The Origin and Evolution of Enamel Mineralization Genes

Jean-Yves Sire[a] Tiphaine Davit-Béal[a] Sidney Delgado[a] Xun Gu[b]

[a]UMR 7138, Université Pierre et Marie Curie-Paris 6, Paris, France; [b]Department of Genetics, Development, and Cell Biology, Iowa State University, Ames, Iowa, USA

Key Words
Enamel · Evolution · Genomics · Mineralization · Tooth

Abstract

Background/Aims: Enamel and enameloid were identified in early jawless vertebrates, about 500 million years ago (MYA). This suggests that enamel matrix proteins (EMPs) have at least the same age. We review the current data on the origin, evolution and relationships of enamel mineralization genes. ***Methods and Results:*** Three EMPs are secreted by ameloblasts during enamel formation: amelogenin (AMEL), ameloblastin (AMBN) and enamelin (ENAM). Recently, two new genes, amelotin (AMTN) and odontogenic ameloblast associated (ODAM), were found to be expressed by ameloblasts during maturation, increasing the group of ameloblast-secreted proteins to five members. The evolutionary analysis of these five genes indicates that they are related: AMEL is derived from AMBN, AMTN and ODAM are sister genes, and all are derived from ENAM. Using molecular dating, we showed that AMBN/AMEL duplication occurred >600 MYA. The large sequence dataset available for mammals and reptiles was used to study AMEL evolution. In the N- and C-terminal regions, numerous residues were unchanged during >200 million years, suggesting that they are important for the proper function of the protein. ***Conclusion:*** The evolutionary analysis of AMEL led to propose a dataset that will be useful to validate AMEL mutations leading to X-linked AI.

Copyright © 2007 S. Karger AG, Basel

Introduction

Living vertebrates possess a great diversity of mineralized elements, comprising not only endochondral and dermal bone (including osteoderms and scutes), mineralized cartilage, and teeth (dentin and enamel), but also scales, fin rays and otoliths, and egg shells [Huysseune and Sire, 1998; Sire and Huysseune, 2003]. The first mineralized elements, which have given rise to the current

Abbreviations used in this paper	
AIH1	amelogenesis imperfecta 1, hypoplastic/hypomaturation, X-linked
AIH2	amelogenesis imperfecta 2, hypoplastic local, autosomal dominant
AMBN	ameloblastin
AMEL	amelogenin
AMTN	amelotin
EMP	enamel matrix protein
ENAM	enamelin gene
KLK4	kallikrein-4 (prostase, enamel matrix, prostate)
MMP20	matrix metallopeptidase 20 (enamelysin)
MYA	million years ago
ODAM	odontogenic, ameloblast associated (APIN, FLJ20513)
SCPP	secretory calcium-binding phosphoprotein
SIBLING	small integrin-binding ligand, N-linked glycoprotein
SPARC	secreted protein, acidic, cysteine-rich (osteonectin)
SPARCL1	SPARC-like 1 (mast9, HEVIN)
C4orf7	chromosome 4 open reading frame 7 (FDC-SP, MGC71894)

Dr. Jean-Yves Sire
Equipe 'Evolution et Developpement du Squelette', UMR 7138
Université Pierre et Marie Curie-Paris 6, 7 quai St-Bernard, Case 5
FR–75252 Paris (France)
Tel./Fax +33 1 44 27 35 72, E-Mail sire@ccr.jussieu.fr

skeletal diversity in vertebrates, were identified as early as 500 million years ago (MYA) [Sansom et al., 1992, 1994; Janvier, 1996; Donoghue, 1998, 2001]. In fact, the occurrence of mineralized tissue in vertebrates was a major innovation, which was fundamental to the radiation of modern vertebrates in relation to the important roles of the skeletal elements in protection, predation and locomotion [Reif, 1982; Smith and Hall, 1990; Janvier, 1996; Donoghue, 2002; Donoghue and Sansom, 2002; Donoghue et al., 2006].

Our understanding of the mechanisms by which organisms form mineralized elements is still at a rudimentary stage, but we know that biomineralization is mediated by the organic matrix, either through its biological activity or in controlling nucleation, growth and microarchitecture of the mineral deposited [Carter, 1990]. It is assumed that the basic processes of biomineralization are common to all systems and that mineral formation by any individual biological system may diverge from this common pathway. This general definition applies to vertebrates in which the main skeletal elements derive from common ancestral elements [Huysseune and Sire, 1998; Sire and Huysseune, 2003] and there is growing evidence that most proteins currently involved in mineralization of skeletal tissues (bone, dentin, and enamel) also have diverged from a common ancestor [Kawasaki and Weiss, 2003; Kawasaki et al., 2004; Kawasaki and Weiss 2006]. The evolutionary analysis of genes coding for these 'mineralizing' proteins not only has the potential to provide insight into the debated question of the origin of mineralization in vertebrates and of its subsequent diversification, but could also bring important information for humans, as mutations of these proteins lead to genetic disorders (bone [Rowe, 2004]; dentin [Zhang et al., 2001], and enamel [Stephanopoulos et al., 2005]).

This review is devoted only to our current knowledge on the origin and evolution of the genes coding for enamel matrix proteins (EMPs). The reader is referred to the paper by Kawasaki et al., published in this issue [pp 7–24], regarding the history of the other mineralizing proteins in vertebrates.

In living and extinct vertebrates, teeth are protected by a hypermineralized tissue, either 'true' enamel (e.g. in tetrapods) or 'enamel-like' tissue, enameloid (e.g. in cartilaginous and ray-finned fish). These hard dental tissues are identified early in the history of the mineralized integument. They were present in the dermal skeleton of various lineages of jawless vertebrates [Ørvig, 1967, 1977; Reif, 1982; Smith and Hall, 1990; Janvier, 1996; Donoghue and Sansom, 2002; Donoghue et al., 2006]. Enamel

and enameloid are homologous tissues that correspond to different aspects of the same hypermineralization process [Donoghue et al., 2006]. Enamel has replaced enameloid in the lineage leading to tetrapods, probably by a process of heterochrony[1] [Smith, 1995], but enameloid was conserved in two important lineages, chondrichthyans[2] and actinopterygians[3]. The close evolutionary relationships, the similar features of the ameloblasts during their formation, and the same maturation process strongly indicate that both enamel and enameloid matrices could contain similar mineralizing proteins, and that some of them (if not all) were already present in tooth-related elements of early vertebrates, 500 MYA. Unfortunately, our knowledge on EMP genes is restricted to the tetrapods, and the road is still long before we will be able to test the hypothesis of an early origin of EMPs.

In mammals, the enamel matrix is composed of three specific proteins secreted by ameloblasts: amelogenin (AMEL), which represents 90% of the matrix deposited, and enamelin (ENAM) and ameloblastin (AMBN), which are components of the remaining 10% organic matrix. Evolutionary analyses have indicated that these three EMPs constitute a family, which, itself, is included into a larger family, the secretory calcium-binding phosphoprotein (SCPP) family. This SCPP family comprises other Ca-binding proteins: some saliva proteins, milk caseins and small integrin-binding ligand, N-linked glycoproteins (SIBLINGs) [Fisher and Fedarko, 2003], which contain five dentin and bone proteins [Kawasaki and Weiss, 2003]. Interestingly, with the exception of AMEL that is located elsewhere (on sex chromosomes X and Y in placental mammals), all SCPP genes are located in two clusters on the same autosomal chromosome. This supports the idea that SCPP genes originated by tandem duplication followed by neofunctionalization.

In humans, several types of amelogenesis imperfecta (AI), leading to enamel hyploplasia or hypomineralization, are related to mutations in AMELX (14 X-linked AI, AIH1, identified to date [Hart et al., 2002; Kim et al., 2004; Stephanopoulos et al., 2005]) or in ENAM (5 autosomal-dominant AI, AIH2 [Hart et al., 2003; Hu and Yamakoshi, 2003; Kim et al., 2005]) genes [review in Stephanopoulos et al., 2005]. In contrast, although being con-

---

[1]   Heterochrony: developmental changes in the timing of events, leading to changes in size and shape from an ancestral state.
[2]   Chondrichthyans: the cartilaginous fishes, including sharks, rays and chimaeras.
[3]   Actinopterygians: the ray-finned fish, which are the dominant group of vertebrates.

sidered as a candidate gene, AMBN was excluded from a causative role within the families studied [Mardh et al., 2001].

Since a few years, we focus our attention on EMP gene relationships (AMEL, AMBN, and ENAM), and more precisely on the origin and evolution of AMEL, the best known member of the family [Delgado et al., 2001; 2005; Sire et al., 2005; 2006; Delgado et al., in press]. Here, (i) we summarize these previous data, (ii) we provide new information on two newly identified genes, amelotin and APIN protein, that are expressed by the ameloblasts, (iii) we provide a date for EMP gene origin and discuss this result in the light of our knowledge of enamel and/or enameloid appearance in vertebrate evolution, and (iv) we show how evolutionary analysis of AMEL can help to identify structural features that might be important for the protein function, and to validate mutations responsible for genetic diseases.

Ameloblast Products: EMPs, and Amelotin and APIN Proteins

In mammals, the synthetic activity of ameloblasts is divided in two successive phases corresponding to two stages of enamel formation: secretion and maturation, separated by a transition stage. To our knowledge, during the former step ameloblasts deposit four proteins in the extracellular matrix: three EMPs (AMEL, ENAM, and AMBN) and a tooth-specific, calcium-dependent peptidase, MMP20 (= enamelysin) [Bartlett et al., 1998; Bartlett, 2004]. During the transition and maturation stages, ameloblasts have been shown to produce a fifth protein, kallikrein 4 (KLK4), a pleiotropic, calcium-independent protease, which is involved in the final proteolysis of the remaining organic matrix [Simmer et al., 1998; Hu et al., 2002; Simmer and Hu, 2002].

Recently, two novel genes were found to be also expressed by ameloblasts during tooth formation: amelotin (AMTN, but annotated UNQ689 in human genome build 36.2) [Iwasaki et al., 2005; Moffat et al., 2006b] and ODAM ('odontogenic, ameloblast associated', previously named APIN or FLJ20513) [Moffat et al., 2006a]. Can the proteins encoded by these two genes be considered EMPs? In other words, although being produced by ameloblasts, are AMTN and ODAM structural proteins playing a role in enamel matrix formation and/or mineralization? In rats, AMTN was localized to the basal lamina of maturation stage ameloblasts [Moffatt et al., 2006b]. This location seems to indicate a possible role of AMTN in cell adhesion, and it also demonstrates the absence of AMTN participation in enamel matrix formation. In humans, ODAM was first identified from extracts of amyloid deposits obtained from calcifying epithelial odontogenic tumors [Solomon et al., 2003]. Transcripts of this gene were also found at a high level in gastric cancer [Aung et al., 2006]. In rats, ODAM is specifically expressed in ameloblasts during maturation stage [Moffatt et al., 2006a], but the location of the protein in the extracellular matrix remains to be shown. However, late expression during tooth formation does not mean that the secreted ODAM protein is not incorporated in the enamel matrix at the end of the mineralization process. Such a location would not be surprising if one considers that the protein was first isolated from calcifying tissues of odontogenic tumors [Solomon et al., 2003]. Therefore, if AMTN cannot be considered an EMP, the few data available to date do not permit to exclude ODAM from this family.

Interestingly, these two genes are located in the same cluster as EMPs, and they share structural similarities with the members of this family (see below). This indicates that AMTN and ODAM were probably created after duplication of an ancestral EMP gene and, therefore, that they belong to the SCPP family [Kawasaki and Weiss, 2006].

In the following, we provide some data on these two newly identified ameloblast-secreted proteins, although concentrating on the evolutionary relationships of EMPs (AMEL, AMBN, and ENAM).

Evolutionary Relationships of AMTN, ODAM, and EMP Genes

EMPs are evolutionarily related, forming a gene family that belongs to a super-gene family called SCPP [Kawasaki and Weiss, 2003]. All SCPP genes probably derive from a common ancestor by gene duplications. The key gene could be SPARC-like1 (SPARCL1, also called HEVIN or SC1), which was created after a duplication of SPARC (osteonectin) [Kawasaki et al., 2004]. Four lines of evidence have permitted to establish SCPP relationships, and SPARCL1 may resemble the ancestral form of SCPP: (i) common gene structure and similar protein characteristics in the N-terminal region, (ii) in most SCPPs, presence of an SXY phosphorylation site encoded in the 3′ region of the second coding exon, suggesting Ca-binding properties, (iii) location on the same chromosome, and (iv) presence of SPARCL1 on this chromosome, adjacent to the dentin-bone protein gene cluster [reviewed by Kawasaki and Weiss, 2006].

```
              exon2          |         exon3            |         exon4          |    exon 5
Ancestral     MKTTILLFCL LGTTQSLPKQ LNPALGLPPT KLGPDQPTLL NQQQPNQVFP SLSQIPLTQM LTLGTDLQLI NPATGMPPGT (80)
Human         .RS....... ..S.R...-. .K........ ...A...G..P ....S..... .......... ....P..H.L ...A..T...
Chimpanzee    .RS....... ..S.R...-. .K........ ..PA...E..P ....S..... .......... ....P..H.L ...A..T...
Rhesus monkey ...M...... ..S.....-. .K........ ...A...G..P ....S..... .......... ....P..H.L ...A..T...
Mouse         ...M...L.. ..SA...... ...S.V.A. .PT.G.V.P P......... .I.......L ....S..P.F ...A.-.H.A
Rat           ...VV..L.. ..SA....R. .S....A.A. .PT.G.V.P T......... .I.......L ....S..P.F ....-..H..
Cow           ..AA...... ..S.L..M. .....V.... ..V..A... .P.......S .........H. ....SN...L ...A...S..
Pig           .......... ..S.L..M. F..V...... ..V..A..R .......... .......... .....S...L ...I..V.SS
Dog           ...M...LY. .S.....A. .......... ...T.H.A. .......... .......... .F..AS...L ...A..AS..
Elephant      .......... ..S....... .......SAA ..V..A... .......... .......... ....L...S...QL ......A...
Opossum       ...AV..... ...I....Q. .Y.GV....P ...LE..A.F TP..S..L.. P.GL...... FSV...M..M T....LL..I

              exon5          |        exon6  |  exon7 |                    exon8
Ancestral     QTLPMTLGDL NIQHQLKPQM LPIIVAQIGA QGAILSSEEL PMAPQIFTGL LIQPLFPGAI LPTSQAGTNP NAQDGALPAG (160)
Human         ..H.L...G. .V.Q..H.HV ...F.T.L.. ..T....... .---...S. I.HS....G. .......A.. DV...S....
Chimpanzee    ..H.L...G. .V.Q..H.HV ...F.T.L.. ..T....... .---...S. I.HS....G. .......A.. DV...S....
Rhesus monkey .AH.L...V. .LPQ..Q.H. ...F.T.L.. ..T....... .---...S. I.HS....G. .......A.. DV...S....
Mouse         H...F...P. .G.Q..Q... .......L.. ...L..... .L.S..... ..H...... P.SG....K. DV.N.V..TR
Rat           ....F...P. .G.Q..Q... .......L.. ...L..... .L.S..... ..H...... Q.SG.T.AK. DV.N....TR
Cow           ....LA..G. KV.Q..Q... ..V...HF.. ..T....... QGTS..L.. IFH....... ....--A.. D..N.I...
Pig           ....L...A. .V.Q..Q... I.V...HL.. .......... .ATR..L.. IFHT...... ..P.P--AK. D..N.IH...
Dog           ....LS..V. .T.Q..Q... ..V...HL.. H......... .GS....... IF........ ...P--A.. D..N.I....
Elephant      ..F.LN..G. T.KQ..QS.L ......L.... .......... .......A.. .......... ....L..AT. EV.E.I....
Opossum       .I......T S.AP.VN... ..VL...... ...VR..... .I........ ....FGT..T ...G.S.IDA .T..A.....

                             exon8                          | ex9
Ancestral     QSGGNPAIQG TPEGQLPTPS GTDD-VFEAT TPAGIQRATH TTEETTTEAP NGTQ  (214)
Human         GA.V...T.. ..A.R..... .....D.AV. ........S.. AI..A...SA ..I.  (210)
Chimpanzee    GA.V...T.. ..A.R..... .....D.AV. ...D...S.R AI..A...SA ..I.  (210)
Rhesus monkey .A.V...... ..A.R..... .....D.AV. ........S.. A.......*. ..I.  (210)
Mouse         .A.AKAVN.. .TP.HVT..G V...DDY.MS ....LR.... ...G..IDP. .R..  (213)
Rat           .A.AS..N.A .TP.H-T..A V...DDY.MS ....L..... ...G..MDP. .R.K  (214)
Cow           .A.A...A.. ...DPFS... .....D.AS. .......GRP .....P.GS. K.I.  (212)
Pig           .A.A...V.. ..R.PF..S. .....D.DV. ...L.G..A ......GS. ..M.  (212)
Dog           .A........ ...SFST... D....D.GV. A......G.. ..Q...SGP. ....  (212)
Elephant      .A.L.....R ...KHPS.S. D..S...GV. ....L..GMR ..G.....S. .EI.  (214)
Opossum       ........W. .S........ SP........ I.V...K..- --..GS..... ...D  (211)
```

Fig. 1. Amelotin (AMTN): alignment of 10 complete mammalian amino acid sequences and of the putative ancestral sequence (shown at the top). Six sequences were inferred from DNA sequences retrieved in databases (blast search against sequenced genomes). These sequences were checked against three published complete coding sequences: human, accession number AY358528; mouse, AK017352, and rat, DQ198381. The pig sequence was obtained from the literature [Moffatt et al., 2006]. The putative ancestral sequence was calculated using PAUP 4.0 and MacCLADE 3.06. Vertical bars indicate the limits between exons. The signal peptide is in a box. The total number of residues in each protein is indicated at the end of each sequence. Unchanged residues are shown on a gray background. · · · · · = Identical residue; – – – = indel.

