Mineralized tissue and vertebrate evolution: The secretory calcium-binding phosphoprotein gene cluster

Kazuhiko Kawasaki and Kenneth M. Weiss*

Department of Anthropology, 409 Carpenter Building, Pennsylvania State University, University Park, PA 16802

Gene duplication creates evolutionary novelties by using older tools in new ways. We have identified evidence that the genes for enamel matrix proteins (EMPs), milk caseins, and salivary proteins comprise a family descended from a common ancestor by tandem gene duplication. These genes remain linked, except for one EMP gene, amelogenin. These genes show common structural features and are expressed in ontogenetically similar tissues. Many of these genes encode secretory Ca-binding phosphoproteins, which regulate the Ca-phosphate concentration of the extracellular environment. By exploiting this fundamental property, these genes have subsequently diversified to serve specialized adaptive functions. Casein makes milk supersaturated with Ca-phosphate, which was critical to the successive mammalian divergence. The innovation of enamel led to mineralized feeding apparatus, which enabled active predation of early vertebrates. The EMP genes comprise a subfamily not identified previously. A set of genes for dentine and bone extracellular matrix proteins constitutes an additional cluster distal to the EMP gene cluster, with similar structural features to EMP genes. The duplication and diversification of the primordial genes for enamel/dentine/bone extracellular matrix may have been important in core vertebrate feeding adaptations, the mineralized skeleton, the evolution of saliva, and, eventually, lactation. The order of duplication events may help delineate early events in mineralized skeletal formation, which is a major characteristic of vertebrates.

New biological function arises from new genes produced by gene duplication (1). Even highly diverged proteins today have been generated from older tools. Enamel matrix proteins (EMPs), caseins, and some salivary proteins are secretory Ca-binding phosphoproteins (SCPPs), secreted from the epithelium-derived tissues formed by epithelium-mesenchymal interactions (2, 3). If duplicated genes such as SCPP genes retain common functional and sequence features even after extensive divergence, we infer that the genes were generated from a common ancestor with a related function.

EMPs are responsible for organizing hydroxyapatite crystallization in the enamel organ and are principally coded by three genes: amelogenin (*AMEL*), ameloblastin (*AMBN*), and enamelin (*ENAM*). The evolutionary relationship among the EMP genes was unknown previously; no paralogous genes for any one of these genes were identified (4).

Milk caseins provide mammalian infants with Ca-phosphate to help bone and tooth development as well as to provide required amino acids. Bovine caseins are coded by four distinct genes: αS1 (*CSN1S1*), αS2 (*CSN1S2*), β (*CSN2*), and κ (*CSN10*) (5). The α- and β-caseins are termed Ca-sensitive, because they precipitate in the presence of Ca ions but are stabilized in colloidal suspension by their interaction with the Ca-insensitive κ-casein (6). Ca-sensitive casein genes (*CSN*) arose from a common origin by gene duplication (7, 8).

Statherin and histatins are salivary proteins that protect teeth by regulating the spontaneous precipitation of Ca-phosphate salts on enamel surface (9). Histatins also show antibacterial and antifungal properties (10). These proteins are coded by three distinct genes, *STATH*, *HTN1*, and *HTN3*, which originated from a common origin (11).

We have found that these SCPP genes including two EMPs (*AMBN* and *ENAM*), four caseins, and three salivary proteins as well as four other salivary proteins [three proline-rich proteins (*PROL1*, *-3*, and *-5*) and a mucin (*MUC7*)] all are clustered on human chromosome 4q13, whereas two distinct *AMEL*s are located in nonpseudoautosomal regions of the X and Y chromosomes (12–16). We investigated the origin and functional divergence of these SCPP genes in vertebrate evolution.

Methods

Nucleotide Sequences. The following nucleotide sequences were obtained from GenBank: human casein αS1 (NM001890), β (AF027807, M86237), κ (NM005212), *STATH* (NM003154), *HTN1* (NM002159), *HTN3* (NM000200), *FLJ20513* (AK000520), *NYD-SP26* (AF380838), *PROL5* (D89501), *PROL3* (NM006685), *PROL1* (NM021225), *MUC7* (L13283), *AMBN* (AF209780), *ENAM* (NM031889), *AMEL* (AF436849 and M86933), and *SPARCL1* (secreted protein, acidic, cysteine-rich related, X86693); *DSPP* (dentin sialophosphoprotein, AF163151); *DMP1* (dentine matrix acidic phosphoprotein 1, U89012); *IBSP* (integrin-binding sialoprotein, NM004967); *MEPE* (matrix, extracellular, phosphoglycoprotein, AJ276396); *SPP1* (secreted phosphoprotein 1, BC022844); cow (M16644) and pig (X54975) *CSN1S2*; caiman (AY043290) *AMBN*; pig *ENAM* (U52196); alligator (AF095568) and African clawed toad (AF095569) *AMEL*; and human genome sequence for 4q13 (NT006216) and 4q21 (NT006204).

Computer Analysis. We used DOTTER, a dot-matrix analysis program (17) to determine exon–intron boundaries by searching identical sequences between cDNA and genome sequences. Multiple sequence alignments were generated by using CLUST-ALW (18) through the DNA Data Bank of Japan web site (www.ddbj.nig.ac.jp) and manually modified. Phylogenetic and molecular evolutionary analyses were conducted by using MEGA 2.1 (19). Both SIGNALP (www.cbs.dtu.dk) and PSORT II (http://psort.nibb.ac.jp) were used to predict cleavage sites and the tripartite structures within signal peptides (SPs) (20, 21).

