 are strictly related to this region.[3–5] The other mutations (either in the N-ter or in the region encoded by exon 6) provoke frame shifts, which change the amino acids encoded downstream. The residues located in the variable region are also changed but, given the crucial role of many residues that are modified, either in the N- or in the C-ter regions, it seems unlikely that changes in the variable region itself could cause AIH1. To date no AIH1 has been demonstrated to be directly related to specific mutations occurring in this region. In contrast, in this less constrained region, indels and substitutions are numerous, and most of the residues have been changed during evolution without any problem for the protein function.[9]

The high variability of the central region in mammalian sequences strongly suggests the possible existence of polymorphism in human *AMEL*. Such an hypothesis is supported by several examples, which show a relationship between a highly variable region in mammalian evolution and the presence of polymorphisms in this region in humans. For instance, such a relationship does exist for another enamel protein, enamelin (ENAM), in which the large exon 10 is variable (Hu and Yamakoshi, 2003; Sire et al., unpublished results)[17]. In human *ENAM*, 12 SNPs are found in databases (http://www.ncbi.nlm.nih.gov/SNP/).

To date, there is no study focusing on this question for *AMELX*. The genetic diagnostic of AIH1 will certainly increase in the coming years and the analysis of variation in this region of the coding *AMELX* sequence is of fundamental importance for the validation of the reported substitutions.

The aim of the present study was to test the hypothesis of the presence of polymorphism in the central region of amelogenin (exon 6). To this end we have (i) searched in databases the presence of *AMEL* variants, (ii) studied this particular region in a human population (100 alleles) using dHPLC (denaturing high-performance liquid chromatography), and (iii) compared this region to other primate sequences. Surprisingly, we have found that this particular *AMELX* region, although highly variable in mammalian evolution, is not subjected to polymorphism in humans, indicating that human *AMELX* is highly constrained.

2. Material and methods

2.1. Materials

Genomic DNA was obtained randomly from a human population. The latter was composed of 33 females (2× 33 *AMELX*) and 34 males (1× 34 *AMELX*). Most of them were from European origin and they did not present a AIH1 phenotype. This sample provided 100 alleles, which allows statistical analysis.

2.2. Sequences in databases

NCBI and EnsEMBL databases were screened for single-nucleotide polymorphisms (SNPs). Only coding synonymous and nonsynonymous SNPs were retained. The human *AMELX* coding sequence was used as the reference sequence (accession number: AF436849). The following primate sequences were found in databases (some were checked and completed): chimpanzee, *Pan troglodytes* (AB091781, amended: EF537869); rhesus monkey, *Macaca mulatta* (ENSMMUG00000008256, amended: EF537871); squirrel monkey, *Saimiri sciureus* (AB091783); bushbaby, *Otolemur garnetti* (AB091787); ring-tailed lemur, *Lemur catta* (AB091785). Three sequences were obtained using in silico analysis: orangutan, *Pongo pygmaeus* (EF537870); marmoset, *Callithrix jacchus* (EF537872); tarsier, *Tarsius syrichta* (EF537873).

2.3. Methods

2.3.1. PCR reaction

Two primers (sense and antisense) were designed (using Primer3 software) on both sides of the target region to obtain a 210 bp long sequence (Fig. 1). sense: GATCCCCAGCAACCAA; antisense: CTGCATGGGGAACATCGG.

These primers are specific for *AMELX* sequences and cannot link *AMELY* sequences at the defined annealing temperature (*AMELY* variants are underlined in the above primer sequences).

Genomic DNA (25 ng/μl, 2 μl) was amplified in a mixture composed of 2.5 μl GeneAmp® PCR Buffer II (10×) pH 8.3, 1 μl $MgCl_2$ 25 mM, and 1 μl dNTP 5 mM, in the presence of sense and antisense primers, and 0.2 μl AmpliTaq™ Gold DNA Polymerase (Applied Biosystems). Amplification was performed in a thermocycler (Primus, MWG Biotech Inc.) The reaction started with an initial denaturation of 7 min at 95 °C, followed by 40 cycles, each cycle consisting of 30 s of denaturation at 94 °C, 30 s of annealing at 53 °C, and 1 min of extension at 72 °C. The final extension was for 5 min at 72 °C.

