

Evolutionary Analysis of “Hagfish Amelogenin”

MARC GIRONDOT,* SIDNEY DELGADO, AND MICHEL LAURIN

URA 1137, Evolution et Adaptations des Systèmes Ostéo-musculaires, CNRS and
Université Paris 7, Case 7077, 75251 Paris cedex 05, France

ABSTRACT

Hagfishes lack mineralized tissues and teeth. Part of a cDNA strand, allegedly from amelogenin, the major gene involved in enamel formation in mammals, has recently been cloned in a hagfish (Slavkin and Diekwish, *Anat. Rec.*, 1996;245:131–150). This cloning is of great interest because it could change the current view about the evolution of mineralized tissues, but no phylogenetic analysis of this piece of DNA has been made by the authors.

Phylogenetic analysis of this part of cDNA has been conducted using both phenetic and cladistic methods.

The cDNA amplified in hagfish does not fit with a nonmammalian origin but fits well with a degraded rodent sequence.

The gene cloned in hagfish is probably of mammalian origin due to contamination during PCR. *Anat. Rec.* 252:608–611, 1998.

© 1998 Wiley-Liss, Inc.

Key words: amelogenin; teeth; evolution; hagfish; PCR contamination

Hagfishes are eel-shaped, jawless craniates belonging to the Hyperotreti, the sister group of vertebrates (Janvier, 1996). All extant hagfishes and all known fossil Hyperotreti lack mineralized tissues; their skeleton is cartilaginous, and their mouth is armed with horny teeth (odontoids or toothlets). The absence of mineralized tissues, which are otherwise found in all gnathostomes and in many groups of fossil jawless vertebrates, has been variously interpreted. While many early phylogenies implied that hagfishes had lost the ability to produce mineralized tissues, the current consensus is that the ancestors of hagfishes never possessed this ability (Janvier, 1993). The recent cloning of a piece of hagfish (*Eptatretus stoutii*) cDNA by RT-PCR using amelogenin primers (Slavkin and Diekwish, 1996) is of great interest in the perspective of the evolution of mineralized tissues but is quite surprising. Indeed, amelogenin has been known for a few years to be involved in the formation of mammalian tooth enamel (review by Deutsch, 1989), and recently we have shown that the amelogenin gene is probably absent in toothless sauropsids, such as turtles and birds (Girondot and Sire, 1998). This reinforces the idea that the only role of amelogenin in amniotes is to contribute to enamel formation and that it is lost in taxa that lack selective pressure to maintain its integrity (in less than 200 My in the case of turtles and in less than 100 My in the case of birds). The presence of amelogenin in hagfishes implies either that this gene has another function (at least in hagfishes) or, if

we assume that the ancestors of hagfishes once had a mineralized skeleton (as suggested by most early phylogenies), that this apparently inactive gene has been retained for over 300 My (the age of the oldest known hagfish).

The existence of an amelogenin gene in hagfishes has been used by Slavkin and Diekwish (1996) to validate the observed cross-reactivity in hagfish toothlets of polyclonal antibodies against mammalian amelogenin (Slavkin et al., 1982, 1983, 1991). However, the published 50 amino-acid sequence of the hagfish amelogenin is very similar to the known eutherian sequences. The authors have not performed any quantitative measure of divergence between these sequences, although such an analysis either would have permitted confirmation of the nonmammalian origin of the sequence or would have detected any source of contamination during the PCR phase of the cloning protocol. Therefore, to determine the origin of the hagfish amelogenin, we have performed a phylogenetic test.

Grant sponsor: Alexander von Humboldt Foundation.