Results

A Gene Cluster for EMP, Casein, and Salivary Proteins. We localized 12 functional genes for the secretory proteins for milk, saliva (and/or tear), and enamel within a cluster spanning >776 kb on 4q13 in humans (Fig. 1). Five nonfunctional pseudogenes related to these proteins (*CSN1S2L1*, *CSN1S2L2*, *STATHL1*, *STATHL2*, and *HTNL*) and two genes with unknown functions (*FLJ20513* and *NYD-SP26*) were also identified. *CSN1S2L2* had a termination codon within exon 4 but retained all canonical
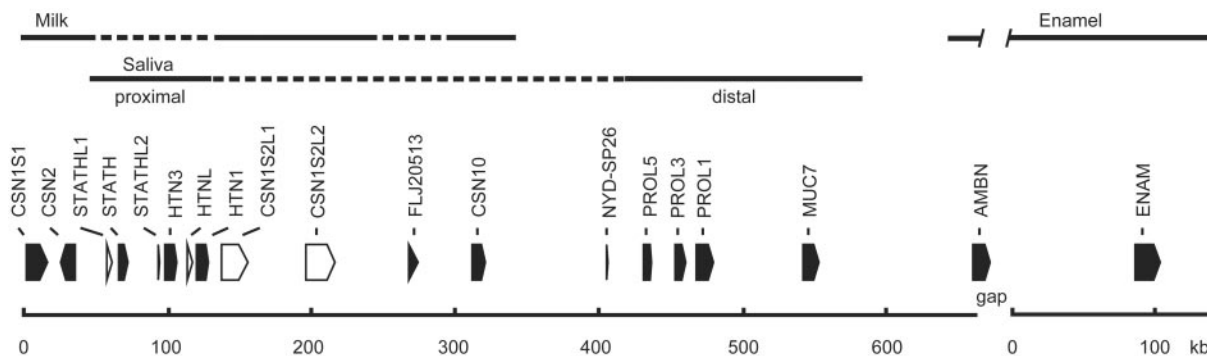

Fig. 1. Casein, salivary, and EMP gene cluster on human 4q13. Locations of the gene products (milk, saliva, and enamel) are shown on the top. Two salivary protein gene complexes, proximal and distal, are shown. Gene symbols, locations, and transcriptional polarities are indicated at the bottom. Filled and open boxes represent the locations of functional genes and pseudogenes, respectively. The scale represents sequence contigs. The 3′ half of *AMBN* is not included in the human draft genome sequence (gap). The length of *AMBN* was inferred from the mouse gene.

splice sites, suggesting recent loss of function. Bovine *CSN*s are arranged in the same order and polarity as in humans, but *CSN1S2* is functional (22). The proximal salivary-protein gene complex (*STATH/HTN3/HTN1*) is located within the Ca-sensitive *CSN* complex and is separated from the distal complex (*PROL5*, -*3*, -*1*, and *MUC7*). *AMBN* is located 109 kb distal to *MUC7* but the 3′ half of this gene resides within a clone gap in the human draft genome sequence (23). The organization of this gene cluster is conserved in the mouse.

Structure of the EMP Genes. The introns of the three EMP genes are all phase-0, that is, their boundaries do not disrupt codons (Fig. 2*A*). A recent study has revised the translation initiation site of mammalian *AMBN* (24). As a result, *AMEL* (both X and Y chromosomes) and *AMBN* show similar gene structure: both exon 1 and the 5′ end of exon 2 (12 or 15 bp) constitute 5′ UTR; exon 2 also codes 16 aa of SP and 2 aa of the N-terminal mature protein (25).

The translation initiation site has not been determined experimentally for *ENAM*, but we found an authentic SP sequence of 16 aa consisting of a canonical tripartite structure: short, positively charged n-region, a central hydrophobic h-region, and a more polar c-region with a cleavage site (26, 27). Mouse *Enam* has an additional 5′-untranslated exon, which was detected within the human first intron but has not been identified in the transcripts (28). If these two exons corresponding to mouse exons 1 and 2 or exons 2 and 3 were inserted secondarily, the initial *ENAM* consisted of a single untranslated exon, and the entire 5′ structure was exactly the same as *AMBN* and *AMEL* (Fig. 2*A*).

Sequence Similarities Among EMPs. Database searches and attempts by hybridization-based library screening failed to detect related sequences for any one of the three EMP genes (29). However, through analysis of the structure of these genes, we now can show that they all are in a single gene family. Large numbers of identical or similar residues in all these proteins were detected in the N-terminal sequences, at least up to the intermediate regions coded by exon 4 (see Fig. 5, which is published as supporting information on the PNAS web site, www.pnas.org). Seven residues completely conserved across known species were identified. Among them, the SXE motif (Ser-Xaa-Glu, where Xaa represents any residue) coded by the 3′ end of exon 3 is a putative phosphorylation site (refs. 30–32; Fig. 2*C*).

A phylogenetic tree was constructed (Fig. 3*A*). The highest sequence similarity was obtained between human *AMBN* and *ENAM*: 36.2% identity and 68.1% similarity for 47 aa. This suggests that the first gene duplication generated the *AMEL* and the *AMBN/ENAM* lineages. *AMEL* was translocated away at this

time or later. We assume a primordial EMP gene that shows the following features: exon 1 and the 5′ end of exon 2 (12–15 bp) constitute 5′ UTR; exon 2 also codes SP (16 aa) and the N-terminal mature protein (2 aa); and exon 3 codes an SXE motif at the 3′ end. In addition, the introns are exclusively phase-0 (Fig. 2*C*).