2.3.2. dHPLC screening, and DNA sequencing

dHPLC allows to detect SNPs with sensitivity and specificity exceeding 96%[10] The amplicons were denatured and analysed by dHPLC in specific conditions predicted by computation (Fig. 2): 66 °C, buffer B: 53%, TS = 0.

DNA sequencing was performed from six DNA samples taken at random within the human population examined, using the dideoxynucleotide chain termination reaction (Big Dye™ Kit, Applied Biosystems).

2.3.3. Sequence alignment

Alignment of primate *AMELX* sequences was done by hand using Se–Al software.

3. Results

3.1. AMELX SNPs in databases

Two SNPs are reported in the databases for *AMELX* coding sequences in humans (Fig. 1). rs2106416 is a C/T synonymous substitution (His/His), amino acid 87. This nucleotide substitution is located in exon 6, but not in the highly variable region examined. The other SNP, rs 6639060 is also a C/T synonymous substitution (Leu/Leu), amino acid 152. This substitution is located in the region studied.

```
 M  G  T  W  I  L  F  A  C  L  L  G  A  A  F  A  M  P| L  P  P  H  P  G  H  (25)
ATGGGGACCTGGATTTTATTTGCCTGCCTCCTGGGAGCAGCTTTTGCCATGCCTCTACCACCTCATCCTGGGCAC
1                                                                          75
 P  G  Y  I  N  F  S  Y  E| N  S  H  S  Q  A  I  N  V  D  R  T  A  L| V  L  (50)
CCTGGTTATATCAACTTCAGCTATGAGAACTCACATTCTCAGGCTATCAATGTTGACAGGACTGCATTAGTGCTT
76                                                                        150
 T  P  L  K  W  Y  Q  S  I  R  P  P| Y  P  S  Y  G  Y  E  P  M  G  G  W  L  (75)
ACCCCTTTGAAGTGGTACCAGAGCATAAGGCCACCGTACCCTTCCTATGGTTACGAGCCCATGGGTGGATGGCTG
151                                                                       225
 H  H  Q  I  I  P  V  L  S  Q  Q  H  P  P  T  H  T  L  Q  P  H  H  H  I  P  (100)
CACCACCAAATCATCCCCGTGCTGTCCCAACAGCACCCCCCGACTCACACCCTGCAGCCTCATCACCACATCCCA  C/T
226                                                                       300
 V  V  P  A  Q  Q  P  V  I  P  Q  Q  P  M  M  P  V  P  G  Q  H  S  M  T  P  (125)
GTGGTGCCAGCTCAGCAGCCCGTGATCCCCCAGCAACCAATGATGCCCGTTCCTGGCCAACACTCCATGACTCCA
301                                                                       375
 I  Q  H  H  Q  P  N  L  P  P  P  A  Q  Q  P  Y  Q  P  Q  P  V  Q  P  Q  P  (150)
ATCCAACACCACCAGCCAAACCTCCCTCCGCCCGCCCAGCAGCCCTACCAGCCCCAGCCTGTTCAGCCACAGCCT
376                                                                       450
 H  Q  P  M  Q  P  Q  P  P  V  H  P  M  Q  P  L  P  P  Q  P  P  L  P  P  M  (175)
CACCAGCCCATGCAGCCCCAGCCACCTGTGCACCCCATGCAGCCCCTGCCGCCACAGCCACCTCTGCCTCCGATG  C/T
451                                                                       525
 F  P  M  Q  P  L  P  P  M  L  P  D  L  T  L  E  A  W  P  S  T  D  K  T  K  (200)
TTCCCCATGCAGCCCCTGCCTCCCATGCTTCCTGATCTGACTCTGGAAGCTTGGCCATCAACAGACAAGACCAAG
526                                                                       600
 R  E  E  V| D  *  (225)
CGGGAGGAAGTGGATTAA
601             618
```

Fig. 1 – Human AMELX sequence. Underlined: location of the primers; in grey background: target region amplified in exon 6 corresponding to the highly variable region in mammals; boldface underlined: two synonymous SNPs found in databases (His/His: rs2106416; Leu/Leu: rs6639060).