*Correspondence to: Marc Girondot, URA 1137, Evolution et Adaptations des Systèmes Ostéo-musculaires, CNRS and Université Paris 7, Case 7077, 2 place Jussieu, 75251 Paris cedex 05, France. E-mail: mgi@ccr.jussieu.fr

Received 12 March 1998; Accepted 25 June 1998

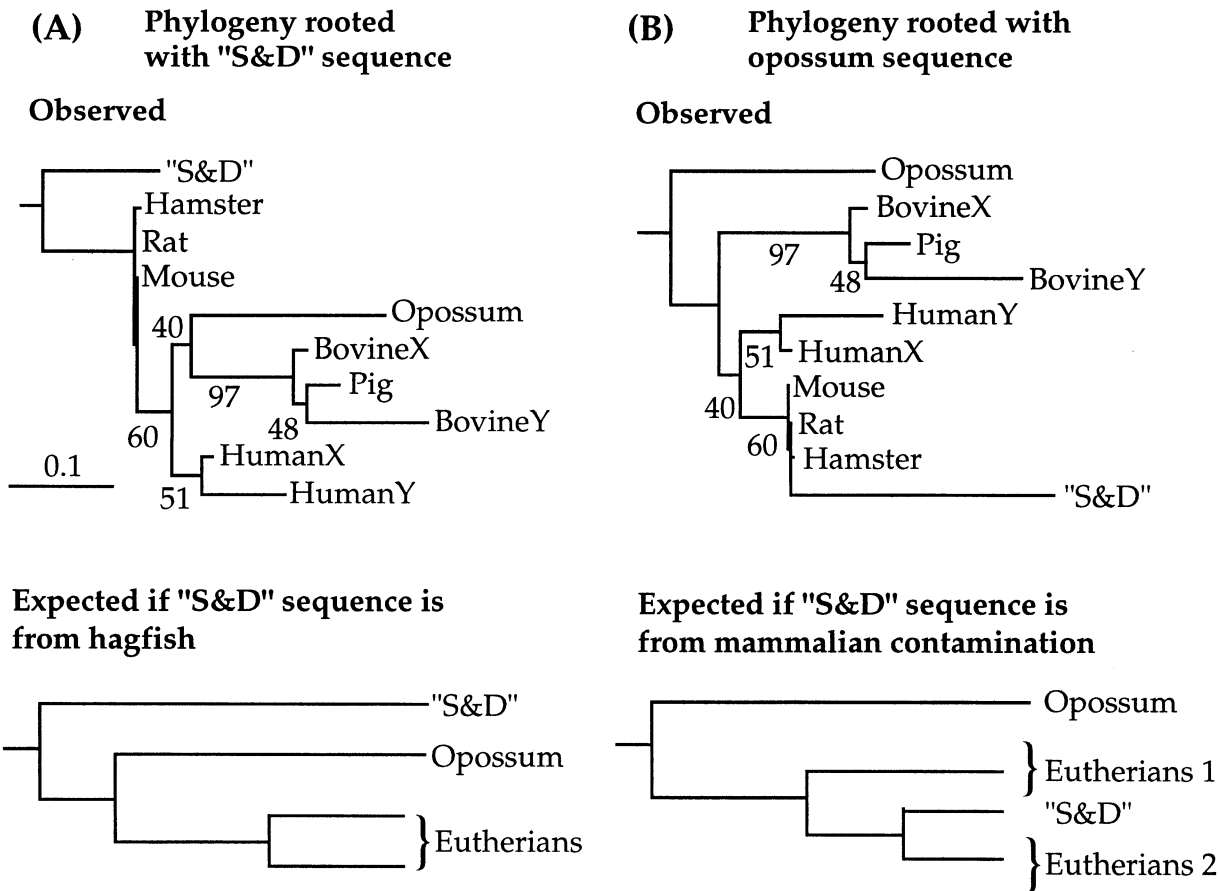


Fig. 1. Tree obtained by the BIONJ algorithm (Gascuel, 1997) with PAM 350 matrix for amino-acid distances and visualized using TreeView 1.4 (Page, 1996). The putative hagfish sequence is indicated by "S&D." The trees are rooted at the divergence between "S&D" and other sequences in panel **A** or rooted at the divergence of metatherian and

other sequences in panel **B**. Indicated below each tree is the expected topology if the "S&D" sequence is a hagfish sequence (**A**) or if the "S&D" sequence is a eutherian sequence (**B**). Bootstrap values in % were obtained using 1,000 replicates.