Structural Similarity Between EMP Genes and Other Genes in the Cluster. In the casein and the salivary protein genes in the cluster, all but one intron in *CSN1S1* are phase-0 (Fig. 2*A*). The last intron of *CSN1S1* located within the termination codon is phase-2, which is conserved in cow (33). This phase-2 intron was probably generated by a mutation that abolished the termination codon and elongated the protein-coding region, because the last exon of *CSN1S1* has sequence homology to the last exons of *CSN1S2* and *CSN2*.

Each of *CSN1S2*, *CSN2*, *STATH*, and *HTN1* has an SXE phosphorylation motif coded by the end of exon 3 (11, 34, 35). An SEE motif was found in *CSN1S1* but is coded by the end of exon 4 (33). The phosphorylated residue in casein directly associates with a Ca ion and is responsible for the Ca-binding property (6). The SHE motif in *HTN1* is shifted upstream by 1 aa, and the corresponding sequence has been substituted to SHA in *HTN3* by a point mutation (36).

All genes for casein and salivary protein in the cluster show similar 5′ structure, although the C terminus of the SP spans to exon 3 in *CSN10* and all salivary protein genes (Fig. 2*A*). *CSN1S2* and *CSN2* show especially high similarity to the primordial EMP gene: exon 1 and 5′-exon 2 (12 bp) constitute 5′ UTR, exon 2 encodes SP (15 aa) and the N-terminal mature protein (2 aa), and exon 3 codes an SXE motif at the 3′ end (8). In addition, their SPs show high sequence similarities to the EMPs in the n- and h-regions. Thus, we conclude that the ancestral Ca-sensitive *CSN* was derived from one of the three EMP genes, probably either from *ENAM* or *AMBN* inferred from their chromosomal locations.

Phylogenetic Analysis of Caseins, Statherin, and Histatins. The Ca-sensitive casein genes and *STATH/HTN* share an additional feature; the last exon consists of only 3′ UTR except for *CSN1S1* as described above (Fig. 2*A*). Comparison of the last exons revealed as high as 59.4% sequence homology for 251 bp between bovine *CSN1S2* and human *HTN1*, although the protein-coding regions of these genes showed no sequence similarity except in the N-terminal half of the SPs. Lower sequence homology in coding regions rather than intergenic regions has been observed between *STATH* and *HTN1* (11). A phylogenetic tree based on sequences of the last exon shows
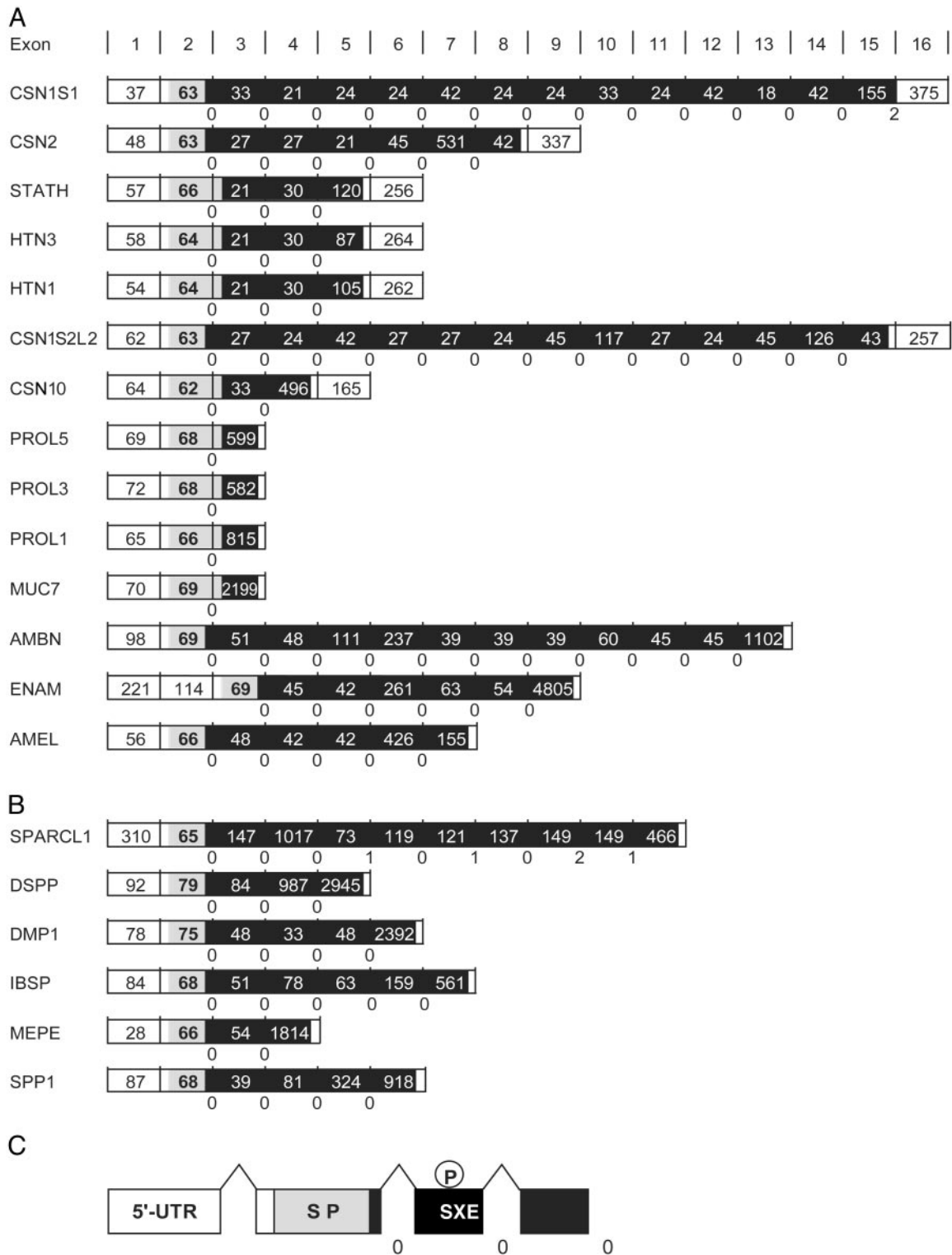
Fig. 2. Structures and intron phases. White, gray, and black regions represent UTR, SP, and the mature protein, respectively. (*A*) Exon–intron structures of EMP, casein, and salivary protein genes on 4q13 and *AMEL* on Xp22 are shown. The length of each exon (bp) is shown in the boxes. The phases of introns are indicated at the bottom of exon boundaries. *CSN1S2L2* is a pseudogene, but the exon–intron boundaries were determined unambiguously by comparing bovine and porcine *CSN1S2* sequences with the human genome sequence. (*B*) Exon–intron structures of *SPARCL1* and dentine/bone ECM protein genes on 4q21 are shown. The structure of *SPARC* is the same as *SPARCL1* except that exon 4 is missing in the 5′ region of *SPARC*. (*C*) The structure of the first four exons of the primordial EMP gene is shown. A protein kinase phosphorylates the Ser residue (P) in the SXE motif coded by the 3′ end of exon 3. The introns are exclusively phase-0.