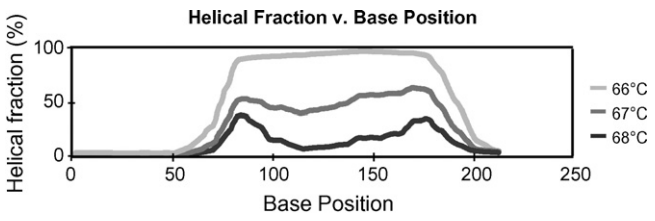

Fig. 2 – Modelisation of the denaturation of the AMEL DNA for dHPLC analysis allowing to determine the appropriate temperature (66 °C) to obtain 90% of hybridized DNA molecules.

3.2. *dHPLC*

No heteroduplexes were identified in the sample of 100 *AMELX* alleles studied. This means that no SNP were present in this particular region of *AMELX* exon 6. In addition, sequencing six of these *AMELX* at random did not reveal any mutation, confirming the dHPLC results.

3.3. *Sequence comparison in primates*

The alignment of the nine coding sequences of primate *AMELX* exon 6 reveals that the target region is highly conserved (Fig. 3). Compared to human *AMELX* (366 bp), we found 100% nucleotide identity in the chimpanzee sequence, 99.45% in

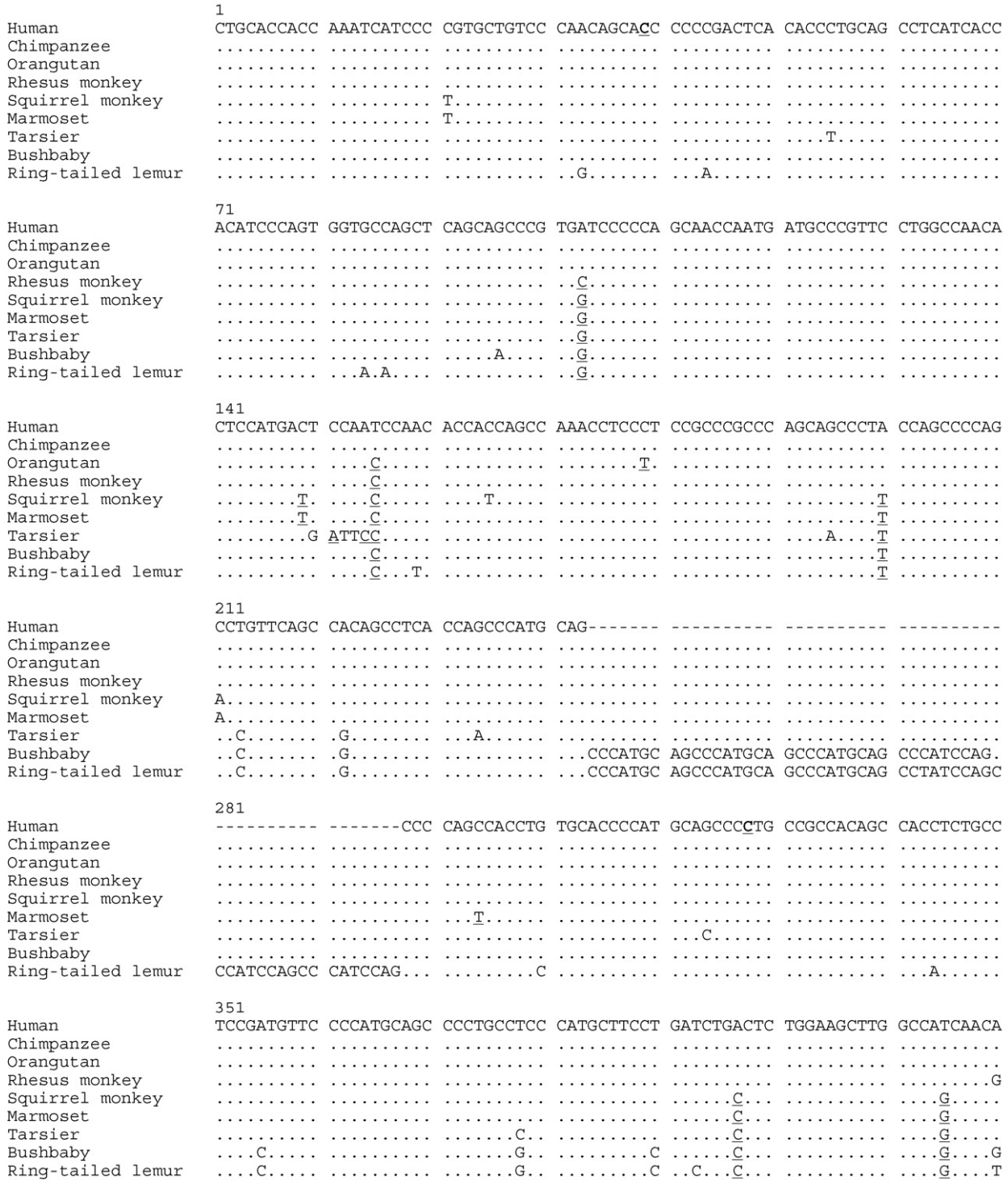