Recently we have demonstrated that the amelogenin gene can be useful for studies of mammalian phylogeny (Girondot and Sire, 1998). It is possible to demonstrate a nonmammalian origin of any sequence by establishing its basal position in the phylogeny, although no outgroup is available because only mammalian sequences are known. The phylogenetic relationships between hagfishes, metatherians (marsupials), and eutherians (placental mammals) within craniates are known without ambiguity, and this phylogeny can be used as a true phylogeny (Janvier, 1996). The mammalian amelogenin phylogeny should include an early divergence between the opossum and the eutherian sequences, and a hagfish amelogenin sequence should fit outside Mammalia. However, since no outgroup is available for this analysis, several rooting options are possible, but only two are biologically meaningful. First, the trees can be rooted between hagfishes and mammals (Figs. 1A, 2A). If the reported sequences truly belong to hagfishes, the "correct" dichotomy between metatherians and eutherians should be found. Second, the trees can be rooted between metatherians (the opossum) and the other taxa (Figs. 1B, 2B). If the reported sequences were the result of contamination from a eutherian, the "hagfish"

sequences should form a clade with the species from which the DNA actually originated.

Alignment has been done using Clustal X 1.64b (Thompson et al., 1994), but the sequences can be aligned without any ambiguity by hand. Only one gap is required in the human X gene (the amelogenin gene is located on the heterosomal part of the sex chromosomes in eutherians [Girondot and Sire, 1998]). Analyses have been performed using both a distance method (phenetic) and a parsimony method (cladistic) (see legend of figures for details of computing procedures). For the distance method, a recent algorithm (BIONJ [Gascuel, 1997]) has been used to minimize the effect of different substitution rates in the X and Y mammalian chromosomal lineage of amelogenin (Huang et al., 1997).

Using both a BIONJ distance tree (Fig. 1) and a consensus parsimony tree (Fig. 2), the topology of the inferred phylogeny is never consistent with a noneutherian origin of the putative hagfish sequences reported by Slavkin and Diekwish (1996). In both cases (Figs. 1A, 2A), if the trees are rooted between the presumed hagfish sequences and the other taxa, the opossum is deeply nested in Eutheria.

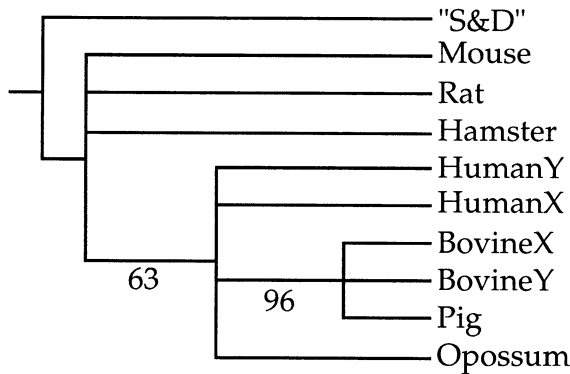
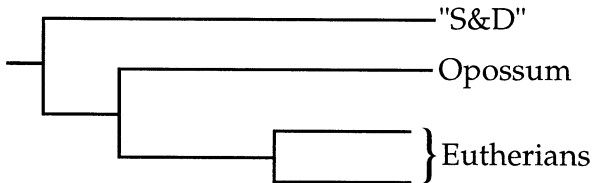
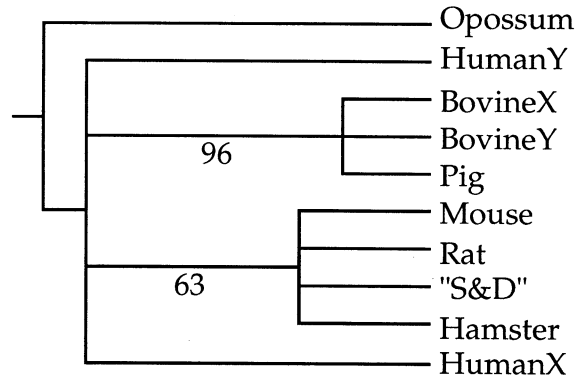
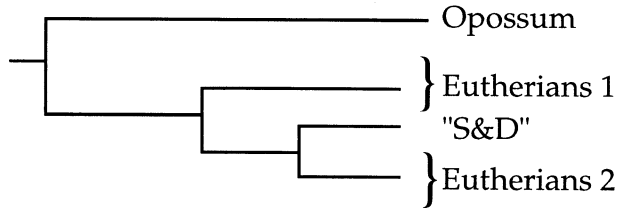
(A) Phylogeny rooted with "S&D" sequence**Observed****Expected if "S&D" sequence is from hagfish****(B) Phylogeny rooted with opossum sequence****Observed****Expected if "S&D" sequence is from mammalian contamination**