Recent studies on the origin and evolution of AMEL in tetrapods have extended our knowledge on EMP relationships [Sire et al., 2005, 2006]. A phylogenetic analysis using a large set of sequences demonstrated that AMEL and AMBN are sister genes, and that AMEL was created from a duplication of AMBN. In addition, it was shown that both genes are related to ENAM, which was recognized as a more ancient member of the EMP family. The calculation of putative ancestral sequences of EMP genes and the use of SPARCL1 as an outgroup were helpful for this phylogenetic analysis. Putative ancestral sequences permit to go back to the gene origin, while the whole dataset of sequences is less informative to reveal possible relationships. Indeed, although they are phylogenetically related, EMP genes show large sequence variations when comparing evolutionary distant lineages. Moreover, since their creation, hundreds of million years ago, AMEL, AMBN, and ENAM have acquired specific functions and their sequences diverged rapidly. However, currently available sequences permit to calculate putative ancestral sequences of EMPs at the origins of the mammals only, i.e. when monotremes[4] and therians[5] diverged, 225 MYA [van Rheede et al., 2006]. Using amniote or tetrapod sequences was not possible because of the few sequences

[4] Monotremes: egg-laying mammals: extant members are the echidnas and the duck-billed platypus.
[5] Therians: marsupials and placental (eutherian) mammals.

```
                     exon2     |      exon3     |      exon4      |              exon5
Ancestral       MKTTILLGLL GATMSAPLIP QHLLSASNSN ELLLNLNNAR LRPLQ--LQG PANSLIPPFP GILQQQQ-T QTPGLSQFSL (80)
Human           ..II....F. ...L...... .R.M...... ........GQ .L....... .L.W....S .........A .I......
Chimpanzee      ..II....F. ...L...... .R.T...... ........GQ .L....... .L.W....S .........A .I......
Orangutan       ..II....F. ...L...... .R.M...... ........GQ .L....... .L.W....S .........A .I......
Rhesus monkey   ..II....F. ...L...... .R.M...... ........GQ .L..R.... .L.W....S .V.......A .I...A....
Mouse           ..II.....I ..SS....S .R......H ........GQ .L.....F. AF..W..... .F.....-.A .VS.RP..T.
Rat             ..II.....I ..S.....T .R........H ........GQ .L.....F.S AF..W..... .L.....QA .VS.HP..P.
Cow             .R.L....I. .......... .......M.. ......... .......Q .......F. .F..WF.... .......N .V...P...
Dog             ...I...... .......... .R.M...... ........Q .Q..P..F. .. .F..W....S .........A .I......
Tenrec          ..II..... ...S....T. .......... .......... .L.....F. .F...T...S ...H.....A .V..AP....
Opossum         .RAA....F. .VALA...L. .P.......R ...MG.G... ..G.PPG..A S.P..F.L. .A.HHG--.P RP--.GL-.W

                            exon5                      |        exon6         |      exon7
Ancestral       PTLDQFAGLV PNQIPLPGQA KLAQRTQAAQ QEPSQPQTPQ QNQQDPNQMI PYVFSFKVPQ DQAQMLQYYP VYMYRPWEQP (160)
Human           SA.......L .......T.E. SF..GA..G. VD.L.L...P .T.PG.SHVM .......M.. E.G..F.... ...VL.....
Chimpanzee      SA.......L .....F.... SF..GA..GH VD.L.L...P .T.PG.SHVM .......M.. E.G..F.... ...LL.....
Orangutan       SA.......F .....F.... SF..GA..G. VD.L.L...P .T.PG..HV. .......M.. E.G..F.... ...LL.....
Rhesus monkey   SA.......F .....F.... SF..GA..G. VD...A...P .T.PG..HVM .......M.. E.G..FE... ..VLL.....
Mouse           S..ES....F .....SR.V G...GG..G. PDL..Q...P .T..SASP-M S..VPV.... ..T..F.... ...LL.....
Rat             S..ES....F .....FSR.V GF..GG..G. PDF..Q...S .T..-ASP-M S..VPV.... ..T..F.... ...LL.....
Cow             S.REW..... ...FV...V SF..G...G. LD........ .T.RG.KNVM .S..-..M.. E.......... ...FL.....
Dog             SA..R....F ...T.F..RV SF..G..VG. .D.......P .T..S..HVM .......M.. E.......... ...LL.....S
Tenrec          .SQ.L....F .....F...T NF..ESRPG. LDF......L .I..GT.PVL ..S....L.. E.T.R.H... .FL.F.....
Opossum         .S.GH.G... ..EV...... R.....S..V. ..APRL.ML. ......Y.I. .CF...G... VWG..VP... .CV.GA----

                         exon7         |      exon8      |        exon9            |         exon10
Ancestral       QQTPT--QLP QQAGQQQPEE QVPFYTQFGY IPQQAQPVIP GGQQQIAFDP LRGTAPETPA MPTEKVIPYT QKEMINLRHP (240)
Human           ...VP..RS. ..TR...Y.. .I..A.... ...L.E.A.S .....L.... QL.....IAV .S.GEE...L ..A..F..D
Chimpanzee      ...VP..RS. ..TR...Y.. .I..A.... ...L.E.ATS .....L.... QL.....IAV .S.GEE...L G..A..F..D
Orangutan       ...VP..RS. ..TRE..Y.. .....A.... ...L.E.A.. .....L...T QL.D..IAV .S.GEE...L ...V..F..D
Rhesus monkey   ...VP..RS. P.TR...Y.. .I..D.... ...L.E.A.. .....L.... QL.....IAV .S.GEE...L ...V..F.RD
Mouse           .-.V...SS. .HT...LF.. .I..N...F A.P..E.GV. ....HL...S FV.......G ..V.GSLL.P ...P.SFK.D
Rat             ...V...SS. ..T...LY.. .I..N...F V....E.GV. ....HLVL.S FV.......G ..AVEGPL.P ...P.GFKQD
Cow             ...VA...S. P.TRE.LF.K .M....E... ...VE..M. VE...PV... FL.....IA. .A-E.S..L ......FQ.T
Dog             ...AP...S. P.T...F.. .M....... V.V.VE..M. ....L.L.. VL.......VV .VRSR...L R..V..FK.A
Tenrec          L..GPST... ......F.V E.......E. .....D..L. ..H..L.... .I......TI .AGG..THS ...RT.S...
Opossum         .DP.L....S AL..PP..P. ...S.PEL.C LT......L. ....EM...L ...DV..S.. ..I.N.L... .R..V..GY.

                         exon10           |     exon11
Ancestral       NAGIFMPSNS PKHSTTNIFA SPTDKTITPE LMEEKTNTDS LKEP (279)
Human           S..V....T. ..P....V.T .AV.Q..... .P...DK... .R.. (279)
Chimpanzee      S..V....T. ..P....A.T .AV.Q..... .P...DK... .R.. (279)
Orangutan       ...V....T. ..P....V.T .AI.R.L... .P...***** **** (279)
Rhesus monkey   S..VL...T. ..P....V.T .AI.R...AK FP...AK..G .R.. (279)
Mouse           ...V...TT. ..P.D.F.T .GI.P..A.- --.Q.VK... .R.. (273)
Rat             .V.VST..T. ..PD.G.F.T .EINP..A.L .P.Q.V.A.. .R.. (278)
Cow             .....I..T. Q.P...IF.T .AV.PI..R. .T.K.AK... .... (277)
Dog             .G...V..T. QTP....Y.. PAI.P..... ...K.AK..Y .... (279)
Tenrec          G..MSR..A. ..P.IAT..T .NI.P.MD.. F..A.AT... .R.. (281)
Opossum         Y.......YP L..Q.A..L. ....NVVPL. .L..EI.P.L .... (276)
```

Fig. 2. ODAM (APIN protein): alignment of 10 complete mammalian amino acid sequences and of the putative ancestral sequence (shown at the top). Seven sequences were inferred from DNA sequences retrieved in databases (blast search against sequenced genomes and trace archive-Whole Genome Shotgun). The sequences were checked against three published complete coding sequences: human, NM17855; rat, DQ198380, and mouse, NM27128. For further information, see legend to figure 1. * = Unknown residue.

available in reptiles, and amphibians are not representative enough of EMP evolution in these lineages (see below).

Here, we use the same approach to try to identify the origins of the two newly identified genes, AMTN and ODAM, with regard to the EMPs. Ten complete coding therian [a metatherian (opossum) + nine eutherian species] sequences of both genes were retrieved from databases and the literature. The inferred protein sequences were aligned using CLUSTALX and hand-checked using the sequence alignment editor Se-Al 2.0 (fig. 1, 2). The putative ancestral sequences of therian AMTN and ODAM (i.e. 190 million years old [van Rheede et al., 2006]) were calculated with PAUP 4.0 (Sinauer, Sunderland, Mass., USA), taking into account the current mammalian phylogeny [Madsen et al., 2001; Murphy et al.,

2001; Delsuc et al., 2002; van Rheede, 2006] using Mac-CLADE 3.06 (Sinauer; fig. 1, 2). Given the small number of sequences available and the lack of sequences of representative species in some important mammalian lineages (e.g. Perissodactyla, Insectivora, Xenarthra, and prototherians: platypus or echidna), it was not possible to perform an evolutionary analysis. However, some findings from these alignments reveal some interesting points.

AMTN Analysis in Mammals

The amino acid sequences (ranging from 210 to 214 residues) were easily aligned without including numerous gaps (fig. 1). The presence of large conserved regions when comparing eutherian and metatherian AMTN, and the large differences with the other members of the family suggest that this gene was created long before mammalian lineage divergence, which occurred 310 MYA [Murphy et al., 2001; Hedges, 2002]. As a consequence, a functional AMTN gene might be found in reptile genomes. The putative mammalian ancestral sequence comprised 214 residues. Four residues were lost during primate evolution. Only a few residues (22%) remained unchanged during mammalian evolution (47 out of 214, fig. 1). Such relaxed selective constraints on AMTN suggest that some polymorphism could be encountered in humans. This idea is supported by the comparison of chimpanzee and human sequences: four amino acids (1.9%) were substituted within a time period of 5–7 million years, which separates the two lineages [Kumar et al., 2005]. In addition, most of the unchanged residues are dispersed through the sequence. This means that the number of conserved positions will almost certainly drop when sequences from other mammalian and reptilian species become available [Sire, unpubl. res.]. However, three important features emerge from this alignment.

(i) In the N-terminal region encoded by exon 2, 55% of the residues (10 out of 18) are unchanged. This region is similarly organized as in the other SCPPs, and it is mainly composed of the signal peptide, which plays an important role in the extracellular secretion of the proteins.

(ii) In positions 55–58 (exon 4), an IPLT motif is conserved, which means that this predicted O-glycosylated site could be important for the function of AMTN. Two other predicted O-glycosylated sites (threonines) are also conserved, but isolated, in exon 8.

(iii) In positions 116–120 (exon 6), a SSEEL motif is well conserved. This is a putative CK2 serine phosphorylation site [Moffatt et al., 2006b]. Surprisingly, in contrast to the condition observed in EMPs, there are no con-served residues in the C-terminal region of AMTN. It is clear that a further study, including new mammalian and reptilian sequences, is necessary to reveal further details on gene ancestry and to perform an accurate evolutionary analysis.

ODAM Analysis in Mammals

The amino acid sequences, which contain 273–281 residues depending on the species, were easily aligned without including numerous gaps (fig. 2). The absence of large sequence variations and the large differences compared with the other members of the family indicate that ODAM, like AMTN and the EMPs, arose before the mammalian/reptilian split. The putative ancestral mammalian sequence comprised 279 amino acids. Regarding AMTN, only a few residues (16.8%) are unchanged (47 out of 279, fig. 2), and this low selective pressure suggests that some polymorphism could occur in human ODAM (seven amino acid variations, 2.5%), are found between human and chimpanzee). Most of the conserved residues are dispersed along the sequence, but four features emerged from this alignment:

(i) the N-terminal region (signal peptide) in which 47% of residues (8 out of 17) are unchanged (exon 2);

(ii) in positions 25–33 (exon 3 and beginning of exon 4), a SASNSxELL motif is well conserved; this is a probable phosphorylation site;

(iii) in positions 147–150 (exon 7), a YYPV motif is kept unchanged, but its function remains to be discovered, and

(iv) in contrast to AMTN, four residues are conserved in the C-terminal region (exon 11). Here too, further sequences from species representative of other tetrapod lineages are needed to perform an accurate evolutionary analysis.

Relationships of Ameloblast-Expressed SCPP Genes

The structure and organization of the two newly identified ameloblast-expressed genes, AMTN and ODAM, were compared to the putative ancestral sequences previously calculated for the three EMP genes and SPARCL1 (fig. 3). A previous analysis of the putative ancestral sequences of EMPs had shown that:

(i) AMEL exon 4 was created during eutherian evolution (it is present in some eutherian lineages only), and two additional exons 8 and 9, that are unique to the mouse and rat, were created by duplication of exons 4 and 5 [Bartlett et al., 2006a];

(ii) AMBN exons 8 and 9 have appeared in primates only, as duplications of exon 7;

Fig. 3. Gene structure of the putative ancestral coding sequences calculated for the EMP genes (AMEL, AMBN, and ENAM), the two other ameloblast-expressed SCPP genes (ODAM and AMTN) and SPARCL1, the supposed SCPP ancestor. The reference to exon number on top of the boxes is that of the human sequences. Empty boxes indicate exons lacking in the basal mammalian taxa. The nucleotide number of each exon is indicated within the boxes (not to scale). Dark gray = Signal peptide.
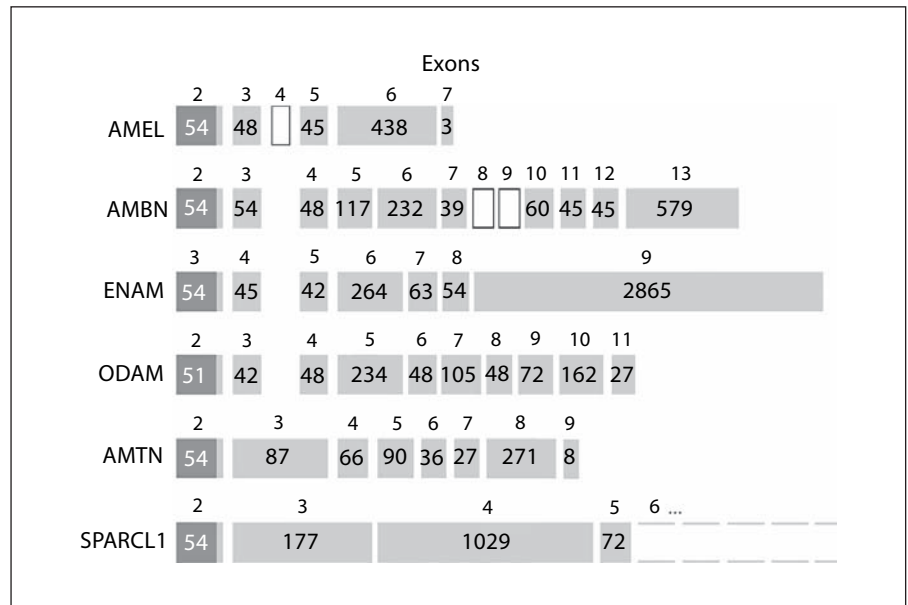
(iii) ENAM exon 3 can be considered as homologous to exon 2 of the other genes, and

(iv) although considered the probable ancestor of SCPPs, the N-terminal organization of SPARCL1 is different from that of the EMPs, except for the first coding exon, exon 2 [Sire et al., 2005; 2006].

