

Molecular Evidence for Precambrian Origin of Amelogenin, the Major Protein of Vertebrate Enamel

Sidney Delgado,* Didier Casane,† Laure Bonnaud,*‡ Michel Laurin,* Jean-Yves Sire,* and Marc Girondot*§

*UMR 8570, Evolution et Adaptations des Systèmes Ostéomusculaires, Paris, France; †UMR 7622, Biologie Moléculaire et Cellulaire du Développement, Equipe "Phylogénie, Bioinformatique et Génome," Université Pierre et Marie Curie, Paris, France; ‡Institut Jacques Monod, Equipe "Evolution du Développement des Nématodes," Université Denis Diderot, Paris, France; and §UPRESA 8079, Ecologie, Systématique et Evolution, Université Paris-Sud, Orsay, France

Although molecular dating of cladogenetic events is possible, no molecular method has been described to date the acquisition of various tissues. Taking into account the specificity of the major protein in enamel in formation (amelogenin), we were able to develop such a method for enamel. Indeed, because the amelogenin protein is exclusively involved in enamel formation and mineralization and because it lacks pleiotropic effects, this protein is a good candidate to estimate the date of acquisition of this highly mineralized tissue. We searched DNA banks for similarities between the amelogenin sequence and other sequences. Similarities were found only to exon 2 of SPARC (osteonectin) in two protostomians and in eight deuterostomians, and to exon 2 of three SPARC-related deuterostomian genes (SC1, hevin, and QR1). The other amelogenin exons did not reveal significant similarities to other sequences. In these proteins, exon 2 mainly encodes the peptide signal that plays the essential role in enabling the protein to be ultimately localized in the extracellular matrix. We tested the significance of the exon 2 similarities. The observed values were always significantly higher than the expected randomly generated similarities. This demonstrates a common evolutionary origin of this exon. The phylogenetic analyses of exon 2 sequences indicated that exon 2 was duplicated to amelogenin from an ancestral SPARC sequence in the deuterostomian lineage before the duplication of deuterostomian SPARC and SC1/hevin/QR1. We were able to date the origin of the latter duplication at approximately 630 MYA. Therefore, amelogenin exon 2 was acquired before this date, in the Proterozoic, long before the so-called "Cambrian explosion," the sudden appearance of several bilateralian phyla in the fossil record at the Proterozoic-Phanerozoic transition. This sudden appearance has been often suggested to reflect intensive cladogenesis during this period. However, molecular dating of protostomian-deuterostomian divergence and of the cladogenesis among several major clades of Bilateria lead to a different conclusion: many bilateralian clades were already present during the late Proterozoic. It has previously been proposed that these bilateralians were not mineralized and that they had low fossilization potential. Our results strongly suggest that late Proterozoic fossils possessing a mineralized tissue homologous to enamel might be found in the future.

Introduction

Amelogenin is quantitatively the major protein in the forming mammalian enamel, in which it represents approximately 90% of the organic content (Termine et al. 1980; Fincham et al. 1982; Sasaki and Shimokawa 1995). Several other gene products have also been found in the enamel matrix, various proteases and anionic proteins (known as nonamelogenins). Some of these proteins have been characterized in mammals, e.g., tuftelin (Deutsch et al. 1995) and ameloblastin (also named amelin or sheathlin; Fong, Slaby, and Hammarström 1996; Krebsbach et al. 1996). Once mineralization starts, the enamel proteins are progressively degraded by proteases and are removed from the matrix; this process continues until complete maturation of the enamel, and it results in a highly mineralized tissue (Deutsch 1989). The mineral matrix of enamel is composed of hydroxyapatite crystals oriented perpendicularly to the ameloblast surface. This particular arrangement, although slightly variable (e.g., either prismatic or not) in the enamel and enameloid tissues of the various verte-

brate lineages, is highly characteristic and allows one to distinguish these tissues from other mineralized tissues. Before it is degraded, amelogenin is thought to play an important role in this organization of the mineral crystals by preventing random proliferation of crystal nuclei and by regulating the growth kinetics, orientation, and size of the enamel crystals (Aoba 1996; Lyngstadaas et al. 1996).

All of the above features are specific to vertebrate enamel formation and are not encountered in other mineralized tissues (even in nonvertebrate lineages). Another striking character of most of these enamel proteins is that they have not been found expressed in other mineralizable or nonmineralizable tissues: these proteins are specific to enamel, or, in other words, they have no pleiotropic effects.

We have been interested in the evolutionary origin of the tissues composing the dermal skeleton, and the specificity of the enamel proteins has drawn our attention. Because the enamel nonamelogenin genes are known only from a few mammalian species, they cannot be used to study long-term evolution. In contrast, amelogenin sequences are available in several vertebrate lineages (Girondot and Sire 1998; Ishiyama et al. 1998; Toyosawa et al. 1998).

The present study was undertaken to date the appearance of the amelogenin gene in the vertebrate genome. Amelogenin is currently believed to be an orphan

Key words: amelogenin, enamel, vertebrates, Cambrian explosion.

Address for correspondence and reprints: Marc Girondot, UPRESA 8079, Ecologie, Systématique et Evolution, Université Paris-Sud, Orsay, Bâtiment 362, 91405 Orsay Cedex, France. E-mail: marc.girondot@epc.u-psud.fr.

Mol. Biol. Evol. 18(12):2146–2153. 2001

© 2001 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

gene without any known homologs. If an amelogenin paralog gene were found, knowledge of the date of duplication of the ancestral gene, or of part of this gene, and its standard deviation would provide a range of possible ages for the acquisition of this protein and of the enamel-like tissue in which it was expressed. Because the amelogenin gene encodes a protein that plays an important role in enamel mineralization and shows no known pleiotropic effect, it is an excellent candidate to independently estimate the date of acquisition of a mineralized tissue in vertebrates. Indeed, this date is currently debated.