that *STATH/HTN* arose from *CSN1S2* (Fig. 3*B*). This result is supported by a high bootstrap value and is consistent even if *CSN* sequences from other species were added. This is also supported by the order and polarity of these genes (Fig. 1). The close relationship between *CSN1S2* and *CSN2* corroborates the previous observation based on the 5′-flanking sequences (7).
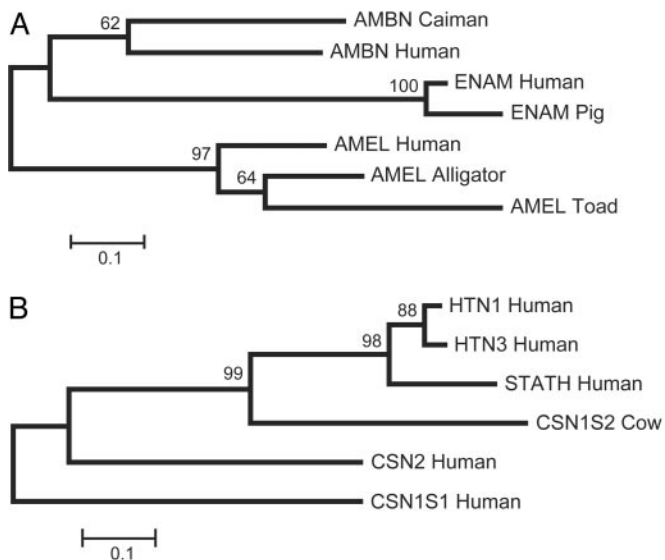
Fig. 3. Phylogenetic trees for EMPs and casein-salivary proteins. MEGA II was used to construct gene trees based on the neighbor-joining method and to calculate substitution rates and bootstrap values. Both trees were drawn based on the timing of the inferred first gene duplication. (*A*) A phylogenetic tree for the three EMPs was constructed based on the amino acid sequences coded by exons 2–4, which are conserved among all EMP genes. The topology was the same when using sequences coded by exons 2 and 3. (*B*) A phylogenetic tree for Ca-sensitive *CSN* and three salivary protein genes was constructed based on the nucleotide sequences of the last exons (3′ UTR): exon 16 of *CSN1S1*, exon 18 of bovine *CSN1S2* (human has two pseudogenes), exon 9 of *CSN2*, and exon 6 of *STATH/HTNs*. The same topology was obtained when using human *CSN1S2L2*. No ubiquitous repeat sequences were identified in these regions.

Dentine and Bone Extracellular Matrix (ECM) Proteins. In a 375-kb region on 4q21, ≈15 Mb distal to *ENAM*, there is the small integrin-binding ligand, n-linked glycoprotein (SIBLING) gene cluster for dentine and bone ECM proteins: *DSPP*, *DMP1*, *IBSP*, *MEPE*, and *SPP1* (37). All these genes have the same transcriptional polarity and the structural features common to the primordial EMP genes (Fig. 2*B*). Their SPs show sequence similarity to the EMP genes. In addition, each of *SPP1*, *DMP1*, and *IBSP* codes an SXE motif and two following amino acids at the 3′ end of exon 3. Indeed, the same protein kinase phosphorylates Ser residues of casein and SPP1 (38). These data suggest that all these dentine/bone ECM protein genes and the primordial EMP gene arose from a common ancestor by gene duplication and comprise the SCPP gene family.

Discussion

Early in the vertebrate lineage, genes for transcription factors and signaling molecules were duplicated extensively (39). Epithelial–mesenchymal interactions facilitated these partly redundant tools for spatial and temporal regulations of organogenesis, which eventually produced various tissues such as teeth, fish scales, hair, feathers, mammary glands, and salivary glands (3, 40). Duplications of terminal differentiation genes also contributed to the generation of these tissues. By exploiting SP and SXE motifs, SCPPs regulate the Ca-phosphate concentration of the extracellular environment. Crystallization of the extracellular Ca-phosphate eventually created tooth, a mineralized hard tissue. By using common regulatory mechanisms, other SCPP genes could be expressed in ontogenetically similar but distinct tissues and contributed to the functions of mammary glands and salivary glands.