The structural comparison of the six putative ancestral genes, i.e. EMPs, AMTN and ODAM, and SPARCL1, confirms the previous findings that only the first three coding exons share similarities (fig. 3). As already shown for human genes, the strongest similarity of the ancestral sequences concerns exon 2 (exon 3 in ENAM), which encodes a well-conserved signal peptide and the first two residues of the protein [Kawasaki and Weiss, 2003; Kawasaki et al., 2004]. The two following exons in EMPs and ODAM are small and of roughly the same size (42–54 and 42–48 bp, respectively), with the third exon (exon 4 for ENAM) ending with an SXE phosphorylation motif. In mammals, such an organization is not observed in AMTN and in SPARCL1, which exhibit a larger third exon (87 and 177 bp, respectively), and which lack an SXE motif. The sizes of exon 3 in chicken and teleost fish SPARCL1 are small (54–57 bp), similar to the size of SPARC exon 3. This suggests that SPARCL1 originally had a small exon 3. However, in the absence of data for SPARCL1 in amphibians, crocodiles and squamates (lizards and snakes) we cannot claim that a small exon 3 was the condition when actinopterygian and sarcopterygian lineages separated. Our alignment (not shown) indicates

that the third exon in AMTN could correspond to the two short exons 3 and 4 in the other genes. The phylogenetic position of AMTN suggests that this exon could have been created by a fusion of these two short exons (see below). The mere comparison of gene organization already suggests that these genes belong to a single family [Kawasaki and Weiss, 2003]. With the exception of AMTN, the structure of which is somewhat different from the four other genes, the N-terminal region of EMPs and ODAM is similar. In addition, the organization of ODAM is more similar to that of ENAM, which suggests closer relationships of ODAM with ENAM than with the other genes (fig. 3).

Since 2002, the study of EMP (and SCPP) relationships has highly benefited from gene mapping in humans, and new data have progressively accumulated in other tetrapod species (but unfortunately mainly in mammals) [http://www.ncbi.nlm.nih.gov/]. In humans, SCPP genes are located on chromosome 4, on which they form two clusters, separated by 15 Mb: the dentin and bone protein cluster (4q22, approximately 375 kb), to which SPARCL1 is adjacent, and the saliva, milk and ameloblast-secreted protein cluster (4q13, approximately 770 kb; fig. 4). The only exception is AMEL, two copies of which are found on the sex chromosomes. The most important copy, which encodes 90% of the transcripts, resides on chromosome X (fig. 4). In humans AMELX is located in antisense in the intron 1 of the ARHGAP6 gene. As AMEL belongs to the EMP family, it is clear that it was translo-
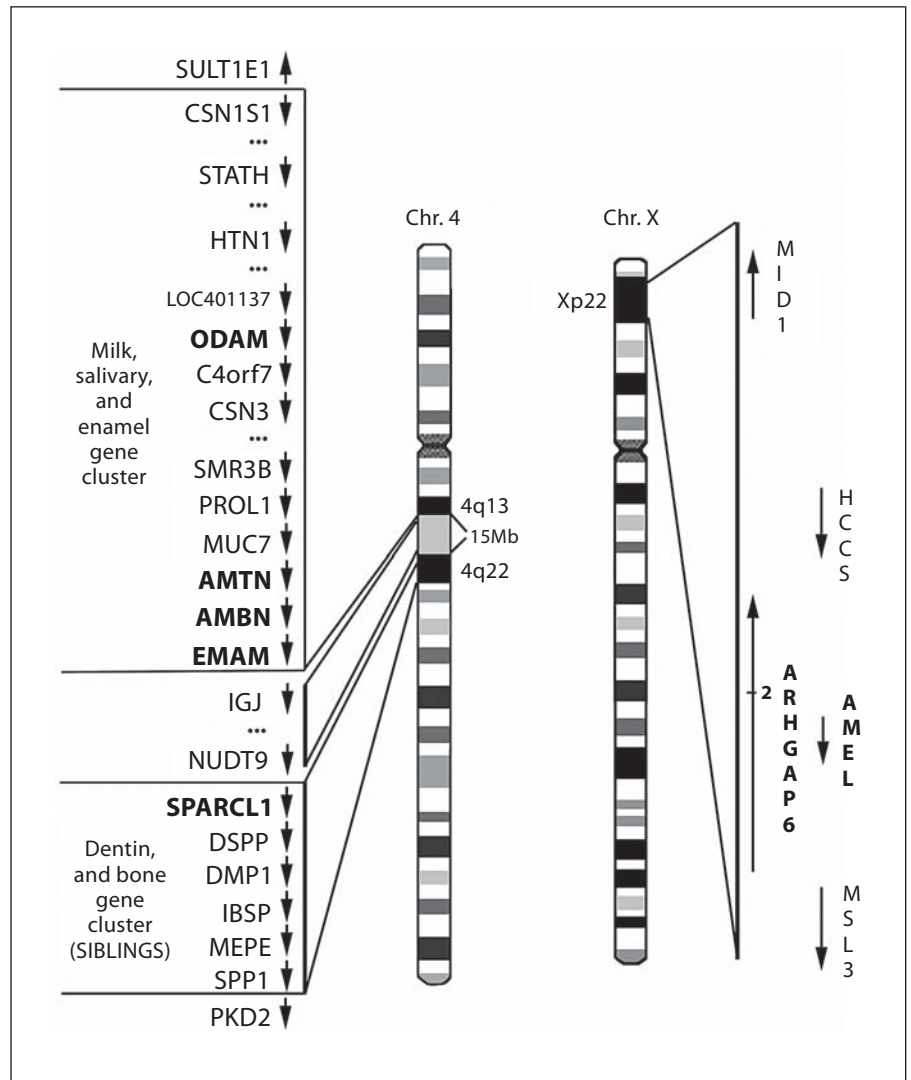
Fig. 4. Location of the ameloblast-expressed SCPP genes and of SPARCL1 on human chromosomes. ENAM, AMBN, AMTN, ODAM, and SPARCL1 are located on chromosome 4, in two clusters separated by 15 Mb. AMEL is the only SCPP found elsewhere, on the sex chromosomes. The most important AMEL copy is on chromosome X, located in antisense within ARH-GAP6 intron 2. SCPP genes are identically oriented on chromosome 4.

cated from the 'EMP family' chromosome to another chromosome (ARHGAP6 gene intron), either immediately after its duplication, or during a particular event, which occurred some time after a tandem duplication. ENAM, AMBN, and AMTN are adjacent genes on chromosome 4, while ODAM is located between C4orf7 (follicular dendritic cell secreted peptide) and LOC401137 (a hypothetical protein), at some distance from the three ameloblast-expressed genes, and separated from them by some salivary protein and milk casein genes (fig. 4). This syntheny is conserved in the few mammalian species for which genes are mapped [http://www.ncbi.nlm.nih.gov/map view/]. In birds, which lost teeth approximately 80–100 MYA [Huysseune and Sire, 1998], the SIBLING genes are found in syntheny, while the enamel, saliva, and milk

protein gene cluster is lacking [Kawasaki and Weiss, 2006]. In amphibians *(Xenopus)* the syntheny is roughly conserved, but some mineralizing protein genes, known to be important in mammals, are apparently lacking. However, this absence could be related to the currently incomplete assembly of this frog genome [Kawasaki and Weiss, 2006].

The five ameloblast-expressed genes (ENAM, AMBN, AMTN, ODAM, and AMEL) were created by tandem duplication from a common ancestor [Kawasaki and Weiss, 2003, Kawasaki et al., 2004; Kawasaki and Weiss, 2006]. These duplications were probably asymmetric, i.e. after each duplication one copy kept the former function of the protein and did not diverge much from the ancestral sequence, while the other copy differentiated rapidly and

acquired new functions (neofunctionalization) [Chung et al., 2006; Steinke et al., 2006]. These functions were positively selected, but they are still to be uncovered for most of these genes. This finding is deduced from comparison of the gene structure (fig. 3) and from sequence analysis (fig. 1, 2 and Sire et al. [2006]). Indeed, the roughly conserved features of the N-terminal region suggest not only a common origin but also some functional similarities (they are all ameloblast-expressed proteins). In contrast, the rest of the sequence (the largest part) houses the specificities of each protein (i.e. its proper functions) and, therefore, is strongly divergent. The specific function of each protein could reside either in this whole sequence, as for instance for most part of the region coded by AMEL exon 6 [Sire et al., 2006] (see below), or in some particular important loci, as for instance the conserved motifs that emerge from the alignment of AMTN and ODAM mammalian sequences (fig. 1, 2).

The next questions now are: how are these ameloblast-expressed genes related and which evolutionary scenario can be proposed for their origins in vertebrates?

AMEL and the Evolutionary Origin of EMP Genes

The current knowledge on the relationships and evolutionary origin of EMPs was acquired in several steps, and this study represents the last (but not least) one. This story can be briefly reconstructed as follows.

In 2001, Delgado et al. showed a high sequence similarity of the 5′ region (exon 2, which mainly encodes the signal peptide) of AMEL, SPARC, and SPARCL1, suggestive of a common origin of this region after duplication. Using a molecular-clock method to estimate SPARC/SPARCL1 divergence, these authors proposed that AMEL exon 2 was created >600 MYA (i.e. at the end of the Precambrian). This meant that AMEL could have been present before the origin of vertebrates, 530 MYA [Shu et al., 1999, 2003], and of the first evidence of mineralized elements in euconodonts, 500 MYA [Sansom et al., 1992; 1994; Janvier, 1996].

Two years later, taking advantage of the availability of the sequenced human genome and gene mapping, Kawasaki and Weiss [2003] convincingly demonstrated that (i) EMPs comprise a subfamily, (ii) EMP, milk casein, and salivary protein families together are regrouped into a cluster on chromosome 4, forming a larger family, and (iii) this family also contains the SIBLING gene cluster, which is located in another locus on the same chromosome. The SCPP family was now a fact.

Another chapter was added to the story when SPARCL1 was proposed to be the common ancestor of SCPP genes on the basis of its location, adjacent to the SIBLING cluster on chromosome 4, and of the structure of its N-terminal region [Kawasaki et al., 2004]. Therefore, although SPARC still remains at the origin of the mineralizing protein gene story, it was SPARCL1 that gave rise to the SCPP gene ancestor. SPARC is present in both protostomes and deuterostomes[6], where it influences cell behavior and interactions with the extracellular matrix, rather than being involved in the generation of mineralized tissues. Several runs of duplications, and subsequent sub- and/or neofunctionalization have occurred and led to the current diversity of this family. Using a molecular-clock method, the divergence date between SPARC and SPARCL1 was found to be inferior or equal to the current divergence date of cartilaginous fishes (estimated at 528 ± 56 MYA using molecular dating [Kumar and Hedges, 1998]). This led to the conclusion that the SCPP genes probably emerged after this date [Kawasaki et al., 2004]. This dating is more recent than the >600 MYA previously calculated by Delgado et al. [2001].

Taken together, these findings suggest that AMEL is more distantly related to SPARC and/or SPARCL1 than hitherto believed before, and that at least five duplication events took place from SPARC to AMEL [Sire et al., 2006]:

SPARC → SPARCL1 → SCPP ancestor → ENAM → AMBN → AMEL

Below, we briefly review the current scenario for EMP gene relationships, which was established in the course of studies dealing with AMEL origins [Sire et al., 2005, 2006]. The previously published dataset is completed by additional information on AMTN and ODAM (fig. 1, 2), with the aim to clarify the relationships of all ameloblast-secreted SCPP proteins.

The Evolutionary Origin of AMEL
This study was performed in three steps:

Step 1: Evolutionary Analysis of AMEL Sequences in Tetrapods
A total of 80 AMEL sequences (including mammals, reptiles, and amphibians) were compiled (published se-

---

[6] Protostomes and deuterostomes: the two main divisions of bilateria mostly comprising animals with bilateral symmetry and three germ layers (endoderm, mesoderm, and ectoderm).
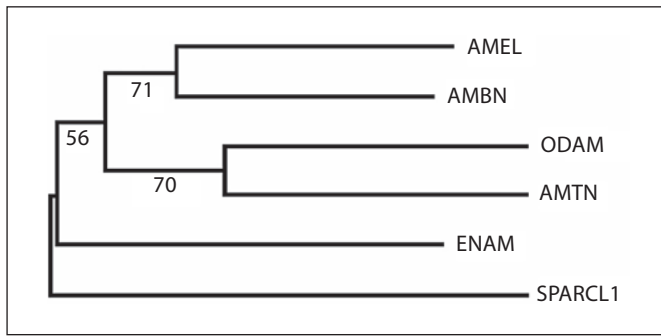
Fig. 5. Phylogenetic analysis (distance analysis with maximum likelihood using neighbor-joining method) of the five ameloblast-expressed SCPP genes (AMEL, AMBN, AMTN, ODAM, and ENAM) based on the 5′ region (288 bp) of their putative ancestral sequences. The ancestral sequence of SPARCL1, the probable ancestor of SCPP genes, was used to root the tree. Bootstrap values are indicated (1,000 replicates).

quences, sequences retrieved in the databases, and new sequences; see Sire et al. [2006] for the species list). The sequences were aligned as described above for AMTN and ODAM, and a putative AMEL ancestral sequence was calculated using PAUP 4.0. The conserved versus variable regions were determined and used for the next step.

Step 2: Search for Sequence Similarity in Databases

A PSI-blast search (National Center for Biotechnological Information) of statistically significant similar peptides was performed in GenBank [Sire et al., 2006]. The well-conserved regions of the putative ancestral AMEL were used, i.e. the N-terminal region: exon 2 (signal peptide), exon 3, exon 5, and beginning of exon 6. Sequence similarities were detected with AMBN, then with ENAM and, finally, with SPARCL1. It is noteworthy that the first non-AMEL sequence to be found using PSI-blast was crocodile AMBN, indicating that the latter is closer to ancestral AMEL than mammalian AMBN. This would mean that crocodile AMBN is more conservative of an ancestral state, and could have been subjected to a slower rate of evolution than mammalian AMBN after reptile/mammal divergence. At this time (July 2004), neither AMTN nor ODAM sequences were available in databases [Sire et al., 2005].

Step 3: Sequence Analysis

The putative ancestral sequences of AMEL, AMBN, ENAM, and SPARCL1 were calculated as described above for AMTN and ODAM. The dataset comprised AMEL

sequences, 30 AMBN, 28 ENAM, and 20 SPARCL1 (entire and partial sequences), and those obtained here from 10 AMTN and 10 ODAM (fig 1, 2). The N-terminal region of SPARCL1 was only used because EMPs and the other SCPPs are supposed to be derived from this region [Kawasaki et al., 2004]. The N-terminal regions of these putative ancestral sequences were aligned to the same region of AMEL (i.e. the first 62 residues, from exon 2 to the TRAP proteolytic site at the beginning of exon 6) with CLUSTALX and hand-checked using Se-Al 2.0. The phylogenetic analysis was performed using maximum likelihood (neighbor-joining method) in PAUP 4.0 and the tree was rooted on SPARCL1, since this is the probable ancestor of the SCPPs. This analysis confirms with a good statistical support the previous finding that AMEL and AMBN are sister genes [Sire et al., 2006] (fig. 5). The two newly identified ameloblast-expressed genes, ODAM and AMTN, appear as two sister genes (this is well supported statistically), and their group is the sister group of the AMEL/AMBN group. ENAM is the sister gene of the two groups AMEL/AMBN + ODAM/AMTN, and SPARCL1 is the sister gene of the three. However, the relationships of ENAM and SPARCL1 are not strongly supported by our bootstrap analysis. This phylogenetic analysis means that AMEL/AMBN and ODAM/AMTN have a common ancestor, which was probably issued from a duplication of the ENAM ancestor, itself deriving from a copy of the SPARCL1 ancestor.