Fig. 3 – Alignment of the variable region of *AMELX* exon 6 of nine primates (420 bp corresponding to positions 223–578 of the entire human sequence). Note the presence of a large insertion (bp 243–297) in the two lemuriform *AMELX*. (˙) = identical residue; (—) lacking residue. Underlined: nonsynonymous substitutions. Bold: the two polymorphisms reported for human *AMELX*.

orangutan (2 substitutions: 2 nonsynonymous), 99.18% in rhesus-monkey (3:2 nonsynonymous), 97.54% (9:6) in squirrel monkey and marmoset (9:7), 96.72% (12:6) in bushbaby, 95.36% (17:7) in tarsier and 94.81% (19:11) in ring-tailed lemur.

The only large difference concerns the two lemuriform (busbaby and ring-tailed lemur) sequences, which exhibit a large insertion in the same locus (36 bp in galago and 54 bp in lemur) (Fig. 4).

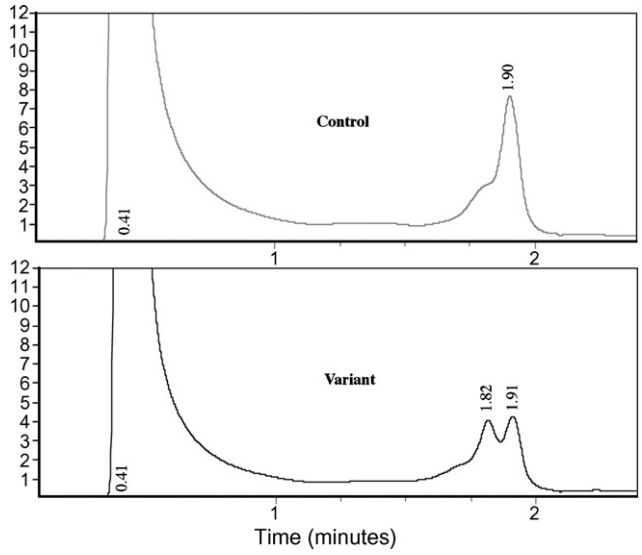

Fig. 4 – Example of identification of a variant using dHPLC; here a variant detected in a DSPP sequence.