The introns for these SCPP genes are exclusively phase-0. Such nonrandom distribution of intron phases is mainly due to exon duplications (41). Exon duplication is common in Ca-sensitive *CSN* and contributed considerably to increasing the capacity of Ca-phosphate transport in milk (8). Our observations support this model; an ancient duplication of exon 3 coding an SXE motif generated the major phosphorylation sites. Exon duplications were also detected in exons 7–9 of human *AMBN* (42). Exon skipping is common in both casein and EMP genes (2, 43). This suggests that exon deletions also contributed to the current structure of these genes. Due to recurrent exon duplications and deletions, the number of exon is different among these genes (Fig. 2*A*).

CSN10 and four salivary genes in the distal complex show several similarities with the SCPP genes (Fig. 2*A*). Previously, amino acid sequence homology between *CSN10* and fibrinogen γ-polypeptide (*FGG*) has been reported (44). However, the "homologous" regions are coded by six different exons in *FGG*, whereas these are solely coded by exon 4 in *CSN10*. We found that *CSN10* has sequence similarity to the mouse mucin 10 gene (*Muc10*) in 5′ UTR (exon 1), SP, and portions of exon 3 (45). *Muc10* is located in the distal salivary gene complex and has the same structural features with genes in the complex, whereas no *MUC10* was identified in human. *PROL1*, *-3*, and *-5* have sequence homologies to one another (46, 47) and show similar structure to *MUC7* (Fig. 2*A*). We speculate that both Ca-sensitive and -insensitive *CSN*s and salivary protein genes in both proximal and distal complexes belong to the SCPP gene family.

Phase-0 introns are the most abundant in nature and account for 43% of all introns (41). However, genes exclusively consisting of phase-0 introns are uncommon. We searched intron phases of a number of other genes: whey proteins (α-lactalbumin, lysozyme, β-lactoglobulin, and lactotransferrin), salivary proteins (mucin 5B, proline-rich proteins located on 1q21 and 12p13, *DMBT1*, and cystatins), enamel proteins (tuftelin, matrix metalloproteinase 20, and kallikrein 4), two unknown genes within the same gene cluster, Ca-binding protein genes, and genes involved in bone formation, but none of these genes exclusively consisted of phase-0 introns. Furthermore, none of them showed the features common to the 5′ portion of the SCPP genes. Thus, no other SCPP genes were identified elsewhere in the genome outside of the cluster on chromosome 4 with the exception of *AMEL*s.

It has been suggested that milk emerged as a cutaneous secretion that protects eggs from microorganisms and subsequently shifted to a directly nutritional product in mammal-like reptiles (48). Lactation enabled immature birth of altricial infants, and shortened the time during which mothers had to carry large fetuses. This led to the key characteristic adaptation of mammals (49). Casein made milk supersaturated with Ca-phosphate. Because available tools could produce carbohydrates, proteins, and lipids, the development of *CSN* may be critical to the mammalian adaptation. Indeed, *CSN* has been identified exclusively from mammals. The primordial Ca-sensitive *CSN* probably diverged either from *ENAM* or *AMBN* before the appearance of monotremes in the Jurassic (Fig. 4). Marsupials have two Ca-sensitive caseins, α and β (50, 51). Comparison of the 3′-UTR sequences revealed a close relationship between *CSNα* and *CSN1S1*, whereas *CSNβ* showed the highest homology to marsupial *CSNα*, not to eutherian *CSN*s. *CSNβ* may be specific to marsupials. Recurrent gene duplications of Ca-sensitive *CSN* imply the significance of the Ca-phosphate requirement for the postnatal growth in mammals.

As α-lactalbumin arose from lysozyme, gene duplications developed more adaptive milk (52). Likewise, *STATH/HTN* arose from *CSN1S2* and developed more adaptive saliva. Statherin and histatin have been identified only from primates, suggesting the recent origin of these salivary proteins (11). We
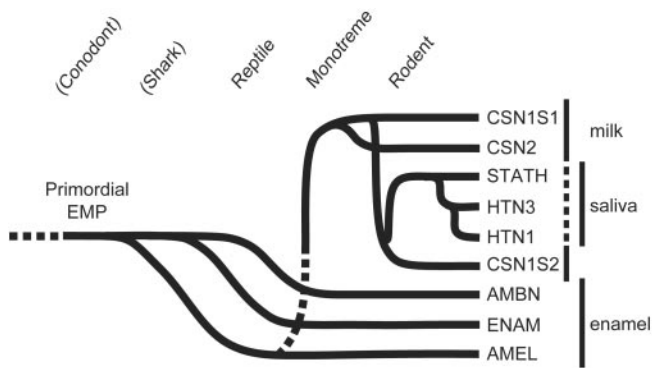
Fig. 4. Divergence of the primordial EMP gene. The phylogeny of Ca-sensitive *CSN*s (milk), *STATH*/*HTN* (saliva), and EMP genes (enamel) was based on the two phylogenetic trees in Fig. 3. The primordial EMP gene appeared early in vertebrate evolution, perhaps in conodont (parenthesized). Immunohistochemical studies suggested the possibility that *AMEL* and *ENAM* emerged before the divergence of shark (parenthesized). *AMBN* diverged before caiman (reptiles). The primordial Ca-sensitive *CSN* diverged from one of the EMP genes, probably either from *ENAM* or *AMBN* before the emergence of monotremes. *STATH* then diverged from *CSN1S2* before the divergence of rodents. Except for *AMEL*, genes are arranged in this order on chromosome 4.