This phylogeny corresponds to our relatively weak knowledge of ameloblast-expressed genes and must be interpreted with caution. Indeed, even though a large number of sequences were used, most of them are from mammals, and even from eutherians only. Only a few AMEL and AMBN sequences are available in reptiles and amphibians, and no ENAM, AMTN, and ODAM sequences are known in these lineages. This lack of data in non-mammalian lineages does not allow to obtain representative putative ancestral sequences at the amniote and tetrapod levels. This means that the phylogenetic signal (i.e. gene relationships) is probably reduced by (i) the long evolutionary period (hundreds of million years) that separates each gene from its closest relative, (ii) the different evolution rate for each gene in each lineage, and (iii) the rapid divergence of some gene regions in relation to their proper functions. This phylogeny will become more accurate in the near future, when more ameloblast-expressed SCPP gene sequences will be known in reptiles and amphibians. Nevertheless, the present analysis supports AMBN/AMEL relationships and the hypothesis that both genes derive from ENAM. It furthermore indi-

Fig. 6. Current probable scenario for the origin and evolution of SCPP genes and, in particular, of ameloblast-expressed genes (AMEL/AMBN, AMTN/ODAM, and ENAM). Early in deuterian evolution, SPARC duplicated into SPARCL1. During successive rounds of genome and gene duplication, SPARCL1 and its descendants were copied several times on the same chromosome, giving rise to two clusters: the ameloblast-expressed/milk/saliva protein gene cluster and the bone/dentin protein gene cluster (SIBLINGs). The ENAM ancestor duplicated from an SCPP ancestor and one ENAM copy was duplicated again, giving rise to the ancestors of AMBN/AMEL and of AMTN/ODAM. After its duplication from AMBN, AMEL was translocated to another chromosome.
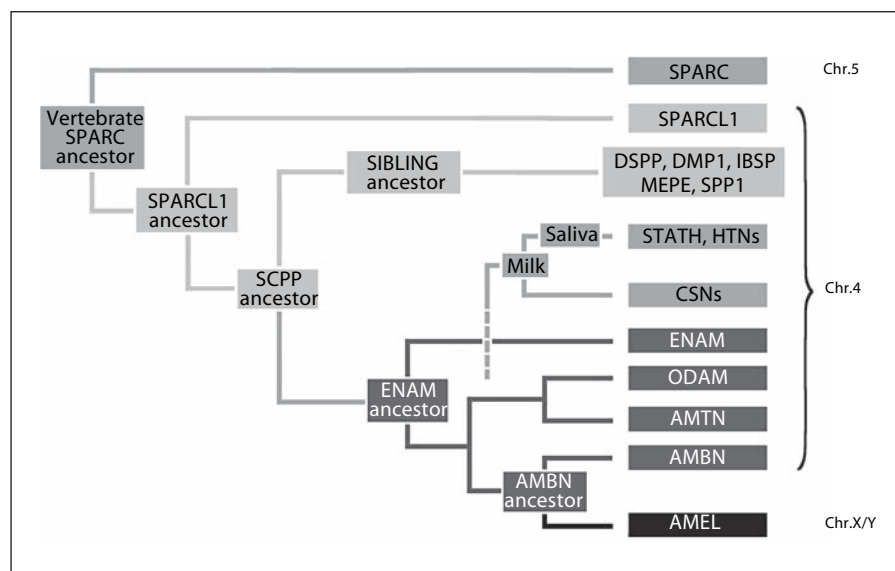
cates that ODAM and AMTN could also be derived from ENAM. This implies that an additional duplication event has occurred between ENAM and the other ameloblast-expressed SCPP genes (fig. 6).

A preliminary, schematic scenario for SCPP evolution and for the place of the ameloblast-secreted actors (to which AMTN and ODAM are now added) can be drawn, but the story is far from complete (fig. 6). In particular, the relationships between SPARCL1 and the two gene clusters (SIBLINGs and enamel-milk-saliva protein genes), and among the SIBLINGs are not established. In contrast, within the salivary SCPPs, histatins 1 and 3 derive from statherin duplication, and the latter was created from a copy of a milk casein ancestor (CSN1S2) [Kawasaki and Weiss, 2003]. The evolutionary story of salivary SCPPs is relatively recent (they are known in some eutherians only), while the origin of milk caseins is more ancient in mammalian evolution. Indeed, α-, β- and κ-caseins are identified in the milk of metatherians (marsupials) [Ginger et al., 1999; Stasiuk et al., 2000]. Milk casein family members are also evolutionarily related and, given their structural similarity with EMP genes, the ancestral Ca-sensitive casein gene was probably derived from the duplication of an EMP [Kawasaki and Weiss, 2003], which remains to be found (fig. 6).

In summary, depending on the branches of the tree, SCPP relationships are either strongly or weakly supported. Strong relationships are: SPARC/SPARCL1; STATH/HTHs; CSN/STATH/HTHs; AMEL/AMBN, and AMEL/AMBN/ENAM. In contrast, there are (i) no clear rela-

tionships established within the SIBLING cluster, and between this cluster and SPARCL1; (ii) no clearly identified connection between CSNs and EMPs; (iii) weak (lack of non-mammalian sequences) relationships between ODAM/AMTN, and ENAM/ODAM/AMTN, and (iv) no clear relationship between the ameloblast-expressed genes (AMEL/AMBN, ODAM/AMTN, and ENAM) and SPARCL1.

Sequencing these SCPP genes in non-mammalian species [reptiles (crocodiles, lizards, and snakes) and amphibians (salamanders, caecilians, and frogs)] will help to improve our knowledge on the relationships in the family.

Dating of AMBN/AMEL Duplication

Now that AMEL and AMBN are clearly established sister genes, the last questions are: was the ancestral gene AMBN or AMEL and is it possible to date this duplication event? The stronger support to AMBN ancestry is indirectly suggested by the location of AMEL on sex chromosomes. Indeed, it is difficult to imagine that an AMEL copy (that would have become AMBN) was translocated by mere chance, on the chromosome housing the other SCPP genes, and close to ENAM, their close relative. In contrast, the close location of AMBN and ENAM on the same autosomal chromosome (fig. 4) strongly supports that AMBN was created from a copy of ENAM, and, as a consequence, that AMEL originated after a duplication of the ancestral AMBN, and then translocated to another chromosome. One could argue that AMEL translocation
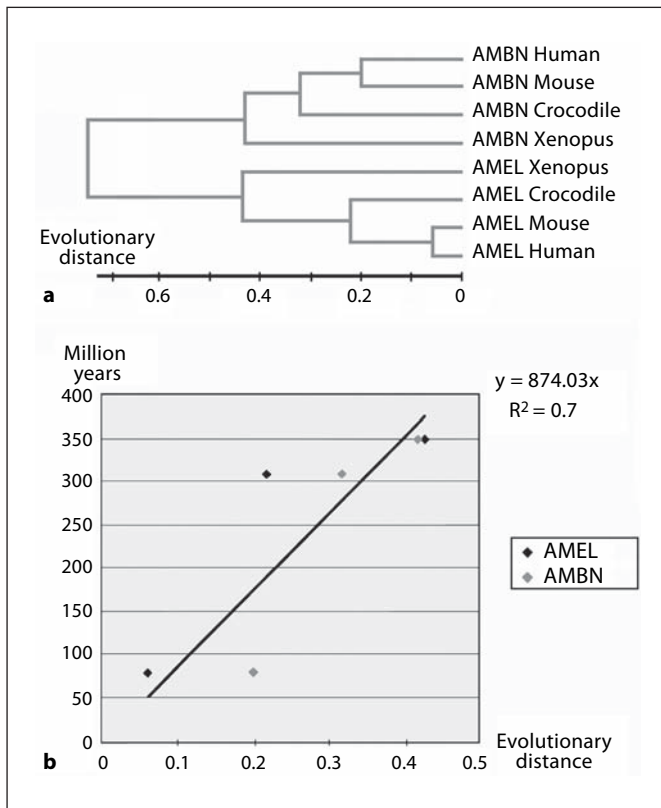
Fig. 7. a Linearized tree obtained from the phylogenetic analysis of AMBN and AMEL sequences in human, mouse, crocodile, and *Xenopus*. The calibration time used is: human/mouse: 90 MYA; human/crocodile: 310 MYA; human/*Xenopus:* 360 MYA [Hedges, 2002]. **b** Linear regression of time versus distance (y-x). Each point has two evolutionary distances of AMBN and AMEL. The duplication time of AMBN/AMEL can be estimated when we add the evolutionary distance of duplication to this linear equation, i.e. it occurred >600 MYA.

occurred after its duplication from the ENAM ancestor and that the copy remained close to ENAM and differentiated into AMBN. This scenario cannot be maintained since the similarities found in gene organization (fig. 3) and in amino acid pattern indicate that AMBN is closer to ENAM than AMEL is. Therefore, AMBN is the 'mother' of AMEL and not the opposite.

In summary, all ameloblast-expressed genes are phylogenetically related, and ENAM could be the ancestor of all of them. AMEL, which codes for the major protein of the forming enamel matrix in mammals (90% of the protein content) is the youngest EMP gene. This strongly suggests that AMEL divergence after AMBN duplication was an important innovation for enamel, at least in mammals. To date, the relationships of EMP genes with SPAR-

CL1 are difficult to establish and more data are needed to test the hypothesis of SPARCL1 ancestry.

The availability of AMEL and AMBN sequences in various mammalian species, in a crocodile and in an amphibian *(Xenopus)* allowed to envisage a molecular dating of AMBN/AMEL duplication. A phylogenetic tree was inferred from the amino acid sequences using the neighbor-joining method (fig. 7a). From the phylogeny, it is apparent that the duplication event was much earlier than the speciation events such as the mammal/amphibian split, or the mammal/reptile split, and roughly two times of these events. To give an approximate estimate of when this duplication event occurred, we utilized the molecular dating technique developed by Gu et al. [2002], calibrated by the fossil record: primate/rodent split (around 90 MYA), mammal/reptile split (310 MYA), and amniote/amphibian split (360 MYA) [Hedges, 2002]. Our results are as follows.

1 If the amniote/amphibian split is used alone, the date of duplication (T) = 627 MYA.
2 If the mammal/reptile split is used alone: T = 896 MYA.
3 If the primate/rodent split is used alone: T = 480 MYA.
4 If all three calibrations are used: T = 682 MYA.

This is a molecular dating of gene duplication, so it should be compared to other molecular date profiling [Gu et al., 2002]. Here, (2) and (3) are unreliable because the distance between human-mouse or human-crocodile differs considerably in AMBN/AMEL genes. In contrast (1) is mostly reliable and (4) takes the average, but both give similar results, i.e. AMBN/AMEL duplication occurred >600 MYA (fig. 7b). This result confirms the previous dating of AMEL origins during the Precambrian period [Delgado et al., 2001]. A major peak of genome and gene duplication occurred around 700–500 MYA [Gu et al., 2002]. Therefore, like many developmental genes, EMPs were duplicated during this period, which preceded vertebrate diversification and skeletal mineralization.

In summary, two unrelated molecular dating methods of EMP origins (SPARC/SPARCL1 divergence date: Delgado et al. [2001] and AMBN/AMEL duplication date: this study) indicate that the genes encoding them were created from several duplication rounds that have occurred before the currently accepted dates of the appearance of the first vertebrates in the fossil record (>600 MYA). In contrast, the molecular dating of SPARC/SPARCL1 divergence proposed by Kawasaki et al. [2004] supports an emergence of EMPs after the di-

vergence of cartilaginous fish (approximately 500 MYA Kumar and Hedges [1998]). The knowledge of the divergence date of SPARC/SPARCL1 is of importance as SPARCL1 is considered the probable ancestor of SCPPs. However, the apparent different evolutionary rates of SPARC and SPARCL1 in various taxa, together with the fact that various gene regions were compared within each species or each clade, does not allow an accurate prediction of the divergence date. Indeed, these two paralogs share a well-conserved C-terminal region which is not easy to differentiate from one gene to the next in the vertebrate species examined. In contrast, their N-terminal region is not only extremely different but also, when comparing this region in various species, difficult to align due to a large number of sequence variations. Nevertheless, the N-terminal region of SPARCL1 is considered the probable ancestor of SCPPs. The divergence date of AMBN/AMEL seems to be more reliable because the relationships of these two genes are now well established. Also, the presence of enamel-like tissues in early vertebrates indicates that the divergence of SCPP genes might have preceded the origin of vertebrate tissue mineralization.

It is important to realize the following.

(i) The molecular dating of AMBN/AMEL duplication does not indicate the presence of these molecules in forming enamel, 600 MYA. After the duplication, several dozens of millions of years were probably necessary before one copy acquired its new function (new gene structure and new expression). This divergence could have occurred before, during or after the vertebrate diversification, reported to be in the Cambrian as demonstrated in the fossil record. Moreover, genetic evidence suggests that most animal phyla evolved dozens of millions of years before they started to leave behind fossil evidence, although this is debated by paleontologists. Given the lack of a temporal association between the birth of a gene (e.g. AMEL 600 MYA) and the advent of mineralized 'teeth' >50–100 millions of years later, the confidence in the assigned dating should be softened.

(ii) Tissue mineralization could not have occurred if the necessary tools were not already present. This implies that EMPs could have had other functions before the first enamel/enameloid tissues mineralized and before EMPs were recruited for mineralization later in vertebrate evolution. This novel trait (mineralization) therefore probably evolved by employing already existing materials.

Enamel/Enameloid and the Origin of EMPs

Morphological studies of enamel and enameloid in living taxa have shown that they are different in their mode of formation. The enamel organic matrix is secreted by the ameloblasts, and contains enamel-specific proteins. In contrast, enameloid organic matrix is mostly deposited by odontoblasts and contains a large amount of collagen, but the ameloblasts contribute to its formation, too [Prostak and Skobe, 1984; Sasagawa, 1984; Prostak and Skobe, 1988; Prostak et al., 1993; Sasagawa, 1995, 2002]. However, in functional teeth, the structure of both tissues is similar, i.e. highly mineralized with only a little organic matrix left (<5%). Given the same location, the same final structure and the same evolutionary origin, most authors have considered enamel and enameloid as homologous tissues. Enamel and enameloid matrices are only partially mineralized when laid down, and their final hardness is acquired during a second stage, maturation, during which the matrix is lost through the activity of proteolytic enzymes. This process creates space, allowing mineral crystal growth to eventually achieve a highly mineralized structure. Because they are highly mineralized, enamel and enameloid are easily recognizable in the fossil record and their relationships can be traced back deep in vertebrate evolution.

The question of which tissue appeared first, enamel or enameloid, has been long debated and it is not clearly answered yet. It is, however, accepted that enamel progressively replaced enameloid during evolution in various lineages (e.g. in tetrapods) [Smith, 1995; Donoghue, 2002; Donoghue and Sansom, 2002; Donoghue et al., 2006]. Odontoblasts progressively reduced their production of loose collagenous matrix, which characterizes forming enameloid, while ameloblast activity increased with the secretion of large amounts of enamel-specific products at the dentin surface. This evolutionary 'transition' between enameloid and enamel was, in fact, probably an enameloid-dentin transition, as recently demonstrated in the ontogeny of caudate amphibians [Davit-Béal et al., 2007]. However, enamel did not replace enameloid in all vertebrate lineages. A particular type of enameloid is present in chondrichthyans (cartilaginous fish [Prostak et al., 1993; Sasagawa, 2002]), and this supports an ancient origin for this tissue, at least for the gnathostome lineage. Enamel and enameloid were certainly present in basal actinopterygians (ray-finned fish), as in polypterids and lepisosteids [Sire et al., 1987; Sire, 1990, 1994, 1995]. This supports the idea that enamel was already present in early osteichthyans, which also indicates an ancient origin.
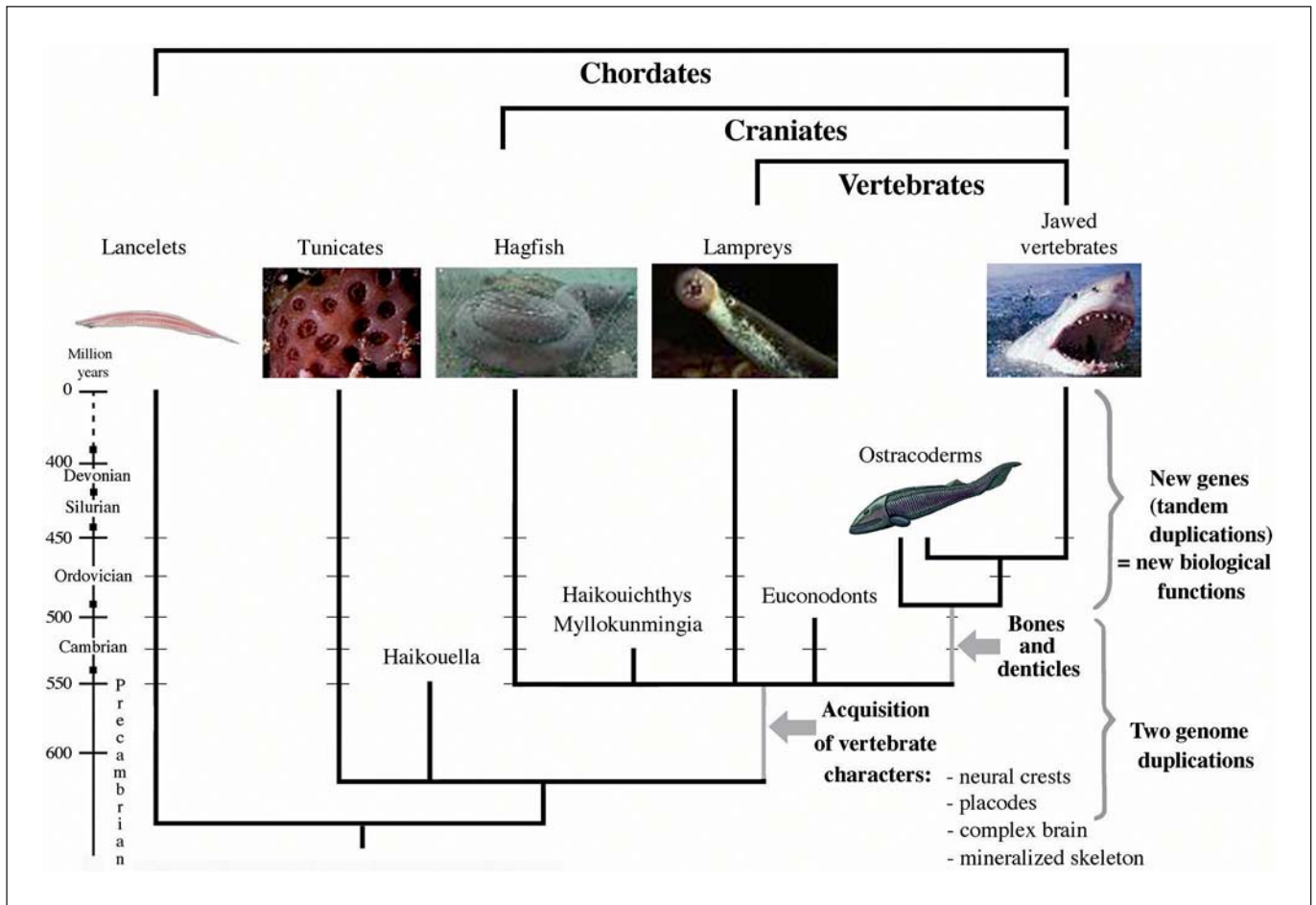
Fig. 8. Chordate relationships and the origin of the mineralized skeletal elements in vertebrates (adapted from Shimeld and Holland [2000]). Chordates are deeply anchored in the Precambrian era (>700 MYA). The acquisition of a mineralized skeleton, a major event for vertebrate radiation, occurred 600–500 MYA, a period which post-dates the two genome duplications [Gu et al., 2002]. Bone and dental tissues are clearly recognized in early, jawless vertebrates, 450 MYA. Skeletal diversification in jawed vertebrates was next favored by the appearance of new genes after tandem duplication.