4. Discussion

4.1. *Lack of polymorphism in human AMELX*

Our evolutionary-supported hypothesis of probable polymorphism in human *AMELX* exon 6 is not confirmed by the study of 100 randomized alleles in a European population using dHPLC analysis, a highly sensitive technique (Fig. 4). This absence of variation does not mean that variants may not be found in this region of *AMELX* in a world-wide population, but statistically they might occur at a low percentage. Indeed, to date only two SNPs are reported in databases for the *AMELX* coding sequence in humans. rs2106416 (C/T, synonymous) was detected in the course of a study of polymorphism in three human populations (42 Caucasian, 42 Asian and 42 black American individuals). rs6639060 (C/T, synonymous) was detected during the human sequencing project by comparing chromosome and DNA specific sequences, but the number of alleles tested is unknown. Recently, in an evolutionary study of sex-linked mammalian AMEL genes, the only substitution reported in exon 6 of a worldwide sample of 32 human AMELX (Coriell Cell Repositories, USA) was one of the two SNP reported above.[11] This means that such a polymorphism is well spread in humans.

We found neither of these two C/T substitutions in the 100 alleles from the European population studied. These two SNPs do not change the amino acid, which confirms that *AMELX* is highly constrained in humans.

It appears that the only mean to find a polymorphism in human *AMELX* is to increase the sampling of human DNA from various regions of the world. Otherwise, the only variations that could be found are those related to a phenotype (enamel defaults). Among the 15 cases of amelogenesis imperfecta reported to date, six are nonsynonymous substitutions changing important residues.[3–5]

The question is why *AMELX* exon 6 is so invariable in human.

4.2. *Low variation frequency in primate AMELX*

The lack of polymorphism in human *AMELX* exon 6 is surprising in view of a previous evolutionary analysis showing a high variability of this particular region of *AMELX* in mammals.[6–8] The analysis of primate *AMELX* reveals a high constraint as illustrated by the high conservation of residues. The only variation is the presence of a large insertion in the two lemuriform *AMELX*. This insertion is located at the level of the hot spot of mutation previously identified in mammalian amelogenin.[6] The whole coding sequence of chimpanzee *AMELX* is identical to the human one, although both lineages have diverged 5–6 million years ago.[12] The sequences of the other simians, orangutan, rhesus monkey, marmoset and squirrel monkey are also close to human *AMELX* (99.3–97.1% nucleotide identity), although these lineages have diverged approximately 15, 22 and 34 million years ago, respectively.[13] In prosimians, bushbaby, tarsier and ring-tailed lemur, the number of variations compared to human *AMELX* (96.4–94.8% nucleotide identity), is also small when considering the large evolutionary distance (63 my) which separates simians and prosimians.[13] Such a low level of substitution in exon 6 is not expected when considering that this region of *AMELX* is not subjected to strong functional constraints.[6]

As already notified in the introduction, 12 SNPs are reported in databases for *ENAM* which encodes another enamel matrix protein but is located on an autosomal chromosome (chr. 4). When comparing the human and chimpanzee ENAM sequences, 30 variations were found among which 10 are nonsynonymous (Sire et al., unpublished results). What did occur during the differentiation of the primate lineage that could explain such a strong fixation of *AMELX* sequence and the resulting low level of polymorphism?

The answer to this question could be found in the particular features of sex chromosome evolution, and especially of the distal part of the short arm of chromosome X, housing *AMELX*. Successive steps of multigene inversions occurred in the homologous region in the Y chromosome during therian evolution. These events led to a progressive inhibition of homologous recombination of X and Y during eutherian evolution. The differentiation of the X from the Y chromosome occurred into four distinct steps revealed on the human X chromosome by different evolutionary strata. SRY and RBMY are thought to be the probable candidate genes for this inhibition.[14] The last inhibition step seems to have occurred a little time before the emergence of primates, and concerned the region of the *AMEL* gene location, the *AMELY* loci becoming nonrecombining.[14–16] More precisely, Iwase et al.[16] showed that the boundary on the human X chromosome was located in AMEL intron 2. This could explain why *AMELX*, and particularly the coding region downstream intron 2, the so-called pseudoautosomal region,[14] is more constrained in primates than expected from the evolutionary analysis of other mammalian sequences (approximately 88% of nucleotide identity between human and mouse; 82% between human and cow). In humans, the extent of sequence differences per site on both sides of the pseudoautosomal boundary is important: 10% in the nonrecombining region versus 0.1% in the pseudoautosomal region.[14] We speculate that the presence of such a nonrecombining locus could be responsible for the high, and unexpected constraint

acting on the variable region of *AMELX*. This would mean that, even if the amino acid substitutions in this region of exon 6 would not be important for the correct function of the protein, they are not allowed. It is possible that *AMELX* is more constrained than similar genes located on autosomes because it is present as a single copy only.