identified *Stath* but not *Htn* pseudogenes in the mouse genome at the locus corresponding to human. This indicates that *STATH* arose from *CSN1S2* at least before the divergence of rodents 96 million years ago (53), and *HTN* subsequently descended (Fig. 4). A previous study estimated the *STATH-HTN* duplication date at 40–50 million years ago and the subsequent *HTN1-HTN3* duplication date at 15–30 million years ago based on the nucleotide substitution rates (11). Rapid sequence divergence of *HTN* suggests the functional divergence from *STATH*. Indeed, *HTN3* lost the SXE motif and cannot work as a Ca-binding phosphoprotein but developed novel antimicroorganism properties (10).

The transition from protochordates to vertebrates was associated with a shift from a passive to an active mode of predation (54). Mineralized feeding apparatus such as oropharyngeal denticles and teeth enabled food apprehension; hence this innovation was critical to the adaptive vertebrate lineage (55). The extant agnathans, hagfish and lamprey that diverged early from the stem craniate, have no mineralized exoskeleton. Conodonts (naked agnathan in the Late Cambrian; 510 million years ago) seem to have developed the earliest mineralized exoskeleton as an oral feeding apparatus. The conodont elements are almost entirely composed of enamel (lamellar crown), and dentine (basal body) underlays the crown in at least some species (56–58), whereas dentine is more common in dermal exoskeleton of ostracoderm (59). Thus, the earliest vertebrate history of the mineralized exoskeletal formation is controversial. We speculate that the primordial EMP is one of the most likely components used in the earliest enamel, conodont elements (Fig. 4). Replaceable sets of specialized teeth on the mandibular arch are confined to jawed vertebrates (60). Antibodies against mouse amelogenin and pig enamelin detected distinct shark (*Heterodontus francisci*) antigens (61). This suggests the possibility that *AMEL* and *ENAM* had diverged by the emergence of chondrichthyans in the Early Silurian (62). However, these EMP genes need to be isolated from chondrichthyans to identify this divergence (Fig. 4). The three EMP genes were fully established by the Carboniferous, as shown by the isolation of *AMBN* from the caiman (24).

The Ordovician agnathans such as *Astraspidae* developed dermal exoskeleton consisting of enamel, dentine, and bone (aspidin) (63). Today, these tissues interact in the development

of tooth, each cooperatively leading to different tissue layers. Functionally and phylogenetically related genes need not retain close linkage alignment, but highly conserved linkage may be an indicator of current function or phylogenetic history. Novel functions can be developed relatively easily in ontogenetically similar tissues, because it requires small modifications to the original regulatory elements. Genes for enamel, dentine, and bone ECM protein genes arose from a common ancestor. We speculate that their tandem linkage arrangement facilitated a common spatial and temporal control of these paralogous genes in embryogenesis, that is, coordinately expressed in focal denticle germs (tubercles or scales on dermal exoskeleton). This supports a scenario in which bone initially appeared as attachment bone closely associated with dentine in scales, and later a continuous sheet such as a head shield emerged. The long-hypothesized relationship between teeth as feeding apparatus and ancestral exoskeletal scales is reinforced by the nature of these genes and their arrangement (55, 64, 65), and additional salivary function reinforces the idea of common usage of related genes for feeding function in vertebrate evolution. Lactation function came later but is evolutionarily related.

Interestingly, enamel and dentine are produced by different, interacting, and juxtaposed cell types (3). The colocalization of enamel and dentine ECM protein genes may also be related to the dental patterning mechanism by which dental epithelium and mesenchyme communicate, but nothing is as yet known about this aspect of the regulation of these closely syntenic genes expressed in a coordinated way in the two intercommunicating tissues. Similarly, it is not known whether variation in these genes is related to variation in dental or related patterning (L. J. Hlusko, J. Rogers, M. C. Mahaney, K.K., and K.M.W., unpublished data).

SPARC is located on 5q31.3-q32 and encodes a major noncollagenous protein of bone matrix (66, 67). We found that *SPARCL1* (*SPARC*-related gene) is located 79 kb proximal to the dentine and bone ECM protein gene cluster (68). In addition, exons 1 and 2 of both *SPARC* and *SPARCL1* show similar structural features to the other SCPP genes (Fig. 2*B*). Previously, nucleotide sequence homology was observed between exon 2 of *AMEL* and *SPARC/SPARCL1* (29). However, this study did not consider the limits of sequence variation in the SP and a subsequent amino acid residue (26). Although *SPARCL1* seems to be of recent origin, *SPARC* is identified in nematode and fruit fly (69). By exon shuffling, *SPARC* obtained a Follistatin-like domain (70), which is not seen in the other dental (enamel/dentine) and bone ECM protein genes in the cluster. *SPARC* also encodes a Ca-binding protein but does not have an SXE motif. Nevertheless, *SPARC* is the only gene that has been suggested as having a possible evolutionary relationship to the ECM protein genes in the cluster and that has been identified in protostome. The 5′ exons of *SPARC* may have originated the primordial dental/bone ECM protein gene.