Enamel is absent in more derived actinopterygian taxa (teleost fish), which possess enameloid only [Sasagawa, 1984; Prostak and Skobe, 1984; Sasagawa, 1995]. The large evolutionary distance between all living representatives of these chondrichthyan and actinopterygian lineages (430–420 MYA, respectively, in the fossil record: Janvier [1996]) explains why the current structure of these enameloids is so different.

Enamel and enameloid appear, therefore, to be merely grades of a hypermineralized tissue that has evolved independently in a number of vertebrate lineages [Donoghue, 2001]. The origin of these tissues can be traced back in early vertebrates, along with the appearance of a bony mineralized skeleton, one of the four main vertebrate character acquisitions, together with neural crest cells and their derivatives, neurogenic placodes, and an elaborate segmented brain (fig. 8). These vertebrate innovations appeared after the divergence between tunicates[7] *(Ciona)* and craniates[8] (recent genetic evidence indicates that tunicates could be closer to vertebrates than cephalochordates [Graham, 2004]), and probably after the divergence between craniates and vertebrates as witnessed by the fossil record. The absence of mineralized tissues in living hagfish and lampreys is probably primitive [Jan-

---

7   Tunicates: subphylum of chordates that feed by siphoning plankton through a filter.
8   Craniates: animals with skull.

Fig. 9. Enamel/enameloid tissues during vertebrate evolution (as reported in the fossil record), and current knowledge of the presence of EMP and SCPP genes in vertebrate lineages. Enamel-like tissues are identified in early vertebrates, the euconodonts, and they display a different evolutionary history in the various lineages. Enameloid was conserved in chondrichthyan and actinopterygian lineages, but disappeared in amniotes. The early presence of enamel/enameloid tissues in vertebrate evolution strongly suggests that EMP divergence predates this time (>500 MYA). However, there is a large gap between this theoretical EMP presence in early vertebrate lineages and the current knowledge of the genes coding for these proteins, which is restricted to the tetrapod level (350 MYA). SCPPs are known, however, from actinopterygian fish.
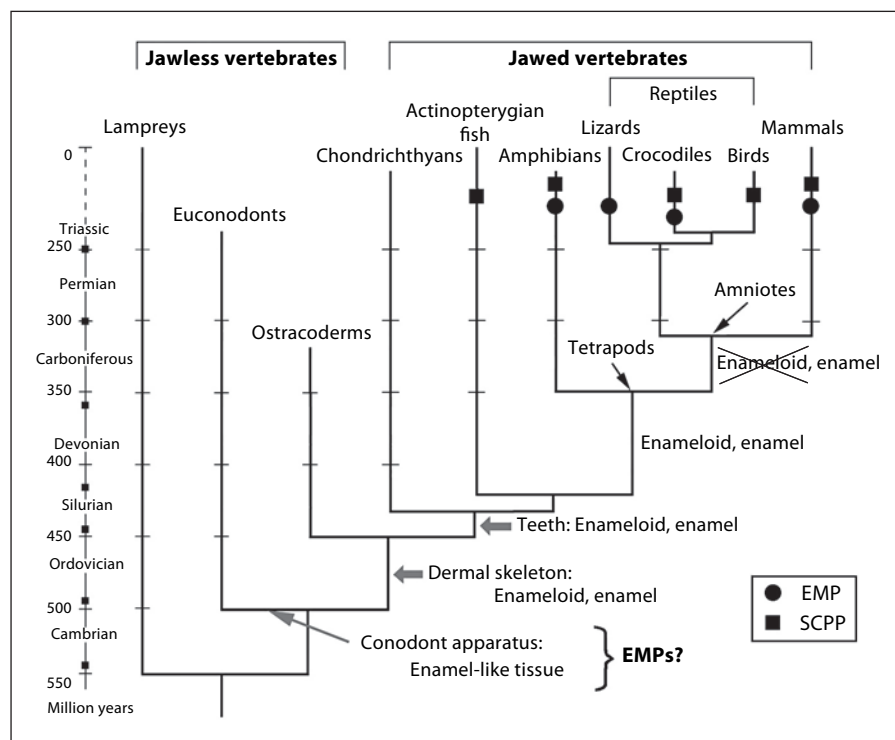
vier, 1996]. Indeed, the most ancient vertebrates discovered in the Lower Cambrian of China (530 MYA), *Haikouichthys* (which looks like a hagfish) and *Myllokunmingia* (which looks like a lamprey), possessed a skeleton composed of unmineralized cartilage only [Shu et al., 1999, 2003].

The first mineralized elements encountered in vertebrates are the tooth-like organs (conodont apparatus) composed of enamel-like and dentine tissue found in euconodonts, fossil marine vertebrates known from the Middle Cambrian (500 MYA) to the Late Triassic (230 MYA) [Sansom et al., 1992, 1994; Janvier, 1996; Donoghue, 1998, 2001] (fig. 9). These minute comb-shaped denticles are located at the entrance of the pharynx (viscerocranium). Bone appears to be absent from these elements [Donoghue, 1998].

Enamel, or enameloid, is clearly identified in the skeleton of early jawless vertebrates (e.g. pteraspidomorphs, heterostracans, thelodonts, and 'ostracoderms') from the Early Ordovician (480 MYA) to Late Devonian (380 MYA) periods and of jawed vertebrates (early chondrichthyans and osteichthyans) [Janvier, 1996; Donoghue et al., 2006] (fig. 8). The earliest skeleton was a dermal skeleton comprising odontodes (tooth-like elements consisting of enameloid and dentine), ornamenting dermal

plates composed of acellular bone [Sansom et al., 2005]. It is noteworthy that our current knowledge of early vertebrates reveals a gap of 30 million years between the appearance of the first vertebrates (530 MYA) and the first evidence of vertebrate mineralized elements (500 MYA).

It is clear that numerous gene families expanded by gene duplication in the vertebrate stem lineage (in particular gene families encoding transcription factors and signaling molecules) [Shimeld and Holland, 2000]. The acquisition of the mineralized skeleton followed the increased genetic complexity (two genome duplications and several gene duplications) which occurred early in vertebrate evolution (during the Precambrian and Cambrian periods) [Dehal and Boore, 2005; Panopoulou and Pouska, 2005]. These large scale genomic events facilitated the evolutionary success of the vertebrate lineage and, probably, led to the diversification of several members of the SCPP family. Additional tandem duplications certainly occurred during the long period of vertebrate evolution and resulted in new gene differentiation and in a further diversification of SCPPs into new biological functions (fig. 8).

The presence of enamel and enameloid tissues in early vertebrates strongly suggests that EMPs (and some other SCPPs) were present in these tissues at least 500 MYA

(fig. 9). This would mean that SCPPs diversified earlier. The hypothetical date of this diversification could be not so distant from the molecular dating of EMP origins (>600 MYA) if we consider that the duplication could have occurred long before the divergence of function/expression of the copies, and that vertebrates possessing a mineralized skeleton could have lived dozens of millions of years before any evidence of them in the fossil record. However, although structurally well-identified enameloid and enamel tissues are present in the teeth of chondrichthyans, actinopterygians, and basal sarcopterygians, EMP genes are known in tetrapods only (fig. 9). However, this statement relates to genes only; there is evidence from immunohistochemical studies or Southern hybridization that AMEL and/or ENAM proteins could be present in sharks [Slavkin et al., 1983; Herold et al., 1989], teleost fish [Lyngstadaas et al., 1990], polypterids [Zylberberg et al., 1997] and lungfish [Satchell et al., 2000].

Whilst the data on EMP genes (mainly in model mammals) slowly accumulated over a period of approximately 15 years, the last years witnessed a rapid increase in our knowledge, mainly because of genome sequencing in numerous species, and in particular in mammals. To date eight well-covered mammalian genomes are available and seven additional genomes are provided at a low coverage level (see http:/www.ensembl.org/). The current mammalian genome project aims to add 11 mammalian species to this list in a phylogenetic perspective (http:/www.broad.mit.edu/mammals). Therefore, within the next few months, we will have access to at least 26 mammalian genomes and, potentially, will be able to perform evolutionary analyses of any gene in the mammalian lineage. Opposite to this large covering of mammalian phylogeny, our knowledge of non-mammalian EMPs is, unfortunately, much less advanced (fig. 10). We can see two reasons: (1) the lack of sequenced genomes and (2) the divergence of EMP sequences.

The Lack of Sequenced Genomes

In toothed reptiles (crocodiles, snakes, and lizards), there is still no sequenced genome available, although the reptilian (sauropsid) lineage is the lineage closest to mammals (fig. 10). However, AMEL sequences are available in a crocodile [Toyosawa et al., 1998], in a snake [Ishiyama et al., 1998], and in two lizards [Delgado et al., 2006; Wang et al., 2006], and AMBN has been sequenced in a crocodile [Shintani et al., 2002]. At present, there are no data on reptilian ENAM but, fortunately, we will soon have access to a lizard genome (*Anolis carolinensis* genome is being sequenced). However, sequencing a croco-

dile genome (a representative of the lineage closest to birds) would be also extremely interesting for evolutionary analyses.

In amphibians, AMEL [Toyosawa et al., 1998] and AMBN [Shintani et al., 2003] have been sequenced in the pipid frog *Xenopus laevis,* and an AMEL sequence is available in another frog (*Rana pipiens* [Wang et al., 2005a]). Moreover, sequencing of a pipid genome *(Silurana tropicalis)* is well advanced (fig. 10). Surprisingly, although as expected AMEL and AMBN are present in this genome, to our knowledge ENAM has not been found yet [Kawasaki and Weiss, 2006]. It is questionable whether this EMP is really absent from this frog genome. Indeed, on the one hand our evolutionary analysis indicates that ENAM is the oldest representative of the EMP family and, on the other hand, ENAM plays important roles in enamel structure and organization as illustrated by AIH2 resulting from ENAM mutations. It is also clear that pipids have a well-formed enamel [Sato et al., 1986]. Therefore, this 'lack' is probably related to the fact that the pipid genome is still not entirely (or correctly) assembled. One should also take into consideration that pipids are highly derived anurans and, as a consequence, EMPs could be divergent compared to more basal amphibian species. Sequencing another frog, salamander/newt or caecilian genome would be, therefore, highly informative for evolutionary analysis.

No EMP is known in basal sarcopterygians, i.e. lungfish and coelacanth, nor in basal actinopterygians (polypterids and lepisosteids), and there is no sequenced genome available nor sequencing project running. However, these taxa possess enamel and they belong to lineages that are crucial to improve our understanding of EMP relationships and evolution. In contrast to this lack of data, the genome has been sequenced in four teleost species, and several SCPPs were identified. However, teleosts are derived actinopterygian lineages, and the long evolutionary distance (>420 million years) between actinopterygians and tetrapods explains the difficulty encountered when trying to identify homology between teleost and tetrapod SCPP genes [Kawasaki et al., 2005]. For instance, no EMP gene can be related to these SCPPs.

No SCPP is known in chondrichthyans (sharks and rays). Here too, the long evolutionary distance (>430 million years) between cartilaginous fish and tetrapods could lead to problems when trying to identify homologous genes, but the synthemy conservation of SCPP genes could help [Kawasaki et al., 2005; Kawasaki and Weiss, 2006].

Fig. 10. Current knowledge of EMP genes in vertebrates. To date only two EMPs are characterized at the tetrapod level (AMBN and AMEL). ENAM is only known in mammals. The lack of data in non-mammalian lineages is clearly related to the absence of sequenced genomes. SCPP genes are identified in teleost fish, but the large evolutionary distance makes their relationships to EMPs uncertain. EMP genes on gray background are potentially accessible to sequencing. Question marks indicate lineages in which sequencing of EMP genes might be a priority to improve our understanding on their origin and evolution. * = Large DNA regions (Whole Genome Shotgun) have been sequenced in a lizard (*A. carolinensis*).
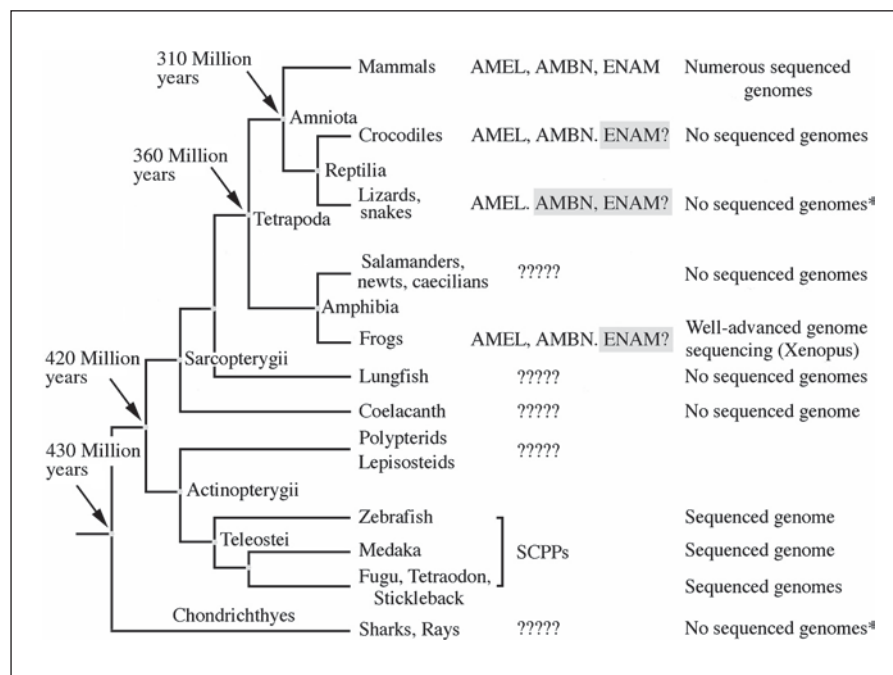
The Divergence of EMP Sequences

The difficulty to find EMP (and other SCPP) genes using PCR or RT-PCR resides in their variability. Indeed, except for the short N-terminal region that is relatively well conserved in each member of the family, the largest part of the sequence is variable. For instance, although they probably conserve their main function, most of the mammalian AMEL exon 6 sequences (the largest part of AMEL) cannot be accurately aligned with the homologous region in reptiles and amphibians due to numerous substitutions and indels [Sire et al., 2006]. These highly variable sequences indicate that SCPPs are intrinsically disordered proteins [Dunker et al., 2001; Kawasaki et al., 2005] and there are only a few conserved residues. Therefore, the only means to find EMPs in evolutionary distant species, such as basal sarcopterygians or actinopterygians, is to study sequenced genomes or sequences of large DNA regions suspected to house these genes. For example, in a teleost fish (fugu), several SCPP genes were identified in a DNA region corresponding to the SIBLING cluster in mammals, meaning that the syntheny of the SIBLING cluster is conserved between fish and tetrapods [Kawasaki et al., 2005; Kawasaki and Weiss, 2006]. These SCPP genes were found not based on their similarity with known SCPP sequences but because they are located adjacent to SPARCL1, and because they share some structural features with tetrapod SCPPs. Fish SCPP genes are so different from tetrapod SIBLINGs that no homology could be recognized. Fish SCPP genes are expressed during tooth formation [Kawasaki et al., 2005] but one can wonder whether they play the same function as EMPs. Moreover, SIBLINGs (DSPP, DMP1, IBSP, and SPP1) are known to be expressed during tooth matrix formation in tetrapods [Fisher and Fedarko, 2003; Qin et al., 2004]. EMP genes could also be conserved in other regions of the teleost fish genome, but they remain to be discovered. Indeed, morphological studies strongly support that EMPs are present in the enamel-like tissue (ganoine) of basal actinopterygian lineages, polypterids and lepisosteids [Sire et al., 1987; Sire, 1994; 1995].