On the one hand, according to the fossil record, the earliest vertebrates are known from the early Cambrian of China (during the so-called "Cambrian explosion"), but they did not possess mineralized tissues (Chen, Huang, and Li 1999; Shu et al. 1999). Mineralization probably appeared later in vertebrates. Indeed, the earliest presumed vertebrate known to possess a mineralized skeleton was *Anatolepis*, found in the Upper Cambrian (520 MYA) (Repetski 1978). Its dermal skeleton was ornamented by small tubercles composed of a tissue identified as a kind of dentin. A mineralized tissue of unknown homology covered it (Smith, Sansom, and Repetski 1996). Enamel and other tissues of uncertain homologies, possibly bone, dentine, and calcified cartilage, were also described in euconodonts from the Upper Cambrian (Sansom et al. 1992, but these interpretations have been disputed; see Schultze 1996). These animals of controversial affinities are now considered craniates by many authors (Aldridge and Donoghue 1998; Janvier 1996b). Enamel, or enameloid tissue covering tubercles, was clearly described in Ordovician vertebrates (450 MYA) such as the ostracoderms (Ørvig 1989). On the other hand, according to molecular data, the Bilateria were estimated to have had their origin during the Proterozoic, between 1,000 and 830 MYA (Bromham et al. 1998; Bromham and Hendy 2000).

Did the mineralized tissues of vertebrates appear in the Cambrian, as the fossil record suggests, or did they arise in the Proterozoic? In the present paper, we have tried to discriminate between these hypotheses using the amelogenin gene. We searched DNA banks for similarities between the amelogenin sequence and other sequences and tested the significance of these similarities, and, using several methods, we were able to date the origin of amelogenin exon 2 in the vertebrate genome.

Materials and Methods

The Amelogenin Genes

Amelogenin genes were cloned (DNA or cDNA) in nine mammalian species (including a marsupial and two monotremes), one squamate, one crocodile, and one anuran lissamphibian. Moreover, amelogenin genes were not detected in turtles and birds, two taxa that lost their teeth 150 and 60 MYA, respectively (Girondot and Sire 1998). This confirms, at least in sauropsids, the previous finding that the amelogenin gene lacks pleiotropic effects that could have enabled this gene to be maintained in the absence of enamel. Amelogenin gene sequences

are not available in nontetrapod vertebrates. However, recent immunocytochemical studies have revealed amelogenin-like proteins in the teeth of a lungfish, a basal sarcopterygian (Satchell, Shuler, and Diekwisch 2000), and in the ganoine of a polypterid, a basal actinopterygian (Zylberberg, Sire, and Nanci 1997). The ganoine that covers the scales in living basal actinopterygians is homologous to enamel (Huysseune and Sire 1998), and this is probably true for the ganoine of most early, extinct actinopterygians (Janvier 1996b). These results suggest that the gene responsible for amelogenin production was present at least in the common ancestor of both the sarcopterygian and the actinopterygian lineages. The presence of amelogenins and/or nonamelogenin proteins in chondrichthyan teeth is still largely controversial (see discussion in Zylberberg, Sire, and Nanci 1997). However, all of these results confirm the absence of pleiotropic effects for amelogenin in vertebrates and its exclusive involvement in enamel formation.

Amelogenin genes are unknown outside vertebrates. However, Slavkin and Diekwisch (1996) published a partial amino acid sequence of a Pacific hagfish amelogenin to support the immunodetection of amelogenin-like proteins in the oral cavity of this jawless craniate (*Hyperotreti sensu* Janvier 1996b). Living hagfish lack mineralized tissues, and this is probably a primitive condition (Janvier 1996a). Therefore, the presence of amelogenin in hagfish would have indicated that this gene was expressed in nonmineralized tissues in a common ancestor of craniates and vertebrates, a result that does not support our hypothesis of an essential role of amelogenin in enamel mineralization. A phylogenetic analysis of the putative hagfish DNA using mammalian amelogenin genes has demonstrated that this sequence was probably due to contamination by rodent DNA during PCR (Girondot, Delgado, and Laurin 1998). This finding was confirmed when the recently available amelogenin sequences in the caiman, the snake and *Xenopus* were added to the analysis (on the MBE website, see *Supplementary Material*, section 1, EMBL ALIGN_000059). Using the ProtML resampling of estimated log likelihoods (RELL) method (Hasegawa and Kishino 1994), if the putative hagfish sequence is constrained to be the sister group of amniotes, the resulting tree ($\ln L = -710.6$) is found to be worse than the maximum-likelihood tree (10,000 replicates) in 97.75% of the cases. Furthermore, a parsimony analysis using PAUP 3.1.1 (Swofford 1993) has indicated that four extra steps are required to locate the putative hagfish sequence outside of amniotes.

Searching for Sequence Similarities

A first BLAST search in GenBank using the various amelogenin exons did not reveal significant similarities to other sequences. Nevertheless, our attention was drawn by the following observation: rat ameloblasts (the epithelial cells that produce amelogenin) show a positive signal when hybridized in situ with an RNA probe of SPARC (secreted protein, acidic and rich in cystein) (Liao et al. 1998). This protein, also