We have hypothesized that the dental/bone ECM protein genes on human 4q13-q21 were created from a single ancestor by gene duplication. The emergence of enamel, dentine, and bone ECM protein genes would suggest when each hard tissue appeared. The sequence divergence of the duplicated genes may reflect complexity of the mineralized tissues and the complex nature of their appearance in various vertebrate lineages. Thus, analysis of the chromosomal loci containing these genes of agnathans and primitive gnathostomes may delineate early events in mineralized skeletal formation, which is a major characteristic of vertebrates.

1. Ohno, S. (1970) *Evolution by Gene Duplication* (Springer, New York).
2. Zeichner-David, M., Diekwisch, T., Fincham, A., Lau, E., MacDougall, M., Moradian-Oldak, J., Simmer, J., Snead, M. & Slavkin, H. C. (1995) *Int. J. Dev. Biol.* **39,** 69–92.
3. Thesleff, I. & Sharpe, P. (1997) *Mech. Dev.* **67,** 111–123.
4. Fincham, A. G., Moradian-Oldak, J. & Simmer, J. P. (1999) *J. Struct. Biol.* **126,** 270–299.
5. Ginger, M. R. & Grigor, M. R. (1999) *Comp. Biochem. Physiol. B Biochem. Mol. Biol.* **124,** 133–145.
6. Farrell, H. M., Jr., & Thompson, M. P. (1988) in *Calcium-Binding Proteins*, ed. Thompson, M. P. (CRC, Boca Raton, FL), Vol. II, pp. 117–137.
7. Yu-Lee, L. Y., Richter-Mann, L., Couch, C. H., Stewart, A. F., Mackinlay, A. G. & Rosen, J. M. (1986) *Nucleic Acids Res.* **14,** 1883–1902.
8. Rosen, J. M. (1987) in *The Mammary Gland*, eds. Neville, M. C. & Daniel, C. W. (Plenum, New York), pp. 301–322.
9. Hay, D. I. & Moreno, E. C. (1989) in *Human Saliva: Clinical Chemistry and Microbiology*, ed. Tenovuo, J. O. (CRC, Boca Raton, FL), Vol. I, pp. 131–150.
10. Amerongen, A. V. & Veerman, E. C. (2002) *Oral Dis.* **8,** 12–22.
11. Sabatini, L. M., Ota, T. & Azen, E. A. (1993) *Mol. Biol. Evol.* **10,** 497–511.
12. Sabatini, L. M. & Azen, E. A. (1989) *Biochem. Biophys. Res. Commun.* **160,** 495–502.
13. vanderSpek, J. C., Wyandt, H. E., Skare, J. C., Milunsky, A., Oppenheim, F. G. & Troxler, R. F. (1989) *Am. J. Hum. Genet.* **45,** 381–387.
14. Lau, E. C., Mohandas, T. K., Shapiro, L. J., Slavkin, H. C. & Snead, M. L. (1989) *Genomics* **4,** 162–168.
15. Rijnkels, M., Meershoek, E., de Boer, H. A. & Pieper, F. R. (1997) *Mamm. Genome* **8,** 285–286.
16. Dong, J., Gu, T. T., Simmons, D. & MacDougall, M. (2000) *Eur. J. Oral Sci.* **108,** 353–358.
17. Sonnhammer, E. L. & Durbin, R. (1995) *Gene* **167,** GC1–GC10.
18. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994) *Nucleic Acids Res.* **22,** 4673–4680.
19. Kumar, S., Tamura, K., Jakobsen, I. B. & Nei, M. (2001) *Bioinformatics* **17,** 1244–1245.
20. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997) *Int. J. Neural Syst.* **8,** 581–599.
21. Nakai, K. & Horton, P. (1999) *Trends Biochem. Sci.* **24,** 34–36.
22. Rijnkels, M., Kooiman, P. M., de Boer, H. A. & Pieper, F. R. (1997) *Mamm. Genome* **8,** 148–152.
23. Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* (2001) *Nature* **409,** 860–921.
24. Shintani, S., Kobata, M., Toyosawa, S., Fujiwara, T., Sato, A. & Ooshima, T. (2002) *Gene* **283,** 245–254.
25. Salido, E. C., Yen, P. H., Koprivnikar, K., Yu, L. C. & Shapiro, L. J. (1992) *Am. J. Hum. Genet.* **50,** 303–316.
26. von Heijne, G. (1985) *J. Mol. Biol.* **184,** 99–105.
27. Hu, C. C., Hart, T. C., Dupont, B. R., Chen, J. J., Sun, X., Qian, Q., Zhang, C. H., Jiang, H., Mattern, V. L., Wright, J. T. & Simmer, J. P. (2000) *J. Dent. Res.* **79,** 912–919.
28. Hu, J. C., Zhang, C. H., Yang, Y., Karrman-Mardh, C., Forsman-Semb, K. & Simmer, J. P. (2001) *J. Dent. Res.* **80,** 898–902.
29. Delgado, S., Casane, D., Bonnaud, L., Laurin, M., Sire, J. Y. & Girondot, M. (2001) *Mol. Biol. Evol.* **18,** 2146–2153.
30. Fukae, M., Tanabe, T., Murakami, C., Dohi, N., Uchida, T. & Shimizu, M. (1996) *Adv. Dent. Res.* **10,** 111–118.