To date the information available for the three EMP genes largely relates to mammals and the few sequences available (or planned to be so) in other tetrapods are not sufficient to perform an evolutionary analysis at this level (fig. 10).

What Can the Evolutionary Analysis of EMP Genes Tell Us? The Case of AMEL

AMEL Evolution

AMEL is the main component of forming enamel and it plays crucial roles in enamel structure and mineralization [Diekwisch et al., 1993; reviews in Bartlett et al.,

2006b; Margolis et al., 2006]. Mutations of the encoding gene lead to AIH1 [Hart et al., 2002; Kim et al., 2004]. Given this importance it is not surprising that AMEL is the best-known EMP. Over the past years, AMEL studies on model animals have provided information on the gene structure and supposed functions of the various regions of the protein [Fincham et al., 1991; Fincham and Moradian-Oldak, 1995; Greene et al., 2002]. AMEL is subject to posttranslational modifications [Fincham and Moradian-Oldak, 1993] and it self-assembles to form nanospheres that are involved in enamel mineralization [Wen et al., 2001; Snead, 2003; Du et al., 2005; Veis, 2005]. The N- and C-terminal regions interact with mineral [Aoba et al., 1989; Aoba, 1996; Hoang et al., 2002; Paine et al., 2003; Snead, 2003] and are involved in adhesion with the ameloblast surface through membrane proteins (e.g. Cd63, annexin A2, and Lamp1 [Wang et al., 2005b; Tompkins et al., 2006]). AMEL interacts also with some keratins in ameloblasts through ligand-binding properties located in the N-terminal region [Ravindranath et al., 1999, 2000, 2001, 2003]. Some splice products have been proposed to be signaling molecules [Veis et al., 2000; Veis, 2003].

From these studies, increasing evidence accumulates to support the idea that the N-terminal, and to a lesser degree the C-terminal, regions are the most important regions for proper AMEL function. This importance is also revealed by several AIH1, caused by mutations modifying the functioning of these regions. The question of a possible role for the central variable region (encoded by most of exon 6) is completely ignored. Is it useless? Certainly not. Evolutionary analyses indicate that this core region of the protein, although intrinsically disordered, could be responsible for the well-ordered microstructure of enamel [Delgado et al., 2005; Sire et al., 2005; 2006]. More data are still needed to understand the relationships between structure and function of this region and, more generally, to reveal the amino acid positions and regions that could play an essential role.

As an alternative to biochemical and in vitro approaches, an evolutionary analysis of mammalian AMEL was performed using 56 sequences constituting a dataset representative of mammalian diversity [Delgado et al., 2005]. Here, we summarize and complete these results in proposing two alignments (fig. 11): one, illustrated with 20 sequences of the N- and C-terminal regions only, reveals the numerous well-conserved residues that are important for the proper function of the protein (interactions with the cell membrane and/or with mineral crystals). The other alignment, comprising 51 sequences, is centered in the variable central region of exon 6, which houses, in mammals, a hot spot of mutation. The putative ancestral sequence has been calculated for both alignments. Briefly, this evolutionary analysis reveals the following points.

(i) A total of 56 residues (out of 74 in the full-length sequence) have remained unchanged in the N- and C-terminal regions of AMEL during mammalian evolution, i.e. during 225 million years [van Rheede et al., 2006] (fig. 11a). This indicates that strong functional constraints act on these amino acids, meaning that they certainly play, either alone or with other conserved residues, an important role. Most variants are found in the C-terminal region of exon 5.

(ii) The hot spot of mutation (large insertions/deletions of residues) has appeared recently in mammals, and independently in several lineages (fig. 11b). Insertions are found in basal primates (lemurs), in tree shrews, in basal rodents (squirrel and guinea pig), in bovids (cow and goat) and cervids (deer), in only one family of carnivores (ursids), in bats (Macrochiroptera), in insectivores (hedgehog), in afrotherians (elephant shrew), and in marsupials (opossums). The perissodactyls (e.g. horse) and prototherians (platypus and echidna) are the only important lineages in which such large insertions are absent. These insertions contain a variable number of three amino acid (triplet) repeats (e.g. PIQ-PMQ-PLQ). These triplet repeats range from two (in the tree shrew) to 12 (in a fruit bat), in which a total of 36 residues (108 bp) are inserted. Within some lineages, e.g. bovids, the number of repeats can vary in closely related species (8 repeats in the African buffalo, 7 in cattle, and 5 in the other members of the family). It is noteworthy that AMELY, that is expressed at a low level in forming enamel (less than 10% [Salido et al., 1992]), does not show insertions in this region. This illustrates the separate evolution of the two AMEL copies on sex chromosomes [Girondot and Sire, 1998], AMELY being subjected to the particular mode of evolution of the Y chromosome [Iwase et al., 2001; Lahn et al., 2001; Iwase et al., 2003]. The lack of triplet insertions in AMELY versus AMELX exon 6 allows to easily discriminate males from females in lineages possessing the hot spot of mutation, e.g. bovids [Weikard et al., 2006] and ursids [Yamamoto et al., 2002]. Large deletions (≥9 residues) are found in dolphin, Weddell seal, panda and roundleaf bat (Microchiroptera). However, we do not know whether these indels have a consequence on enamel microstructure in these species [Delgado et al., 2005]. It is clear, however, that the conservation of such large indels during evolution has no negative results on enamel function as protective tissue.

```
                   exon2       |   exon3    |            exon5    |         exon6      |  ------      exon6  | ex7
AMEL_Ancestral   MGTWILLACL LGAAFAMPLP PHPGHPGYIN FSYEVLTPLK WYQNMIRQQY PSYGYEPMGG WLHHQIIPVL      DLPLEAWPAT DKTKREEVD
Human            .......... .......... .......... .......... ...S-..PP. .......... ..........      ..T.....S. .........
Squirrel_monkey  .......... .......... .......... .......... ...S..PA.. .......... ..........      .......... ---------
Lemur            .......... .......... .......A... .......... ...S...PP. .......... ..........      .......... ---------
Galago           .......... .......... .......... .......V. ...S.L.PP. .......... ..........      .......... ---------
Mouse            .......... .......... ...S...... L......... ...S....P. .......... ..........      E......... .........
Guinea_pig       .......... ..T....... .......... .......... ...S....P. .......... .V...V....      .......... .........
Squirrel         .......... .......... .......... .....I..F. ..........P. .......... ..........      .......... .........
Goat             .......... .....S.... .......... ....P.... ...S...HP. .......... .........V      .VL..D.... .........
Cow              .......... .....S.... .......... .......... ...S...HP. .......... .........V      .......... .........
Pig              .......... .....S.... .......... .......... .......HP. T......... .........V      .......... .........
Horse            .......S. ......I... S......... .......... ...SL...P. T......... ..........      .......... .........
Dog              .......... .......... .......... .......... .......HP. .......... ..........      .......... .........
Flying_fox       .......... .......... .......... ...****** ********** .......... ..........      .......... .........
Hedgehog         .......... .......... .......... .......... .......PP. .......... ..........      .......... .........
Elephant         .......... .......... .......... .......... .......... .......... ..........      .......... .........
Tenrec           .......... .......... .......... .......... ....L..P. .......... ..........      ...M...... .........
Hyrax            .R...L.... .......... .......... .......I. .......-P. .......... .S........      .M........ .........
Opossum          .R...L.... ......I... .......... .......... ...S.M.HE. .......... ..........      .M--...... .........
Wallaby          .R....L... ......I... .......... .....F.. ...S.M.-. .......... ..........      ********** ********
Platypus         .......T. I.....I... ...A...... .........Q ....K.... .......... ..........      .....Q.... .........  a


AMEL_Ancestral   PNLPQPAQQP YQ----PQPP Q-PQ--PHQP IQ------- ---------- ---------- --------PQ APVHPMQPLP PQ-PPLPPMF
Human            ....P..... .........V M......... .......... .......... .......... P......... ..........
Orangutan        ..LP..... .........V M......... .......... .......... .......... P......... ..........
Squirrel_monkey  ....P..... F.......TV M......... .......... .......... .......... P......... ..........
Lemur            ....P..... F........V M.PMQPMQPM QPIQPIQPIQ .......... .......... P.L....... ....H...L.
Galago           ....P..... F........V M.PMQPMQPM QPIQ...... .......... .......... P......... .......L.
Marmoset         ....P..... F.......TV M......... .......... .......... .......... S......... ..........
Tree_shrew       ....P..... F........V ..PIQPIQ.. .......... .......... .......... T......... ..........
Flying_lemur     S.I.M...P. .........V ..L....... M......... .......... .......... P.....H... .......L.
Mouse            ..I.PS.... F.QPFQ..AI M......... .......... .......... .......... S.L......A .......L.
Hamster          ..I.PS.... F.QPFQ...I S.....S... M......... .......... .......... S.L......A .......L.
Guinea_pig       ....PTS... F.QPFPT..V ......H... ..PIQPIQPI QPIQPIQ... .......... S.L..I.... ...QA.....
Squirrel         ....P..... F.QPFQ..SI ......S... M.PMQPMQPM QPMQPVQ... .......... P.L.SLH... ...H......
Goat             ....L..... F.......SI ......H... L.PLQPMQPL QPLQPLQ... .......... P....I.... ......L.I.
Sheep            *******... F.......SI ..S....... L.PLQPMQPL QPLQPLQ... .......... S....I.... .........I
Cow              ....L..... F.......SI ......H... L.PLQPMQPL QPMQPLQPLQ PLQ....... P....I.... .........I
African_buffalo  ....L..... F.......SI ......H... L.PHQPLQPM QPMQPLQPLQ PLQPLQ.... P....I.... .........I
Japanese_serow   ....L..... F.......SI ......H... L.PLQPMQPM QPLQPLQ... .......... P....I.... .........I
Deer             ********... F.......SI ......H... L.PLQPLQPL QPLQPLQ... .......... P....I.... .........I
Pig              ....L..... F........V ......H... L......... .......... .......... S.M..I...L ..........
Hippopotamus     ..F.L..... F........I .......H L......... .......... .......... K P.M..I...L ......S...
Dolphin          .H..V....- --....--.V .........-. -- ........ .......... -- -------.L ..........
Porpoise         ..F.V..... F........V M......... .......... .......... .......... -...I..... ..........
Horse            ....P.V... FH.......V .......... .......... .......... .......... P.L..I.... .........I
Tapir            .HF.P..... F........V .......... .......... .......... .......... P....I.... .........I
Rhinoceros       ....P.V... F........V .......... .......... .......... .......... P....IH... .........I
Wolverine        ******.... F........V ...A...... .......... .......... .......... P.M..I.... ..........
River_otter      ....L..... F........V ...A...... .......... .......... .......... P....I.... ..........
Dog              ....L..... F........V .......... .......... .......... .......... P....I.... ..........
Arctic_fox       *****..... F........V .......... .......... .......... .......... P.M..I...L .E........
Gray_seal        ....L..... F........V .......... .......... .......... .......... P....I.... ..........
Weddell_seal     ....L..... F........V .......... .......... .......... .......... P--------- ..........
Canada_lynx      *****..... F........V .......... .......... .......... .......... S.M..I...L ..........
Tiger            AT..L..... F........I .......... .......... .......... .......... S.M..I...L ..........
Brown_bear       ....L..... F........V ..PIQPIQPI Q......... .......... .......... P.M..I.... ..........
Panda            ....L..... F........H. .......... .......... .......... .......... -- -------- ..........
Flying_fox       ...LP..... F.....HV ..PIQPIQPI QPIQPIQPIQ PIQPIQPMQP MQPMQPIQ.. P....I.... .......L.
Fruit_bat        ...LP..... F....----- ..PQQPVHPI QPQSPVHSMQ .......... .......... S....I.... .......L.
Roundleaf_bat    T..LP..... F........V .......... .......... .......... .......... -- -------- ..........
Hedgehog         S...A..... .........V M.PMAPMQPM Q......... .......... .......... S.M.---... ......H.I.
Shrew            ..V.P..... F........A .......... .......... .......... .......... P.M..I.... Q.........
Armadillo        ..V.P.V... F........A .......... .......... .......... .......... P.M..I.--- --......V.
Elephant         ....P.I... .........V ......Q.. M......... .......... .......... P....I.... ..........
Manatee          ....P.I... .........V .......... .......... .......... .......Q. P....I.... ..........
Tenrec           .H..P.V... .........V MP........ .......... .......... .......... P......... ..........
Golden_mole      AH..P-V... F.PIQ.H..V ..Q....... MH........ .......... .......... P.M....... .......S...
Elephant_shrew   .H..P.V... ........SV M.PMQPMHPM H......... .......... .......... QSM....... ..........
Hyrax            ....P.I... .........V L......... .......... .......... .......... P.M..I.... .........I
Opossum          ......G... .........A .Q........ ..PIQPIQPI QPIQPMQPMQ PMQPMQPMQ. T...AVL... ..........
Aquatic_opossum  ......GH.. .........A .Q..PQ.... ..PIQPIQPI QPMQPMQPMQ PLQPMQPMQP MQ........ T...AVL... ..........
Platypus         S......... F.......V. ..QP...... .......... .......... .......... ..A...P.M- ..........
Echidna          S.....G... F......F. ..KP..T.R. .......... .......... .......... ......P.M- ...Q......  b
```

Fig. 11. Alignment of AMEL amino acid sequences in representative mammals. **a** Well-conserved N- and C-terminal regions in 20 species. Exon 4 is not represented because it is lacking in several species. Partial sequences were removed from this alignment. Vertical bars indicate the limits between exons. Unchanged residues are shown on a gray background. **b** Central region of exon 6 in 51 species emphasizing the region considered a hot spot of mutation. This region is characterized by amino acid triplet insertions or deletions. Identical sequences were not included in these alignments: e.g. human = chimpanzee and rhesus monkey; mouse = rat; cow = European bison. · · · · · = Identical residue; – – – = indel; * = unknown residue.

Fig. 12. Amino acid sequence of human amelogenin highlighting the residues which remained unchanged during the 225 million years of mammalian diversification. The importance of amino acids is inferred from the alignment of 60 mammalian sequences representative of the main lineages, as partially shown in figure 11. Exon 4 (14 residues) was not included because it is missing in several species studied. Signal peptide is on gray background. The protein sequence (191 amino acids) is numbered from methionine (1). Bold characters (n = 75) indicate residues unchanged in mammals, italics (n = 35) residues that can be substituted by an amino a{tb}cid from the same group only, small roman characters residues that can be substituted, characters on gray background (n = 5) residues that are known so far to lead to amelogenesis imperfecta when substituted, and underlined characters indicate (n = 31) residues that are unchanged in amniotes (mammals and reptiles) [Delgado et al., in press].