called osteonectin, BM-40, and 43k protein, is a glycoprotein that is involved in mediating cell-matrix interactions but does not serve structural roles (Brekken and Sage 2000). Liao et al. (1998) have suggested that this signal is probably due to the presence of 10 identical bases in the rat sequences of SPARC and amelogenin. Indeed, the alignment of SPARC and amelogenin sequences has revealed that the complete translation of the second exons of both genes (the first exon is not translated) shows similarities. We searched for other SPARC sequences in DNA banks using Nentrez or BLAST tools. A total of 13 sequences were found: eight complete SPARC sequences in deuterostomians (dSPARC), two complete sequences in protostomians a nematode, and drosophila (pSPARC), and partial sequences (i.e., lacking exon 2) of three other deuterostomians. Three other SPARC-related genes were included in the analysis: SC1 in rodents and hevin, its homolog in humans, and QR1, its homolog in the quail. In the following, these three SPARC-related genes will be collectively called SC1 (the first identified). Moreover, in the course of our search, we found 19 zebrafish (*Danio rerio*) expressed sequence tag (EST) clones showing some similarities to SPARC. After extensive checking, we reconstructed the complete zebrafish SPARC mRNA sequence (on the *MBE* website, see *Supplementary Material*, section 2), and this sequence was added to our analysis (on the *MBE* website, see *Supplementary Material*, section 3, EMBL Align_000006).

Amelogenin sequences were also sought in the DNA banks using the same procedures, and 20 sequences were found. Only 12 amelogenin sequences were complete (with exon 2) and thus retained for the analysis.

In SPARC, SC1, and amelogenin exon 2 is composed of three regions: first region consists of 12 untranslated nucleotides, the second consists of 48 translated nucleotides constituting the signal peptide, and the third consists of six nucleotides that constitute the beginning of the protein itself. Regions 2 and 3 of dSPARC, SC1, and amelogenin exon 2 (a total of 19 sequences) can be aligned with only one tail of three gaps for the 54 total nucleotides (on the *MBE* website, see *Supplementary Material*, section 4, EMBL ALIGN_000005). The signal peptide is highly hydrophobic, and it enables the protein being synthesized (which is hydrophilic) to pass through the rough endoplasmic reticulum membrane. Once localized in the lumen of the endoplasmic reticulum, the protein is transported in the Golgi apparatus, then exocytosed by means of intracellular vesicles.

Testing Similarities

We needed to ensure that the dSPARC, SC1, and amelogenin sequences shared a common ancestor, i.e., that the observed similarity reflected a real homology. Thus, we tested the similarities of dSPARC, SC1, and amelogenin exon 2.

The proportions of identity of the various pairs of sequences cannot be directly compared because these values are not independent. Indeed, we suppose that all amelogenins, on the one hand, and dSPARC and SC1, on the other hand, have a common phylogenetic history. Therefore, the most parsimonious ancestral sequence for the second exon of each gene was calculated separately using MacClade 3.07 (Maddison and Maddison 1992). Due to ambiguities in the ancestral sequence, position i in a sequence j is fully described by the five frequencies f_{Aij} , f_{Gij} , f_{Cij} , f_{Tij} , and f_{ij} , which are, respectively, the probabilities that either an A, a G, a C, a T, or a gap is present at this position for this sequence (on the *MBE* website, see *Supplementary Material*, section 5). Then, the mean probability that two identical bases are present in an alignment of l bases is estimated as

$$\Phi = \sum_{i=1}^l (f_{Ai1} \cdot f_{Ai2} + f_{Ti1} \cdot f_{Ti2} + f_{Gi1} \cdot f_{Gi2} + f_{Ci1} \cdot f_{Ci2}) \div \left(l - \sum_{i=1}^l (f_{i1} \cdot f_{i2}) \right).$$

For each pair, the distribution of Φ under the null hypothesis (i.e., that the observed Φ value is not significantly different from what is expected for random sequences) was established using 1,000 replicates of two random computer-generated sequences sharing the same characteristics as the sequences under analysis (same length, same level of ambiguity, and same base composition). A simple position permutation was also used to generate random sequences. Each pair of computer-generated sequences was aligned to maximize the estimated Φ value. The corresponding value was used to estimate the distribution of Φ under H_0 (Φ_{H_0}). This test was performed with the three bases of the codons as well as with only bases 1 and 2 in order to disregard synonymous substitutions that occur mainly at the third position in each codon.

Searching for the Origin of Amelogenin Exon 2

BIO-NJ trees (Gascuel 1997) using either Kimura (1980) two-parameter or Tajima and Nei (1993) four-parameter distances were produced with only the translated part of SPARC, SC1, and amelogenin exon 2. Several methods were applied because the sequence (54 nt) was short and we needed to ensure that the result was significant. When these trees were rooted using the two pSPARC sequences, three monophyletic groups were always observed: dSPARC, SC1, and amelogenin. However, the relative positions of these three groups were uncertain.

The covarion method was applied to eliminate most of the substitution saturation (Lopez, Forterre, and Philippe 1999). Positions with more than one change inferred by parsimony in any of the four monophyletic groups (pSPARC, dSPARC, SC1, and amelogenin) were changed to missing character states (“?”) in the sequences of that group (equivalent to the H_1 matrix of Lopez, Forterre, and Philippe [1999]), and the gaps were considered “NEWSTATES” for parsimony analysis.