REFERENCES

1. Fincham AG, Hu Y, Lau EC, Slavkin HC, Snead ML. Amelogenin post-secretory processing during biomineralization in the postnatal mouse molar tooth. *Arch Oral Biol* 1991;**36**:305–17.
2. Salido E, Yen P, Koprivnikar K, Yu LC, Shapiro L. The human enamel protein gene amelogenin is expressed from both the X- and Y-chromosomes. *Am J Hum Genet* 1992;**50**:303–16.
3. Hart PS, Hart TC, Simmer JP, Wright J. A nomenclature for X-linked amelogenesis imperfecta. *Arch Oral Biol* 2002;**47**: 255–60.
4. Kim JW, Simmer JP, Hu YY, Lin BPL, Boyd C, Wright JT, et al. Amelogenin pM1T and p. W4S mutations underlying hypoplastic X-linked amelogenesis imperfecta. *J Dent Res* 2004;**83**:378–83.
5. Kida M, Sakiyama Y, Matsuda A, Takabayashi S, Ochi H, Sekiguchi H, et al. A novel missense mutation (pP52R) in amelogenin gene causing X-linked amelogenesis imperfecta. *J Dent Res* 2007;**86**:69–72.
6. Delgado S, Girondot M, Sire JY. Molecular evolution of amelogenin in mammals. *J Mol Evol* 2005;**60**:12–30.
7. Sire JY, Delgado S, Fromentin D, Girondot M. Amelogenin: lessons from evolution. *Arch Oral Biol* 2005;**50**:205–12.
8. Sire JY, Delgado S, Girondot M. The amelogenin story: origin and evolution. *Eur J Oral Sci* 2006;**114**:64–77.
9. Delgado S, Ishiyama M, Sire JY. Validation of amelogenesis imperfecta inferred from amelogenin evolution. *J Dent Res* 2007;**86**:326–30.
10. Xiao W, Oefner PJ. Denaturing high-performance liquid chromatography: a review. *Hum Mut* 2001;**17**:439–74.
11. Iwase M, Kaneko S, Kim H, Satta Y, Takahata N. Evolutionary history of sex-linked mammalian amelogenin genes. *Cells Tissues Organs* 2007;**186**:49–59.
12. Patterson N, Richter DJ, Gnerre S, Lander ES, Reich D. Genetic evidence for complex speciation of humans and chimpanzees. *Nature* 2006;**441**:1103–8.
13. Poux C, Douzery EJP. Primate phylogeny, evolutionary rate variations and divergence times: a contribution from the nuclear gene IRBP. *Am J Phys Anthropol* 2004; **124**:1–16.
14. Iwase M, Satta Y, Hirai Y, Hirai H, Imai H, Takahata N. The amelogenin loci span an ancient pseudoautosomal boundary in diverse mammalian species. *Proc Natl Acad Sci USA* 2003;**100**:5258–63.
15. Lahn BT, Pearson NM, Jegalian K. The human Y chromosome, in the light of evolution. *Nat Rev Genet* 2001;**2**:207–16.
16. Iwase M, Satta Y, Takahata N. Sex-chromosomal differentiation and amelogenin genes in mammals. *Mol Biol Evol* 2001;**18**:1601–3.
17. Hu JCC, Yamakoshi Y, Enamelin. Enamelin and autosomal-dominant amelogenesis imperfecta. *Crit Rev Oral Biol Med* 2003;**14**:387–98.