Fig. 2. Strict consensus tree of the 22 shortest trees (41 steps) obtained by the parsimony method using an exhaustive search (2,027,025 trees analyzed) in PAUP 3.1.1 (Swofford, 1993), with the collapsing zero-length branches option. The putative hagfish sequence is indicated by "S&D." The trees are rooted at the divergence between "S&D" and other sequences in panel **A** or rooted at the divergence of metatherian and other sequences in panel **B**. Indicated below each tree is the

expected topology if the "S&D" sequence is a hagfish sequence (**A**) or if the "S&D" sequence is a eutherian sequence (**B**). Bootstrap values in % were obtained using 1,000 replicates with the branch-and-bound algorithm. The 50% majority-rule consensus of the resulting trees has the same topology as the strict consensus of the shortest trees obtained by the exhaustive search.

This result suggests that the "hagfish" sequence actually represents a eutherian contaminant.

To assess the plausibility of a contaminant origin of the "hagfish" sequence and to identify the actual source of the DNA sequence, we rerooted the trees between the opossum and the other taxa (Fig. 1B and 2B). In this case, the "hagfish" sequences clustered with the rodent sequences (rat, mouse and hamster) (Figs. 1B, 2B). Twenty-two shortest trees (41 steps) are obtained by the parsimony method, and in all these trees the "hagfish" sequence clustered with the rodent ones (Fig. 2B). The "correct" basal topology is obtained in only three of the 118 trees that require up to one extra step, and a strict consensus of these trees gives a topology inconsistent with the actual knowledge on the evolution of amelogenin based on the entire gene sequence (Girondot and Sire, 1998). Furthermore, the "hagfish" sequence still clustered with at least one rodent sequence in the remaining 115 of these 118 trees. This suggests that the position of hagfishes as a sister group of all the placental sequences in three of these 118 trees is not significant. Finally, 63% and 60% of the 1,000 bootstrap replicates using, respectively, parsimony or BIONJ methods link "hagfish" with rodent sequences,

whereas the analyzed sequences are relatively short (Figs. 1B, 2B). The classical neighbor-joining method (Saitou and Nei, 1987) gives exactly the same tree topology as the BIONJ method (not shown).

The divergence time between hagfishes and the other species analyzed here is approximately 470 million years, whereas the divergence time between metatherians (marsupials) and eutherians (placental mammals) is only 120 million years. Thus, the results obtained here cannot be explained by rapid speciation events. More probably, the published "hagfish" sequence is a mammalian sequence obtained by contamination during PCR amplification. This artifact may result from the use of either some potentially degenerate primers, a low annealing temperature, or a high number of PCR cycles (the primer sequences, the annealing temperature, and the number of PCR cycles are not described in the original article). Such a contamination is frequent when cloning genes of distant species using PCR (see, for example, the fungal and angiosperm origin of putative dinosaur ribosomal genes (Wang et al., 1997) or the human origin of putative dinosaur mitochondrial gene [Collura and Stewart, 1995]) and might be detected before publication by a phylogenetic analysis of the produced

sequence. The differences between the putative "hagfish" sequence and the most similar mammalian one could be due to the sequencing of a formerly unsequenced mammalian amelogenin gene or more probably to the sequencing of a contaminant, degraded mouse gene.