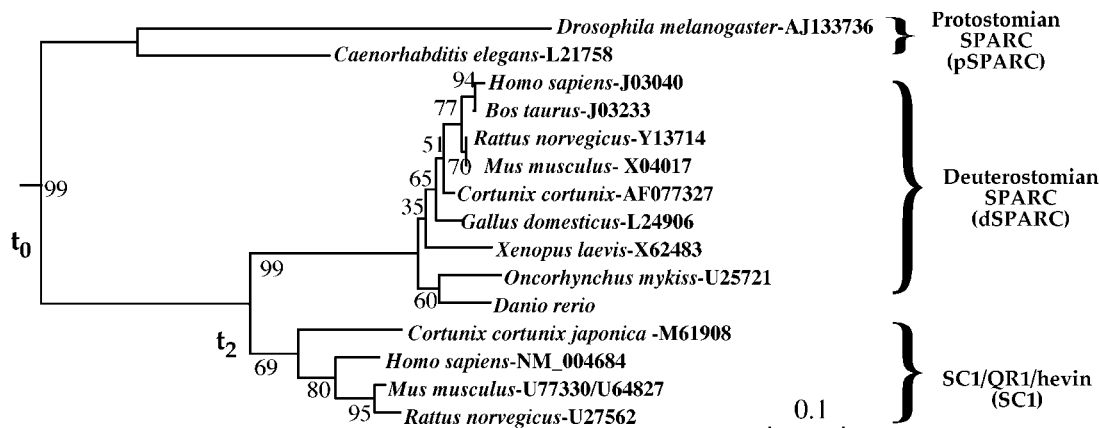


FIG. 1.—Gene tree for SPARC and SPARC-related genes produced using the amino acid sequences. Gonnet distances were applied with the neighbor-joining method of tree construction. Bootstraps are indicated at corresponding nodes. The dates t_0 and t_2 are, respectively, the divergence time between protostomians and deuterostomians, and the duplication that produced dSPARC and SC1 in deuterostomians.

This new alignment was analyzed using parsimony, and the relative bootstraps for the three possible positions of dSPARC, SC1, and amelogenin were established using 100 replicates with a heuristic search using PAUP 3.1.1 (Swofford 1993). Finally, the three possible relative positions of dSPARC, SC1, and amelogenin were tested using the REL method (Hasegawa and Kishino 1994) with default options. The phylogenetic positions of sequences among each monophyletic group (dSPARC, SC1, and amelogenin) were chosen according to the current accepted phylogeny of vertebrates (Kumar and Hedges 1998; Murphy et al. 2001).

Dating the Origin of Exon 2

Because of the small number of nucleotides in exon 2, it was not possible to date its origin. In contrast, dating the origin of the dSPARC/SC1 duplication (more than 1,000 nt available) was possible.

The different functions that have been attributed to amelogenin, SPARC, and SC1 proteins raise the possibility of differences in evolutionary rates for these paralogous genes; therefore, a method that could take into account this potential is needed. Rambaut and Bromham (1998) have proposed a quartet method in which the lineage-specific evolutionary rates are calculated using maximum-likelihood. This method requires that four species be grouped into pairs in a phylogeny and that the two internal dates of divergence be known. The following estimates of divergence dates were used: sauropsids versus synapsids, 310 MYA; lissamphibians versus amniotes, 360 MYA; actinopterygians versus sarcopterygians, 410 MYA (Kumar and Hedges 1998). This method uses internal references for divergence dates and is more suitable for our study than a method that uses external divergence dates, such as the date of divergence between protostomians and deuterostomians, which is highly controversial (Ayala and Rzhetsky 1998; Bromham et al. 1998; Gu 1998; Bromham and Hendy 2000). Moreover, for this phase of metazoan evolution, the fossil record is extremely poor (Valentine, Jablonski, and Erwin 1999).

To estimate the duplication date of SPARC and SC1, all of the quartet combinations involving two dSPARC sequences on one hand and two SC1 sequences on the other were produced ($n = 84$). For each of these quartets, the substitution rate was calculated using the two-parameter substitution model of Hasegawa, Kishino, and Yano (1985). The substitution rate estimated was tested against the five-parameter substitution model. A significant difference implies that the two-parameter substitution model is not sufficient to produce a reliable estimate of the substitution rate because the five-parameter substitution model significantly alters the estimate; none of these 84 tests were rejected at the 5% level. Therefore, minimal and maximal dates at the 95% confidence limit were estimated for each of the 84 quartets using the two-parameter substitution model. These 84 values were not independent of each other because these sequences shared a common phylogenetic history; therefore, all of the results of the estimated duplication dates were used. The molecular-clock method used here allowed a molecule-specific evolutionary rate (SPARC or SC1/QR1/Hevin) (Rambaut and Bromham 1998). The consistency among the 84 estimates suggests that species-specific evolutionary rates for a given molecule (SPARC or SC1/QR1/Hevin) are not widely divergent.

Results

The SPARC tree reveals that a duplication of an ancestral SPARC gene within the deuterostomian lineage gave rise to SC1/hevin/QR1 and to dSPARC and that this event occurred before the divergence between actinopterygians and sarcopterygians (fig. 1). Therefore, SC1, hevin, and QR1 are considered orthologous to each other and they are paralogous to dSPARC. This is confirmed by the absence of SC1/hevin and QR1 in the complete sequenced genomes of two protostomians, *Caenorhabditis* and *Drosophila*.

The Origin of SPARC, SC1, and Amelogenin Exon 2

The Φ values were 0.4787 for dSPARC versus SC1, 0.5064 for dSPARC versus amelogenin, and

Probability of identical bases in alignment

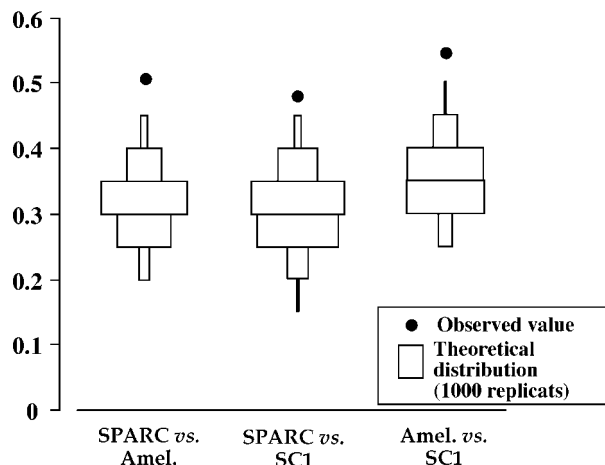


FIG. 2.—Comparison and test for significance of probabilities of identical bases in the exon 2 alignment of SPARC, SC1, and amelogenin when the three bases of the codon are used. The close points are observed values of probability of identical bases in exon 2 for comparisons between pairs of ancestral sequences (reconstructed using parsimony). The histograms show the distribution of this value for the null hypothesis, i.e., that the sequences are not homologous, using computer-generated sequences.