Ex 2	Ex 3	Ex 5	Ex 6	Ex 6	Ex 6	Ex 6	Ex 6	Ex 6	Ex 6	Ex 6	Ex 6	Ex 6	Ex 6
M (1)	L (20)	V	Y	I	H	M (100)	P	P	P	P (170)	P	T
G	P	L	P	P	H	M	N	Q	L	P	P	K
T	P	T	S (50)	V	I	P	L	P	P	Q	M	R
W	H	P	Y	L	P	V	P (120)	H	P	P	L	E
I	P	L	G	S	V	P	P	Q	Q	P	P	E
L	G	K (40)	Y	Q (70)	V	G	P	P	P	L	D (190)	V (190)
F	H	W	E	Q	P	Q	A	M (140)	P	T	L	
A	P	Y	P	H	A	A	Q	Q	L (160)	L	**Ex 7**
C	G	Q	M	P	Q	S	Q	P	P	E	D
L (10)	Y	S	G	P	H	M	P	Q	P	A	
L	I	I	G	T	S	T (110)	Y	P	M	W	
G	N (30)	R	W	H	M	P	Q	P	F	P	
A	F	P	L (60)	T	T	P	P	P	P (180)	S	
A	S	P	H	L	P	I	P	V	M	T	
F	Y		H (80)	Q	I	Q	P	H	Q	D	
A	E		Q	P	Q	Q	P	P	P	K	
M			I	H	P	H	V	M	L		
P					Q	Q	Q (150)	Q			
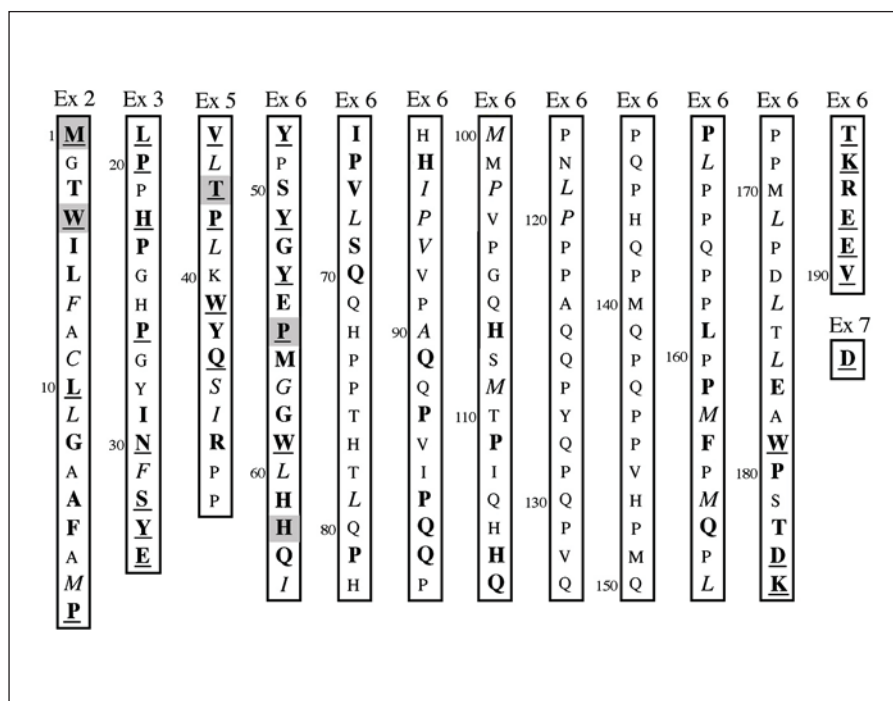
(iii) Although this central region of AMEL exon 6 is variable, it maintains its richness in proline (30%) and glutamine (20%) in all sequences studied. This means that this region is also subject to a functional constraint but that this selective pressure probably acts on the general conservation of the P and Q richness rather than on specific amino acid positions. This strongly suggests that this region could be subject to polymorphism in humans.

(iv) The origin of the largest of AMEL exon 6 has to be found in the repeats of nine nucleotides coding for three residues (triplets) PXQ or PXX [Delgado et al., 2005]. These repeats have not been blurred by substitutions during at least 310 million years of amniote evolution, because such triplet repeats have been identified in crocodile AMEL [Sire et al., 2006]. The triplet insertions found in the hot spot mutation in mammals are probably reminiscent of this mechanism. These repeats are to be found, probably, in the origin of AMEL after AMBN duplication, and also constitute the originality of AMEL compared to the other EMPs and to ameloblast-secreted SCPPs in general. This leads to the hypothesis that AMEL divergence consisted of the loss of most of the C-terminal region of the AMBN ancestor and of the development of exon 6 (probably from AMBN exon 5) through several runs of PXQ triplet repeats. This new protein was posi-

tively selected during enamel evolution in vertebrates because this hydrophobic region, rich in P and Q, improved the resistance of enamel to wear and microbreaks. This could explain why today AMEL represents 90% of the forming enamel matrix in mammals.

Validation of Mutations and Important Residues
The evolutionary analysis of AMEL in mammals reveals >70 residues (out of 191) that are certainly important for a correct function of AMEL because they have remained unchanged during 225 million years of evolution (fig. 12). The number of conserved residues is reduced to 34 when reptilian AMELs are added to this analysis [Delgado et al., in press]. These 34 positions conserved during 310 million years of amniote evolution are considered crucial residues for enamel formation. All of them are located in the N- and C-terminal regions of AMEL, known to play an important role in relation with the environment (interactions with the ameloblast surface and/or with the mineral crystals). The residues conserved only in mammals could indicate that they play new, important roles for enamel formation in this lineage.

As a consequence of their long-lasting conservation, substitution of the important amino acids revealed in this study could result in enamel defects (AIH1) when substi-

tuted in humans (fig. 12). The five substitutions leading to AIH1 are validated when using the mammalian, and four of them when using the amniote dataset. Therefore, this list of conserved residues in the human AMEL sequence (fig. 12) can be useful for the clinical diagnosis of AIH1 since it helps to validate any human AMEL mutation, which could be suspected for AIH1.

Conclusion

Although the origin of enamel can be traced back to early vertebrates, at least 500 MYA in the fossil record, our knowledge of enamel mineralization genes is still restricted to the tetrapod level (350 MYA) for AMEL and AMBN, and to the mammalian level (225 MYA) for ENAM. The difficulty encountered when looking for EMP genes in the vertebrate lineages that diverged earlier in evolution (i.e. chondrichthyans, 430 MYA, and actinopterygians, 420 MYA) resides in their high sequence variations (intrinsically disordered proteins) and in the lack of sequenced genomes in basal lineages such as lungfish, polypterids and sharks, which do not allow looking for EMP genes using syntheny. Our approach using putative ancestral sequences could help to obtain data in closely related but not in evolutionary distant lineages. Molecular dating of AMBN/AMEL duplication indicates that EMP genes probably appeared at the end of the Precambrian era (>600 MYA) after several rounds of genome/gene duplications that took place in this period. ENAM was created first, then AMBN and AMEL. After AMBN duplication, one copy lost a large part of the ancestral 3′ region and accumulated PXQ repeats. These events gave rise to a new protein: AMEL. AMEL was then positively selected (and constrained), probably because it improved enamel microstructure and thickness: it is now the major protein forming enamel in amniotes. The AMEL story is relatively well established now, but some details will be undoubtedly added when the evolutionary analyses in amphibians and reptiles will be achieved. Such a study will probably open the door to access the AMEL sequence in lungfish, the sister group to tetrapods. In contrast to our knowledge on AMEL, the other ameloblast-secreted SCPP proteins (AMBN, ENAM and the newly identified AMTN and ODAM) are poorly known. Efforts have to be made towards better knowledge of the relationships and evolution of these proteins, and the current genome sequencing programs will certainly be of great value in this quest. It is clear that evolutionary analyses are necessary not only for thorough knowledge of each protein (i.e. its origin, relationships, and mode of evolution) but also because they provide insights into residues that play important roles for the correct function of the protein. In addition, as illustrated with AMEL, sequence datasets obtained in a phylogenetic perspective will be helpful to validate mutations responsible for genetic diseases in humans.

References

Aoba, T. (1996) Recent observations on enamel crystal formation during mammalian amelogenesis. Anat Rec 245: 208–218.

Aoba, T., E.C. Moreno, M. Kresak, T. Tanabe (1989) Possible roles of partial sequences at N- and C-termini of amelogenin in protein-enamel mineral interaction. J Dent Res 68: 1331–1336.

Aung, P.P., N. Oue, Y. Mitani, H. Nakayama, K. Yoshida, T. Noguchi, A.K. Bosserhoff, W. Yasui (2006) Systematic search for gastric cancer-specific genes based on SAGE data: melanoma inhibitory activity and matrix metalloproteinase-10 are novel prognostic factors in patients with gastric cancer. Oncogene 25: 2546–2557.

Bartlett, J.D., O.H. Ryu, J. Xue, J.P. Simmer, H.C. Margolis (1998) Enamelysin mRNA displays a developmentally defined pattern of expression and encodes a protein which degrades amelogenin. Connect Tissue Res 39: 101–109.

Bartlett, J.D., E. Beniash, D.H. Lee, C.E. Smith (2004) Decreased mineral content in MMP-20 null mouse enamel is prominent during the maturation stage. J Dent Res 83: 909–913.

Bartlett, J.D., R.L. Ball, T. Kawai, C.E. Tye, M. Tsuchiya, J.P. Simmer (2006a) Origin, splicing, and expression of rodent amelogenin exon 8. J Dent Res 85: 894–899.

Bartlett, J.D., B. Ganss, M. Goldberg, J. Moradian-Oldak, M.L. Paine, M.L. Snead, X. Wen, S.N. White, Y.L. Zhou (2006b) Protein-protein interactions of the developing enamel matrix. Curr Top Dev Biol 74: 57–115.

Carter, J.G. (1990) Skeletal Biomineralization: Patterns, Processes and Evolutionary Trends. New York, Van Nostrand Reinhold, p 832.

Chung, W.Y., R. Albert, I. Albert, A. Nekrutenko, K.D. Makova (2006) Rapid and asymmetric divergence of duplicate genes in the human gene coexpression network. BMC Bioinformatics 7: 46.

Davit-Béal, T., F. Allizard, J.-Y. Sire (2007) Enameloid/enamel transition through successive tooth replacements in *Pleurodeles waltl* (Lissamphibia, Caudata). Cell Tissue Res 328: 167–183.

Dehal, P., J.L. Boore (2005) Two rounds of genome duplication in the ancestral vertebrate. PLoS Biol 3: e314.

Delgado, S., D. Casane, L. Bonnaud, M. Laurin, J.-Y. Sire, M. Girondot (2001) Molecular evidence for Precambrian origin of amelogenin, the major protein of vertebrate enamel. Mol Biol Evol 18: 2146–2153.

Delgado, S., M. Girondot, J.-Y. Sire (2005) Molecular evolution of amelogenin in mammals. J Mol Evol 60: 12–30.

Delgado, S., M.L. Couble, H. Magloire, J.-Y. Sire (2006) Cloning, sequencing and expression of the amelogenin gene in two scincid lizards. J Dent Res 85: 138–143.

Delgado, S., M. Ishiyama, J.-Y. Sire (in press) Validation tools for AIH1 inferred from amelogenin evolution. J Dent Res.

Delsuc, F., M. Scally, O. Madsen, M.J. Stanhope, W.W. de Jong (2002) Molecular phylogeny of living xenarthrans and the impact of character and taxon sampling on the placental tree rooting. Mol Biol Evol 19: 1656–1671.

Diekwisch, T., S. David, P. Bringas, Jr., V. Santos, H.C. Slavkin (1993) Antisense inhibition of AMEL translation demonstrates supramolecular controls for enamel HAP crystal growth during embryonic mouse molar development. Development 117: 471–482.

Donoghue, P.C.J. (1998) Growth and patterning in the conodont skeleton. Philos Trans R Soc Lond Ser B 353: 633–666.

Donoghue, P.C.J. (2001) Microstructural variation in conodont enamel is a functional adaptation. Proc R Soc Lond Ser B 268: 1691–1698.

Donoghue, P.C.J. (2002) Evolution and development of the vertebrate dermal and oral skeletons: unraveling concepts, regulatory theories, and homologies. Paleobiology 28: 474–507.

Donoghue, P.C.J., I.J. Sansom (2002) Origin and early evolution of vertebrate skeletonization. Microsc Res Techn 59: 352–372.

Donoghue, P.C.J., I.V. Sansom, J.P. Downs (2006) Early evolution of vertebrate skeletal tissues and cellular interactions, and the canalization of skeletal development. J Exp Zoolog B Mol Dev Evol 306: 278–294.

Du, C., G. Falini, S. Fermani, C. Abbott, J. Moradian-Oldak (2005) Supramolecular assembly of amelogenin nanospheres into birefringent microribbons. Science 307: 1450–1454.

Dunker, A.K., J.D. Lawson, C.J. Brown, R.M. Williams, P. Romero, J.S. Oh, C.J. Oldfield, A.M. Campen, C.M. Ratliff, K.W. Hipps, J. Ausio, M.S. Nissen, R. Reeves, C.H. Kang, C.R. Kissinger, R.W. Bailey, M.D. Griswold, W. Chiu, E.C. Garner, Z. Obradovic (2001) Intrinsically disordered protein. J Mol Graph Model 19: 26–59.

Fincham, A.G., Y. Hu, E.C. Lau, H.C. Slavkin, M.L. Snead (1991) Amelogenin post-secretory processing during biomineralization in the postnatal mouse molar tooth. Arch Oral Biol 36: 305–317.

Fincham, A.G., J. Moradian-Oldak (1993) Amelogenin post-translational modifications: carboxy-terminal processing and the phosphorylation of bovine and porcine 'TRAP' and 'LRAP' amelogenins. Biochem Biophys Res Commun 197: 248–255.

Fincham, A.G., J. Moradian-Oldak (1995) Recent advances in amelogenin biochemistry. Connect Tissue Res 32: 119–124.

Fisher, L.W., N.S. Fedarko (2003) Six genes expressed in bones and teeth encode the current members of the SIBLING family of proteins. Connect Tissue Res 44(suppl 1): 33–40.

Ginger, M.R., C.P. Piotte, D.E. Otter, M.R. Grigor (1999) Identification, characterisation and cDNA cloning of two caseins from the common brushtail possum (*Trichosurus vulpecula*). Biochim Biophys Acta 1427: 92–104.

Girondot, M., J.-Y. Sire (1998) Evolution of the amelogenin gene in toothed and tooth-less vertebrates. Eur J Oral Sci 106: 501–508.

Graham, A. (2004) Rise of the little squirts. Curr Biol 14: R956–R958.

Greene, S.R., Z.A. Yuan, J.T. Wright, H. Amjad, W.R. Abram, J.A. Buchanan, D.I. Trachtenberg, C.W. Gibson (2002) A new frameshift mutation encoding a truncated amelogenin leads to X-linked amelogenesis imperfecta. Arch Oral Biol 47: 211–217.

Gu, X., Y. Wang, J. Gu (2002) Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. Nat Genet 31: 205–209.

Hart, P.S., M. Aldred, P. Crawford, N. Wright, T. Hart, J.T. Wright (2002a) Amelogenesis imperfecta phenotype-genotype correlations with two amelogenin gene mutations. Arch Oral Biol 47: 261–265.

Hart, P.S., T.C. Hart, J.P. Simmer, J.T. Wright (2002b) A nomenclature for X-linked amelogenesis imperfecta. Arch Oral Biol 47: 255–260.

Hart, P.S., M.D. Michalec, W.K. Seow, T.C. Hart, J.T. Wright (2003) Identification of the enamelin (g.8344delG) mutation in a new kindred and presentation of a standardized ENAM nomenclature. Arch Oral Biol 48: 589–596.

Herold, R.C., J. Rosenbloom, M. Granovsky (1989) Phylogenetic distribution of enamel proteins: immunolocalization with monoclonal antibodies indicates the evolutionary appearance of enamelins prior to amelogenins. Calcif Tissue Int 45: 88–94.

Hedges, S.B. (2002). The origin and evolution of model organisms. Nat Rev Genet 3: 838–849.

Hoang, A.M., R.J. Klebe, B. Steffensen, O.H. Ryu, J.P. Simmer, D.L. Cochran (2002) Amelogenin is a cell adhesion protein. J Dent Res 81: 497–500.

Hu, J.C., Y. Yamakoshi (2003) Enamelin and autosomal-dominant amelogenesis imperfecta. Crit Rev Oral Biol Med 14: 387–398.

Hu, J.C., X. Sun, C. Zhang, S. Liu, J.D.. Bartlett, J.P. Simmer (2002) Enamelysin and kallikrein-4 mRNA expression in developing mouse molars. Eur J Oral Sci 110: 307–315.

Huysseune, A., J.-Y. Sire (1998) Evolution of patterns and processes in teeth and tooth-related tissues in non-mammalian vertebrates. Eur J Oral Sci 106(suppl 1): 437–481.

Ishiyama, M., M. Mikami, H. Shimokawa, S. Oida (1998) Amelogenin protein in tooth germs of the snake *Elaphe quadrivirgata*, immunohistochemistry, cloning and cDNA sequence. Arch Histol Cytol 61: 467–474.

Iwasaki, K., E. Bajenova, E. Somogyi-Ganss, M. Miller, V. Nguyen, H. Nourkeyhani, Y. Gao, M. Wendel, B. Ganss (2005) Amelotin – a novel secreted, ameloblast-specific protein. J Dent Res 84: 1127–1132.

Iwase, M., Y. Satta, N. Takahata (2001) Sex-chromosomal differentiation and amelogenin genes in mammals. Mol Biol Evol 18: 1601–1603.

Iwase, M., Y. Satta, Y. Hirai, H. Hirai, H. Imai, N. Takahata (2003) The amelogenin loci span an ancient pseudoautosomal boundary in diverse mammalian species. Proc Natl Acad Sci USA 100: 5258–5263.