Our conclusion that the putative "hagfish" amelogenin gene sequence published by Slavkin and Diekwisch (1996) probably originates from a mammalian contaminant does not invalidate the results obtained using mammalian antibodies (Slavkin et al., 1982, 1983, 1991), but this sequence cannot be used to validate them. The primary sequence of amelogenin cDNA in hagfish remains to be found.

ACKNOWLEDGMENTS

We thank Patricia Lai for correction of this manuscript and Jean-Yves Sire and Armand de Ricqlès (URA 1137) for critical reading and many valuable suggestions. Michel Laurin was supported by the Alexander von Humboldt Foundation.

LITERATURE CITED

- Collura RV, Stewart C-B. Insertions and duplications of mtDNA in the nuclear genomes of Old World monkeys and hominoids. *Nature* 1995;378:485-489.
- Deutsch D. Structure and function of enamel gene product. *Anat. Rec.* 1989;224:189-210.
- Gascuel O. BIONJ: An improved version of the NJ algorithm based on a simple model of sequence data. *Mol. Biol. Evol.* 1997;14:685-695.
- Girondot M, Sire J-Y. Evolution of the amelogenin gene in toothed and tooth-less vertebrates. *Eur. J. Oral Biol.* 1998;106(Suppl. 1):501-508.
- Huang W, Chang BH-J, Hewett-Emmett D, Li W-H. Sex differences in mutation rate in higher primates estimated from AMG intron sequences. *J. Mol. Evol.* 1997;44:463-465.
- Janvier P. Patterns of diversity in the skull of jawless fishes. In: Hanken J, Hall BK, eds. *The Skull*. Chicago: The University of Chicago Press, 1993: 131-188.
- Janvier P. *Early Vertebrates*. Oxford: Clarendon Press, 1996.
- Page RDM. TreeView: An application to display phylogenetic trees on personal computers. *CABIOS* 1996;12:357-358.
- Saitou N, Nei M. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Mol. Biol. Evol.* 1987;4:406-425.
- Slavkin HC, Diekwisch T. Evolution in tooth developmental biology: Of morphology and molecules. *Anat. Rec.* 1996;245:131-150.
- Slavkin HC, Zeichner-David M, Ferguson MWJ, Termine JD, Graham E, MacDougall M, Bringas P Jr, Bessem C, Grodin M. Phylogenetic and immunogenetic aspects of enamel proteins. In: Riviere GR, Hildemann WH, eds. *Oral Immunogenetic Aspects of Enamel Proteins*. New York: Elsevier, 1982: 241-251.
- Slavkin HC, Graham EE, Zeichner-David M, Hildemann W. Enamel-like antigens in hagfish; possible evolutionary significance. *Evolution* 1983;37:404-412.
- Slavkin HC, Krejsa RJ, Fincham AG, Bringas P Jr, Santos V, Sasano Y, Snead ML, Zeichner-David M. Evolution of enamel proteins: A paradigm for mechanisms of biomineralization. In: Suga S, ed. *Mechanisms and Phylogeny of Mineralisation in Biological Systems*. Tokyo: Springer-Verlag, 1991: 383-389.
- Swofford DL. PAUP: Phylogenetic Analysis Using Parsimony, Vers. 3.1.1. Washington, DC: Smithsonian Institution, 1993.
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, positions-specific gap penalties and weight matrix choice. *Nucleic Acid Res.* 1994;22:4673-4680.
- Wang HL, Yan ZY, Jin DY. Reanalysis of published DNA sequence amplified from cretaceous dinosaur egg fossil. *Mol. Biol. Evol.* 1997;14:589-591.