0.5447 for amelogenin versus SC1 when the three bases of the codons were analyzed, and these values were 0.5625, 0.6038, and 0.6476, respectively, when only bases 1 and 2 were taken into account. When similarities of the deuterostomian exon 2 sequences were tested, these observed values were always significantly higher than the expected randomly generated similarities (Φ_{Ho} ; $P < 0.001$ for dSPARC vs. SC1, dSPARC vs. amelogenin, and amelogenin vs. SC1 whatever the method used to generate them; see *Materials and Methods*) (fig. 2). These similarities clearly demonstrate a common evolutionary origin of this exon. An internal test of the method (i.e., that the method does not overestimate the similarity) was obtained when pSPARC exon 2 was compared with any of the deuterostomian exons 2. Indeed, the genomic organization of protostomian and deuterostomian SPARC exons, the function of the signal peptide, and the hydrophobicity plot of the protein (not shown) suggest that exon 2 of dSPARC and pSPARC are really homologous, but the similarity was not found to be statistically significant. This indicates that mutations have erased most of the phylogenetic signal. It should also be noted that the three-base indels required for a correct alignment of the deuterostomian sequences (dSPARC, SC1, and amelogenin) do not change the open reading frame.

Another argument for a common evolutionary history of exon 2 is obtained from the genomic organization. Indeed, the 3' end of the zone of similarity in the sequences of amelogenin and SPARC/SC1 exactly matches the end of exon 2.

Date for the Presence of Exon 2 in Amelogenin

The tree of exon 2 shows that the signal peptide it encodes was duplicated from a SPARC sequence to

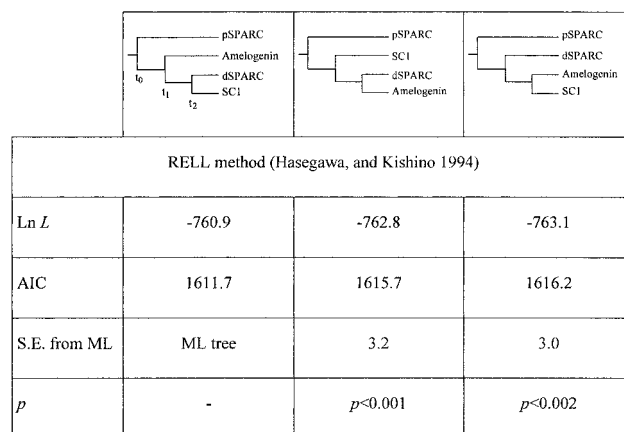


FIG. 3.—Three possible relative relationships between dSPARC, SC1, and amelogenin exon 2 tested using the resampling of estimated log likelihoods method (Hasegawa and Kishino 1994). *L* = likelihood; AIC = Akaike information criterion (Akaike 1974); S.E. from ML = standard error from the maximum likelihood. This represents the distance (i.e., the numbers of standard errors of the likelihood) from one tree to the best one (the ML tree). The associated *P* value is the probability that such a likelihood is observed with the ML tree. The dates t_0 , t_1 , and t_2 are, respectively, the divergence time between protostomians and deuterostomians, and the two duplication events that produced (1) amelogenin and dSPARC and (2) SC1 in deuterostomians.

amelogenin before the duplication between dSPARC and SC1. This conclusion was supported by all phylogenetic analyses of these sequences (fig. 3). This duplication occurred before the separation between actinopterygians and sarcopterygians, because amelogenin exon 2 falls outside the cluster formed by the actinopterygian (*D. rerio* and *Oncorhynchus mykiss*) and the sarcopterygian (which includes the tetrapod sequences) SC1.

Moreover, the acquisition of exon 2 by amelogenin predates the duplication between dSPARC/SC1, which most likely occurred around 630 MYA (fig. 4). Therefore, the latter duplication date may be considered an underestimation of the date of exon 2 acquisition in the amelogenin locus. This date was long before the Cambrian explosion and the sudden appearance of mineralization in the fossil record (543 MYA) and earlier than the Vendian (610 MYA), from which several presumed metazoan fossils are known (Valentine, Jablonski, and

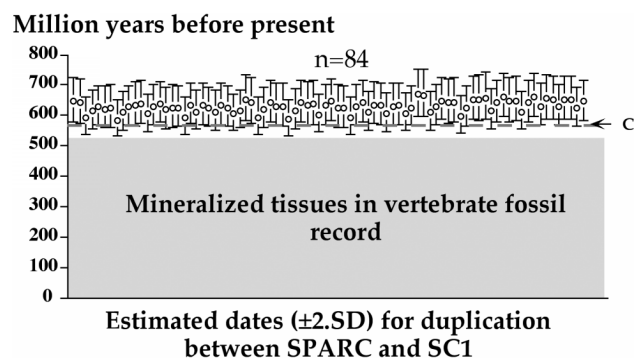


FIG. 4.—Estimated dates of duplication between SPARC and SC1 (t_2 in figs. 1 and 3) using the quartet method (Rambaut and Bromham 1998). "C" indicates the boundary between the Precambrian and the Cambrian.

Erwin 1999). The origin of the other parts of the amelogenin sequence is still unknown. The function of exon 2 (the peptide signal), which consists in passing the proteins through the endoplasmic reticulum membrane, is essential to transfer these proteins out of the cells and into the extracellular matrix. Enamel could not exist if amelogenin were not secreted by the ameloblasts, because this protein plays a crucial role in enamel mineralization (Deutsch 1989; Schwarzbauer and Spencer 1993).