Janvier, P. (1996) Early Vertebrates. Oxford Monographs on Geology and Geophysics. New York, Oxford University Press, vol 33, p 393.

Kawasaki, K., K.M. Weiss (2003) Mineralized tissue and vertebrate evolution: the secretory calcium-binding phosphoprotein gene cluster. Proc Natl Acad Sci USA 100: 4060–4065.

Kawasaki, K., K.M. Weiss (2006) Evolutionary genetics of vertebrate tissue mineralization: the origin and evolution of the secretory calcium-binding phosphoprotein family. J Exp Zoolog B Mol Dev Evol 306: 295–316.

Kawasaki, K., T. Suzuki, K.M. Weiss (2004) Genetic basis for the evolution of vertebrate mineralized tissue. Proc Natl Acad Sci USA 101: 11356–11361.

Kawasaki, K., T. Suzuki, K.M. Weiss (2005) Phenogenetic drift in evolution: the changing genetic basis of vertebrate teeth. Proc Natl Acad Sci USA 102: 18063–18068.

Kim, J.-W., J.P. Simmer, Y.Y. Hu, B.P.-L. Lin, C. Boyd, J.T. Wright, C.J.M. Yamada, S.K. Rayes, R.J. Feigal, J.C.-C. Hu (2004) Amelogenin p.M1T and p.W4S mutations underlying hypoplastic X-linked amelogenesis imperfecta. J Dent Res *83:* 378–383.

Kim, J.-W., F. Seymen, B.P.-L. Lin, B. Kiziltan, K. Gencay, J.P. Simmer, J.C.-C. Hu (2005) ENAM mutations in autosomal-dominant amelogenesis imperfecta. J Dent Res *84:* 278–282.

Kumar, S., S.B. Hedges (1998) A molecular timescale for vertebrate evolution. Nature *392:* 917–920.

Kumar, S., A. Filipski, V. Swarna, A. Walker, S.B. Hedges (2005) Placing confidence limits on the molecular age of the human-chimpanzee divergence. Proc Natl Acad Sci USA *102:* 18842–18847.

Lahn B.T., N.M. Pearson, K. Jegalian (2001) The human Y chromosome, in the light of evolution. Nat Rev Genet *2:* 207–216.

Lyngstadaas, S.P., S. Risnes, H. Nordbo, A.G. Flones (1990) Amelogenin gene similarity in vertebrates: DNA sequences encoding amelogenin seem to be conserved during evolution. J Comp Physiol *160:* 469–472.

Madsen, O., M. Scally, C.J. Douady, D.J. Kao, R.W. DeBry, R. Adkins, H.M. Amrine, M.J. Stanhope, W.W. de Jong, M.S. Springer (2001) Parallel adaptive radiations in two major clades of placental mammals. Nature *409:* 610–614.

Mardh, C.K., B. Backman, D. Simmons, I. Golovleva, T.T. Gu, G. Holmgren, M. MacDougall, K. Forsman-Semb (2001) Human ameloblastin gene: genomic organization and mutation analysis in amelogenesis imperfecta patients. Eur J Oral Sci *109:* 8–13.

Margolis, H.C., E. Beniash, C.E. Fowler (2006) Role of macromolecular assembly of enamel matrix proteins in enamel formation. J Dent Res *85:* 775–793.

Moffatt, P., C.E. Smith, R. Sooknanan, R. St-Arnaud, A. Nanci (2006a) Identification of secreted and membrane proteins in the rat incisor enamel organ using a signal-trap screening approach. Eur J Oral Sci *114(suppl 1):* 139–146.

Moffatt, P., C.E. Smith, R. St-Arnaud, D. Simmons, T. Wright, A. Nanci (2006b) Cloning of rat amelotin and localization of the protein to the basal lamina of maturation stage ameloblasts and junctional epithelium. Biochem J *399:* 37–46.

Murphy, W.J., E. Elzirik, W.E. Johnson, Y.P. Zhang, O.A. Ryder, S.J. O'Brien (2001) Molecular phylogenetics and the origin of placental mammals. Nature *409:* 614–618.

Ørvig, T. (1967) Phylogeny of tooth tissues: evolution of some calcified tissues in early vertebrates; in Miles, A.E.W. (ed): Structural and Chemical Organization of Teeth. New York, Academic Press, vol 1, pp 45–105.

Ørvig, T. (1977) A survey of odontodes ('dermal teeth') from developmental, structural, functional, and phyletic points of view; in Andrews, S.M., R.S. Miles, A.D. Walker (eds): Problems in Vertebrate Evolution. New York, Academic Press, pp 53–75.

Paine, M.L., W. Luo, D.-H. Zhu, P. Bringas, Jr., M.L. Snead (2003a) Functional domains for amelogenin revealed by compound genetic defects. J Bone Miner Res *18:* 466–472.

Paine, M.L., H.J. Wang, M.L. Snead (2003b) Amelogenin self-assembly and the role of the proline located within the carboxyl-teleopeptide. Connect Tissue Res *44:* 52–57.

Panopoulou, G., A.J. Pouska (2005) Timing and mechanism of ancient vertebrate genome duplications – the adventure of an hypothesis. Trends Genet *21:* 559–567.

Prostak, K., Z. Skobe (1984) Effects of colchicines on fish enameloid matrix formation; in Fearnhead, R.W., S. Suga (eds): Tooth Enamel IV. Amsterdam, Elsevier, pp 525–529.

Prostak, K., Z. Skobe (1988) Ultrastructure of odontogenic cells during enameloid matrix synthesis in tooth buds from an elasmobranch, *Raja erinacea.* Am J Anat *182:* 59–72.

Prostak, K., P. Sieffert, Z. Skobe (1993) Enameloid formation in two tetraodontiform fish species with high and low fluoride contents in enameloid. Arch Oral Biol *38:* 1031–1044.

Qin, C., O. Baba, W.T. Butler (2004) Post-translational modifications of SIBLING proteins and their roles in osteogenesis and dentinogenesis. Crit Rev Oral Biol Med *15:* 126–136.

Ravindranath, R.M, J. Moradian-Oldak, A.G. Fincham (1999) Tyrosyl motif in amelogenins binds N-acetyl-D-glucosamine. J Biol Chem *274:* 2464–2471.

Ravindranath, R.M.H., W. Tam, P. Nguyen, A.G. Fincham (2000) The enamel protein amelogenin binds to the N-acetyl-D-glucosamine-mimicking peptide motif of cytokeratins. J Biol Chem *275:* 39654–39661.

Ravindranath, R.M.H., W. Tam, P. Bringas, V. Santos, A.G. Fincham (2001) Amelogenin-cytokeratin 14 interaction in ameloblasts during enamel growth. J Biol Chem *276:* 36586–36597.

Ravindranath, R.M.H, R.M. Basilrose, N.H. Ravindranath, B. Vaitheesvaran (2003) Amelogenin interacts with cytokeratin-5 in ameloblasts during enamel growth. J Biol Chem *278:* 20293–20302.

Reif, W.E. (1982) Evolution of dermal skeleton and dentition in vertebrates. Evol Biol *15:* 287–368.

Rowe, P.S. (2004) The wrickkened pathways of FGF23, MEPE and PHEX. Crit Rev Oral Biol Med *15:* 264–281.

Salido, E., P. Yen, K. Koprivnikar, L.C. Yu, L. Shapiro (1992) The human enamel protein gene amelogenin is expressed from both the X- and Y-chromosomes. Am J Hum Genet *50:* 303–316.

Sansom, I.V., M.P. Smith, H.A. Armstrong, M.M. Smith (1992) Presence of earliest vertebrate hard tissues in conodonts. Science *256:* 1308–1311.

Sansom, I.J., M.P. Smith, M.M. Smith (1994) Dentine in conodonts. Nature *368:* 391.

Sansom, I.J., P.C.J. Donoghue, G.L. Albanesi (2005) Histology and affinity of the earliest armoured vertebrate. Biol Lett *2:* 446–449.

Sasagawa, I. (1984) Formation of cap enameloid in the jaw teeth of dog salmon, *Oncorhynchus keta.* Jpn J Oral Biol *26:* 477–495.

Sasagawa, I. (1995) Fine structure of the tooth germs during formation of the enameloid matrix in *Tilapia nilotica,* a teleost fish. Arch Oral Biol *40:* 801–814.

Sasagawa, I. (2002a) Fine structural and cytochemical observations of dental epithelial cells during the enameloid formation stages in red stingrays *Dasyatis akajei.* J Morphol *252:* 170–182.

Sasagawa, I. (2002b) Mineralization patterns in elasmobranch fish. Microsc Res Techn *59:* 396–407.

Satchell, P.G., C.F. Shuler, T.G.H. Diekwisch (2000) True enamel covering in teeth of the Australian lungfish *Neoceratodus forsteri.* Cell Tissue Res *299:* 27–37.

Sato, I., M. Kobayashi, R. Ueno, T. Sato (1986) The ultrastructure of the teeth in the Amphibia: differentiation of the enamel. Shigaku *73:* 1815–1820.

Shimeld, S.M., P.W.H. Holland (2000) Vertebrate innovations. Proc Natl Acad Sci USA *97:* 4449–4452.

Shintani, S., M. Kobata, S. Toyosawa, T. Fujiwara, A. Sato, T. Ooshima (2002) Identification and characterization of ameloblastin gene in a reptile. Gene *283:* 245–254.

Shintani, S., M. Kobata, S. Toyosawa, T. Ooshima (2003) Identification and characterization of ameloblastin gene in an amphibian, *Xenopus laevis.* Gene *318:* 125–136.

Shu, D.-G., H.-L. Luo, S. Conway Morris, X.-L. Zhang, S.-X. Hu, L. Chen, J. Han, M. Zhu, Y. Li, L.-Z. Chen (1999) Lower Cambrian vertebrates from South China. Nature *402:* 42–46.

Shu, D.-G., S.C. Morris, J. Han, Z.-F. Zhang, K. Yasui, P. Janvier, L. Chen, X.-L. Zhang, J.-N. Liu, Y. Li, H.-Q. Liu (2003) Head and backbone of the Early Cambrian vertebrate *Haikouichthys.* Nature *421:* 526–529.

Simmer, J.P., M. Fukae, T. Tanabe, Y. Yamakoshi, T. Uchida, J. Xue, H.C. Margolis, M. Shimizu, B.C. DeHart, C.-C. Hu, J.D. Bartlett (1998) Purification, characterization and cloning of enamel matrix serine proteinase 1. J Dent Res *77:* 377–386.

Simmer, J.P., J.C. Hu (2002) Expression, structure, and function of enamel proteinases. Connect Tissue Res *43:* 441–449.

Sire, J.-Y. (1990) From ganoid to elasmoid scales in the actinopterygian fishes. Neth J Zool *40:* 75–92.

Sire, J.-Y. (1994) A light and TEM study of non-regenerated and experimentally regenerated scales of *Lepisosteus oculatus* (Holostei) with particular attention to ganoine formation. Anat Rec *240:* 189–207.

Sire, J.-Y. (1995) Ganoine formation in the scales of primitive actinopterygian fishes, lepisosteids and polypterids. Connect Tissue Res *33:* 213–222.

Sire, J.-Y., A. Huysseune (2003) Formation of skeletal and dental tissues in fish: a comparative and evolutionary approach. Biol Rev *78:* 219–249.

Sire, J.-Y., S. Delgado, D. Fromentin, M. Girondot (2005) Amelogenin: lessons from evolution. Arch Oral Biol *50:* 205–212.

Sire, J.-Y., S. Delgado, M. Girondot (2006) The amelogenin story: origin and evolution. Eur J Oral Sci *114:* 64–77.

Sire, J.-Y., J. Géraudie, F.J. Meunier, L. Zylberberg (1987) On the origin of ganoine: histological and ultrastructural data on the experimental regeneration of the scales of *Calamoichthys calabaricus* (Osteichthyes, Brachiopterygii, Polypteridae). Am J Anat *180:* 391–402.

Slavkin, H.C., N. Samuel, P. Bringas, Jr., A. Nanci (1983) Selachian tooth development: II. Immunolocalization of amelogenin polypeptides in epithelium during secretory amelogenesis in *Squalus acanthias.* J Craniofac Genet Dev Biol *3:* 43–52.

Smith, M.M. (1995) Heterochrony in the evolution of enamel in vertebrates; in McNamara, K.J. (ed): Evolutionary Change and Heterochrony. New York, Wiley, pp 125–150.

Smith, M.M., B.K. Hall (1990) Development and evolutionary origins of vertebrate skeletogenic and odontogenic tissues. Biol Rev *65:* 277–373.

Snead, M. (2003). Amelogenin protein exhibits a modular design: implications for form and function. Connect Tissue Res *44(suppl 1):* 47–51.

Solomon, A., C.L. Murphy, K. Weaver, D.T. Weiss, R. Hrncic, M. Eulitz, R.L. Donnell, K. Sletten, G. Westermark, P. Westermark (2003) Calcifying epithelial odontogenic (Pindborg) tumor-associated amyloid consists of a novel human protein. J Lab Clin Med *142:* 348–355.

Stasiuk, S.X., E.L. Summers, J. Demmer (2000) Cloning of a marsupial kappa-casein cDNA from the brushtail possum (*Trichosurus vulpecula*). Reprod Fertil Dev *12:* 215–222.

Steinke, D., W. Salzburger, I. Braasch, A. Meyer (2006) Many genes in fish have species-specific asymmetry rates of molecular evolution. BMC Genom *8:* 7–20.

Stephanopoulos, G., M.E. Garefalaki, K. Lyroudia (2005) Genes and related proteins involved in amelogenesis imperfecta. J Dent Res *84:* 1117–1126.

Tompkins, K., A. George, A. Veis (2006) Characterization of a mouse amelogenin [A-4]'M59 cell surface receptor. Bone *38:* 172–180.

Toyosawa, S., C. O'hUigin, F. Figueroa, H. Tichy, J. Klein (1998) Identification and characterization of amelogenin genes in monotremes, reptiles, and amphibians. Proc Natl Acad Sci USA *95:* 13056–13061.

van Rheede, T., T. Bastiaans, D.N. Boone, S.B. Hedges, W.W. de Jong, O. Madsen (2006) The platypus is in its place: nuclear genes and indels confirm the sister group relation of monotremes and Therians. Mol Biol Evol *23:* 587–597.

Veis, A., K. Tompkins, K. Alvares, K. Wei, L. Wang, X.S. Wang, A.G. Brownell, S.-M. Jengh, K.E. Healy (2000) Specific amelogenin gene splicing products have signaling effects on cells in culture and in implants in vivo. J. Biol Chem *275:* 41263–41272.

Veis, A. (2003) Amelogenin gene splice products: potential signaling molecules. Cell Mol Life Sci *60:* 38–55.

Veis, A. (2005) A window on biomineralization. Science *307:* 1419–1420.

Wang, X., Y. Ito, X. Luan, A. Yamane, T.G.H. Diekwisch (2005a) Amelogenin sequence and enamel biomineralization in *Rana pipiens.* J Exp Zoolog B Mol Dev Evol *304:* 1–10.

Wang, H.J., S. Tannukit, D.H. Zhu, M.L. Snead, M.L. Paine (2005b) Enamel matrix protein interactions. J Bone Miner Res *20:* 1032–1040.

Wang, X., J.L. Fan, Y. Ito, X. Luan, T.G.H. Diekwisch (2006) Identification and characterization of a squamate reptilian amelogenin gene: *Iguana iguana.* J Exp Zool B Mol Dev Evol *306:* 393–406.

Weikard, R., C. Pitra, C. Kuhn (2006) Amelogenin cross-amplification in the family Bovidae and its application for sex determination. Mol Reprod Dev *73:* 1333–1337.

Wen, H.B., A.G. Fincham, J. Moradian-Oldak (2001) Progressive accretion of amelogenin molecules during nanospheres assembly revealed by atomic force microscopy. Matrix Biol *20:* 387–395.

Yamamoto, K., T. Tsubota, T. Komatsu, A. Katayama, T. Murase, I. Kita, T. Kudo (2002) Sex identification of Japanese black bear, *Ursus thibetanus japonicus*, by PCR based on amelogenin gene. J Vet Med Sci *64:* 505–508.

Zhang, X., J. Zhao, C. Li, S. Gao, C. Qiu, P. Liu, G. Wu, B. Qiang, W.H.Y. Lo, Y. Shen (2001) DSPP mutation in dentinogenesis imperfecta Shields type II. Nat Genet *27:* 151–152.

Zylberberg, L., J.-Y. Sire, A. Nanci (1997) Immunodetection of amelogenin-like proteins in the ganoine of experimentally regenerating scales of *Calamoichthys calabaricus*, a primitive actinopterygian fish. Anat Rec *249:* 86–95.