Theoretically, the maximal period during which a nonexpressed amelogenin sequence can persist in a genome while remaining potentially functional was estimated not to exceed 10 Myr (Marshall, Raff, and Raff 1994). Beyond this time, the amelogenin gene would have been lost through accumulation of mutations.

Therefore, we conclude that the date of acquisition of exon 2 for amelogenin probably coincides with the date of amelogenin appearance. This probably means that vertebrate enamel (or its precursor) appeared no later than 630 MYA, even taking into account the standard deviation of the estimate (fig. 4).

Discussion

In developing a dating method using a gene (amelogenin) involved in enamel mineralization and without any known pleiotropic effects, we have demonstrated that exon 2 of the amelogenin gene was recruited from the duplication of the exon 2 of a SPARC gene. This duplication most likely occurred in the Precambrian, 630 MYA, nearly 100 Myr before the Cambrian explosion, a period characterized by the sudden appearance of mineralization in the fossil record. The origin of amelogenin exon 2 is close to the beginning of the Vendian (610 MYA), in which several presumed metazoan fossils are known (Valentine, Jablonski, and Erwin 1999). The molecular-clock method used in the present study (Rambaut and Bromham 1998) allowed a molecule-specific evolutionary rate (dSPARC or SC1). A method allowing a species-specific rate may be more reliable, but none is currently available (Pagel 1999). The consistency between the 84 estimates suggests that species-specific evolutionary rates for a given molecule (dSPARC or SC1) are not widely divergent.

The acquisition of exon 2 in the "pre-amelogenin locus" is slightly anterior to the date of dSPARC and SC1 duplication, which occurred nearly 630 MYA (t_2 in fig. 1). The date calculated for the duplication of dSPARC and SC1 may be considered a reliable estimate for the latest possible date of exon 2 acquisition in the amelogenin locus.

The lack of mineralized fossils in the Proterozoic may be due to the very small sizes of individuals or to the fragility of the mineralized skeleton (e.g., the scales of *Anatolepis* are less than 0.1 mm thick). Metazoan fossils have been found in only a few Proterozoic sites so far. We would not be surprised if mineralized fossils were found in late Proterozoic deposits.

The history of the bilateralian phyla is characterized by their sudden appearance in the fossil record

within the early Cambrian at the Proterozoic-Phanerozoic transition (approximately 543 MYA), in the so-called "Cambrian explosion" (Morris 1997; Knoll and Carroll 1999; Valentine, Jablonski, and Erwin 1999). Vertebrates were no exception to this rule, and the first early Cambrian fossils indisputably belonging to the vertebrate lineage were recently described from a Chinese locality (Chen, Huang, and Li 1999; Shu et al. 1999). One of these fossils may be related to lampreys, and the other is probably a basal vertebrate. The sudden appearance of bilateralian clades during the 10-Myr duration of the Cambrian explosion has long been interpreted as the result of a great and rapid evolutionary radiation (Valentine, Jablonski, and Erwin 1999). However, this finding has been questioned by several molecular studies.

Using various molecular-clock models with several sets of genes, the origins of many bilateralian clades are consistently estimated to be older than the Cambrian explosion (Fortey, Briggs, and Wills 1997), and the divergence between protostomians and deuterostomians has been estimated to have occurred between 830 and 1,000 Myr during the Proterozoic (Wray, Levinton, and Shapiro 1996; Bromham et al. 1998; Gu 1998). However, all of these estimates are subject to caution, because most of them use a molecular clock that is supposed to tick regularly through the studied period. This hypothesis has recently been tested, and the same conclusion applies (Bromham and Hendy 2000).

However, this hypothesis has recently received support from the fossil record: a putative mollusk-like animal (Fedonkin and Waggoner 1998), many burrows (Valentine, Jablonski, and Erwin 1999), and some spiralian embryos (such as annelids or mollusks) (Bengtson 1998; Xiao, Zhang, and Knoll 1998) have been described from the late Proterozoic. Therefore, the great increase in abundance of metazoan fossils in the Cambrian explosion could simply be a taphonomic artifact due to either environmental conditions not being conducive to fossilization or the paucity of Bilateralialia in the Proterozoic, or their small size, or their lack of mineralization. Whatever the condition it made fossilization much less likely and has led to the incorrect conclusion that Bilateralialia diversified in a short period. The hypothesis of a lack of mineralization is supported by the observation of a simultaneous acquisition of mineralization in most animal phyla during the Cambrian (Bengtson and Conway Morris 1992). The few Proterozoic fossils that have been interpreted as belonging to the bilateralian clade lack mineralized tissues. The most conclusive example is that of *Kimberella*, which was reinterpreted as a mollusk (placophore?) without a mineralized shell (Fedonkin and Waggoner 1998). Also, some of early Cambrian bilateralian fossils, which are sister groups of clades that subsequently acquired mineralization, are themselves not mineralized. For example, lobopods from the early Cambrian may be the sister group of arthropods, but they possess a nonmineralized cuticle (Min, Kim, and Kim 1998).

The phylogeny of craniates presented by Janvier (1996a) suggests that the absence of mineralized tissues

in hagfish and lampreys is primitive. If this conclusion is accepted, our results suggest that stem vertebrates (such as *Anatolepis* or Thelodonts) were already present 630 MYA.

Our study is the first attempt to discriminate between hypotheses of a late, rapid evolutionary radiation of metazoans (supported by a literal interpretation of the fossil record) and an earlier, slow evolutionary radiation (supported by molecular dates of divergence of metazoan taxa) using a molecular date of acquisition of amelogenin, a protein involved in enamel mineralization.

Acknowledgments

We thank Professors Armand de Ricqlès and Ann Huyseune for their careful reading of the manuscript and for many valuable suggestions. We are also indebted to Matthew H. Godfrey for English checking.

LITERATURE CITED

- AKAIKE, H. 1974. A new look at the statistical model identification. *IEEE Trans. Automatic Control* **19**:716–723.
- ALDRIDGE, R. J., and P. C. J. DONOGHUE. 1998. Conodonts: a sister group to hagfishes? Pp. 15–31 in J. M. JØRGENSEN, J. P. LOMHOLT, R. E. WEBER, and H. MALTE, eds. *The biology of hagfishes*. Chapman and Hall, London.
- AOBA, T. 1996. Recent observations on enamel crystal formation during mammalian amelogenesis. *Anat. Rec.* **245**:208–218.
- AYALA, F. J., and A. RZHETSKY. 1998. Origin of the metazoan phyla: molecular clocks confirm paleontological estimates. *Proc. Natl. Acad. Sci. USA* **95**:606–611.
- BENGSTON, S., and S. CONWAY MORRIS. 1992. Early radiation of biomineralizing phyla. Pp. 447–481 in J. H. LIPPS and P. W. SIGNOR, eds. *Origin and evolution of the Metazoa*. Plenum Press, New York.
- BENGSTON, S. 1998. Evolution: animal embryos in deep time. *Nature* **391**:529–530.
- BREKKEN, R. A., and E. H. SAGE. 2000. SPARC, a matricellular protein: at the crossroads of cell-matrix. *Matrix Biol.* **19**:569–580.
- BROMHAM, L. D., and M. D. HENDY. 2000. Can fast early rates reconcile molecular dates with the Cambrian explosion. *Proc. R. Soc. Lond. B Biol. Sci.* **267**:1041–1047.
- BROMHAM, L., A. RAMBAUT, R. FORTEY, A. COOPER, and D. PENNY. 1998. Testing the Cambrian explosion hypothesis by using a molecular dating technique. *Proc. Natl. Acad. Sci. USA* **95**:12386–12389.
- CHEN, J.-Y., D.-Y. HUANG, and C.-W. LI. 1999. An early Cambrian craniate-like chordate. *Nature* **402**:518–522.
- DEUTSCH, D. 1989. Structure and function of enamel gene product. *Anat. Rec.* **224**:189–210.
- DEUTSCH, D., A. PALMON, L. DAFNI, J. CATALANO-SHERMAN, M. F. YOUNG, and L. W. FISHER. 1995. The enamelin (tufelin) gene. *Int. J. Dev. Biol.* **39**:135–143.
- FEDONKIN, M. A., and B. M. WAGGONER. 1998. The Late Precambrian fossil *Kimberella* is a mollusc-like bilaterian organism. *Nature* **388**:868–871.
- FINCHAM, G., A. B. BELCOURT, D. M. LYARUU, and J. D. TERMINE. 1982. Comparative protein biochemistry of developing dental enamel matrix from five mammalian species. *Calcif. Tissue Int.* **34**:182–189.
- FONG, C. D., I. SLABY, and L. HAMMARSTRÖM. 1996. Amelin: an enamel-related protein, transcribed in the cells of epithelial root sheath. *J. Bone Miner. Res.* **11**:892–898.
- FORTEY, R. A., D. E. BRIGGS, and M. A. WILLS. 1997. The Cambrian evolutionary ‘explosion’ recalibrated. *BioEssays* **19**:429–434.
- GASCUEL, O. 1997. BIONJ: an improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* **14**:685–695.
- GIRONDOT, M., S. DELGADO, and M. LAURIN. 1998. Evolutionary analysis of “hagfish amelogenin.” *Anat. Rec.* **252**:608–611.
- GIRONDOT, M., and J.-Y. SIRE. 1998. Evolution of the amelogenin gene in toothed and tooth-less vertebrates. *Eur. J. Oral Sci.* **106**(Suppl. 1):501–508.
- GU, X. 1998. Early metazoan divergence was about 830 million years ago [letter]. *J. Mol. Evol.* **47**:369–371.
- HASEGAWA, M., and H. KISHINO. 1994. Accuracies of the simple methods for estimating the bootstrap probability of a maximum-likelihood tree. *Mol. Biol. Evol.* **11**:142–145.
- HASEGAWA, M., H. KISHINO, and T. YANO. 1985. Dating the human-ape splitting by a molecular clock of mitochondrial DNA. *J. Mol. Evol.* **22**:160–174.
- HUYSEUNE, A., and 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, and S. OIDA. 1998. Amelogenin protein in tooth germs of the snake *Elaphe quadrivirgata*, immunohistochemistry, cloning and cDNA sequence. *Arch. Histol. Cytol.* **61**:467–474.
- JANVIER, P. 1996a. The dawn of the vertebrates: characters *versus* common ascent in the rise of current vertebrate phylogenies. *Paleontology* **39**:259–287.
- . 1996b. *Early vertebrates*. Clarendon Press, Oxford, England.
- KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**:111–120.
- KNOLL, A. H., and S. B. CARROLL. 1999. Early animal evolution: emerging views from comparative biology and geology. *Science* **284**:2129–2137.
- KREBSBACH, P. H., S. K. LEE, Y. MATSUKI, C. A. KOZAK, K. M. YAMADA, and Y. YAMADA. 1996. Full-length sequence, localization, and chromosomal mapping of ameloblastin. A novel tooth-specific gene. *J. Biol. Chem.* **271**:4431–4435.
- KUMAR, S., and S. B. HEDGES. 1998. A molecular timescale for vertebrate evolution. *Nature* **392**:917–920.
- LIAO, H., C. BRANDSTEN, C. LUNDMARK, C. CHRISTERSSON, and T. WURTZ. 1998. Osteonectin RNA and collagen $\alpha 1(I)$ RNA in the developing rat maxilla. *Eur. J. Oral Sci.* **106**(Suppl. 1):418–423.
- LOPEZ, P., P. FORTERRE, and H. PHILIPPE. 1999. The root of the tree of life in the light of the covarion model. *J. Mol. Evol.* **49**:496–508.
- LYNGSTADAAS, S. P., S. RISNES, B. S. SPROAT, P. S. THRANE, and H. P. PRYDZ. 1996. A synthetic, chemically-modified ribozyme eliminates amelogenin, the major translation product in developing mouse enamel *in vivo*. *EMBO J.* **14**:5224–5229.
- MADDISON, W. P., and D. R. MADDISON. 1992. *MacClade: analysis of phylogeny and character evolution*. Sinauer, Sunderland, Mass.
- MARSHALL, C. R., E. C. RAFF, and R. A. RAFF. 1994. Dollo’s law and the death and resurrection of genes. *Proc. Natl. Acad. Sci. USA* **91**:12283–12287.
- MIN, G. S., S. H. KIM, and W. KIM. 1998. Molecular phylogeny of arthropods and their relatives: polyphyletic origin of arthropodization. *Mol. Cells* **8**:75–83.

- MORRIS, S. C. 1997. Defusing the Cambrian 'explosion'? *Curr. Biol.* **7**:R71–R74.
- MURPHY, W. J., E. EIZIRIK, W. E. JOHNSON, Y. P. ZHANG, O. A. RYDER, and S. J. O'BRIEN. 2001. Molecular phylogenetics and the origins of placental mammals. *Nature* **409**:614–618.
- ØRVIG, T. 1989. Histologic studies of ostracoderms, placoderms and fossil elasmobranchs. 6. Hard tissues of Ordovician vertebrates. *Zool. Scripta* **18**:427–446.
- PAGEL, M. 1999. Inferring the historical patterns of biological evolution. *Nature* **401**:877–884.
- RAMBAUT, A., and L. BROMHAM. 1998. Estimating divergence dates from molecular sequences. *Mol. Biol. Evol.* **15**:442–448.
- REPETSKI, J. E. 1978. A fish from the Upper Cambrian of North America. *Science* **200**:529–531.
- SANSOM, I. J., M. P. SMITH, H. A. ARMSTRONG, and M. M. SMITH. 1992. Presence of the earliest vertebrate hard tissue in conodonts. *Science* **256**:1308–1311.
- SASAKI, S., and H. SHIMOKAWA. 1995. The amelogenin gene. *Int. J. Dev. Biol.* **39**:127–133.
- SATCHELL, P. G., C. F. SHULER, and G. H. DIEKWISCH. 2000. True enamel covering in teeth of the Australian lungfish *Neoceratodus forsteri*. *Cell Tissue Res.* **299**:27–37.
- SCHULTZE, H. P. 1996. Conodont histology: an indicator of vertebrate relationship? *Mod. Geol.* **20**:275–285.
- SCHWARZBAUER, J. E., and C. S. SPENCER. 1993. The *Caenorhabditis elegans* homologue of the extracellular calcium binding protein SPARC/osteonectin affects nematode body morphology and mobility. *Mol. Biol. Cell* **4**:941–952.
- SHU, D.-G., H.-L. LUO, S. CONWAY MORRIS, X.-L. ZHANG, S.-X. HU, L. CHEN, J. HAN, M. ZHU, Y. LI, and L.-Z. CHEN. 1999. Lower Cambrian vertebrates from south China. *Nature* **402**:21–22.
- SLAVKIN, H. C., and T. DIEKWISCH. 1996. Evolution in tooth developmental biology: of morphology and molecules. *Anat. Rec.* **245**:131–150.
- SMITH, M. P., I. J. SANSOM, and J. E. REPETSKI. 1996. Histology of the first fish. *Nature* **380**:702–704.
- SWOFFORD, D. L. 1993. PAUP: phylogenetic analysis using parsimony. Version 3.1.1. Smithsonian Institution, Washington, D.C.
- TAJIMA, F., and M. NEI. 1993. Unbiased estimation of evolutionary distance between nucleotide sequences. *Mol. Biol. Evol.* **10**:677–688.
- TERMINE, J. D., A. B. BELCOURT, P. V. CHRISTNER, K. M. CONN, and M. U. NYLEN. 1980. Properties of dissociatively extracted fetal tooth matrix proteins. I. Principal molecular species in developing bovine enamel. *J. Biol. Chem.* **255**:9760–9768.
- TOYOSAWA, S., C. O'HUIGIN, F. FIGUEROA, H. TICHY, and J. KLEIN. 1998. Identification and characterization of amelogenin genes in monotremes, reptiles, and amphibians. *Proc. Natl. Acad. Sci. USA* **95**:13056–13061.
- VALENTINE, J. W., D. JABLONSKI, and D. H. ERWIN. 1999. Fossils, molecules and embryos: new perspectives on the Cambrian explosion. *Development* **126**:851–859.
- WRAY, G. A., J. S. LEVINTON, and L. H. SHAPIRO. 1996. Molecular evidence for deep pre-Cambrian divergence among metazoan phyla. *Science* **274**:568–573.
- XIAO, S., Y. ZHANG, and A. W. KNOLL. 1998. Three-dimensional preservation of algae and animal embryos in a neoproterozoic phosphorite. *Nature* **391**:553–558.
- ZYLBERBERG, L., J.-Y. SIRE, and A. NINCI. 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.

PIERRE CAPY, reviewing editor

Accepted July 16, 2001