31. Salih, E., Huang, J. C., Strawich, E., Gouverneur, M. & Glimcher, M. J. (1998) *Connect. Tissue Res.* **38,** 225–235.
32. Salih, E., Huang, J. C., Strawich, E., Gouverneur, M. & Glimcher, M. J. (1998) *Connect. Tissue Res.* **38,** 241–246.
33. Koczan, D., Hobom, G. & Seyfert, H. M. (1991) *Nucleic Acids Res.* **19,** 5591–5596.
34. Jones, W. K., Yu-Lee, L. Y., Clift, S. M., Brown, T. L. & Rosen, J. M. (1985) *J. Biol. Chem.* **260,** 7042–7050.
35. Groenen, M. A., Dijkhof, R. J., Verstege, A. J. & van der Poel, J. J. (1993) *Gene* **123,** 187–193.
36. Sabatini, L. M., Warner, T. F., Saitoh, E. & Azen, E. A. (1989) *J. Dent. Res.* **68,** 1138–1145.
37. Fisher, L. W., Torchia, D. A., Fohr, B., Young, M. F. & Fedarko, N. S. (2001) *Biochem. Biophys. Res. Commun.* **280,** 460–465.
38. Lasa, M., Chang, P. L., Prince, C. W. & Pinna, L. A. (1997) *Biochem. Biophys. Res. Commun.* **240,** 602–605.
39. Holland, P. W. (1999) *Semin. Cell Dev. Biol.* **10,** 541–547.
40. Krejsa, R. J. (1979) in *Hyman's Comparative Vertebrate Anatomy*, ed. Wake, M. H. (Univ. of Chicago Press, Chicago), pp. 112–191.
41. Fedorov, A., Fedorova, L., Starshenko, V., Filatov, V. & Grigor'ev, E. (1998) *J. Mol. Evol.* **46,** 263–271.
42. Toyosawa, S., Fujiwara, T., Ooshima, T., Shintani, S., Sato, A., Ogawa, Y., Sobue, S. & Ijuhin, N. (2000) *Gene* **256,** 1–11.
43. Menon, R. S., Chang, Y. F., Jeffers, K. F. & Ham, R. G. (1992) *Genomics* **12,** 13–17.
44. Jolles, P., Loucheux-Lefebvre, M. H. & Henschen, A. (1978) *J. Mol. Evol.* **11,** 271–277.
45. Denny, P. C., Mirels, L. & Denny, P. A. (1996) *Glycobiology* **6,** 43–50.
46. Isemura, S. & Saitoh, E. (1994) *J. Biochem. (Tokyo)* **115,** 1101–1106.
47. Dickinson, D. P. & Thiesse, M. (1996) *Curr. Eye Res.* **15,** 377–386.
48. Blackburn, D. G. (1989) *Mamm. Rev.* **19,** 1–26.
49. Pond, C. M. (1976) *Evolution (Lawrence, Kans.)* **31,** 177–199.
50. Collet, C., Joseph, R. & Nicholas, K. (1992) *J. Mol. Endocrinol.* **8,** 13–20.
51. Ginger, M. R., Piotte, C. P., Otter, D. E. & Grigor, M. R. (1999) *Biochim. Biophys. Acta* **1427,** 92–104.
52. Brew, K., Vanaman, T. C. & Hill, R. L. (1967) *J. Biol. Chem.* **242,** 3747–3749.
53. Nei, M., Xu, P. & Glazko, G. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 2497–2502.
54. Gans, C. & Northcutt, R. G. (1983) *Science* **220,** 268–274.
55. Smith, M. M. & Coates, M. I. (2000) in *Development, Function, and Evolution of Teeth*, ed. Teaford, M. F., Smith, M. M. & Ferguson, M. W. J. (Cambridge Univ. Press, Cambridge, U.K.), pp. 133–151.
56. Sansom, I. J. (1996) *Zool. J. Linn. Soc.* **118,** 47–57.
57. Smith, M. M., Sansom, I. J. & Smith, M. P. (1996) *Modern Geology* **20,** 303–319.
58. Donoghue, P. C. J. & Aldridge, R. J. (2001) in *Major Events in Vertebrate Evolution*, ed. Ahlberg, P. E. (Taylor & Francis, London), pp. 85–105.
59. Forey, P. & Janvier, P. (1993) *Nature* **361,** 129–134.
60. Smith, M. M. & Coates, M. I. (2001) in *Major Events in Early Vertebrate Evolution*, ed. Ahlberg, P. E. (Taylor & Francis, London), pp. 223–240.
61. Satchell, P. G., Anderton, X., Ryu, O. H., Luan, X., Ortega, A. J., Opamen, R., Berman, B. J., Witherspoon, D. E., Gutmann, J. L., Yamane, A., *et al.* (2002) *J. Exp. Zool.* **294,** 91–106.
62. Sansom, I. J., Smith, M. M. & Smith, M. P. (2001) in *Major Events in Early Vertebrate Evolution*, ed. Ahlberg, P. E. (Taylor & Francis, London), pp. 156–171.
63. Smith, M. M. & Hall, B. K. (1990) *Biol. Rev.* **65,** 277–373.
64. Ørvig, T. (1977) in *Problems in Vertebrate Evolution*, eds. Andrews, S. M., Miles, R. S. & Walker, A. D. (Academic, London), pp. 53–75.
65. Schaeffer, B. (1977) in *Problems in Vertebrate Evolution*, eds. Andrews, S. M., Miles, R. S. & Walker, A. D. (Academic, London), pp. 25–52.
66. Termine, J. D., Kleinman, H. K., Whitson, S. W., Conn, K. M., McGarvey, M. L. & Martin, G. R. (1981) *Cell* **26,** 99–105.
67. Yan, Q. & Sage, E. H. (1999) *J. Histochem. Cytochem.* **47,** 1495–1506.
68. Girard, J. P. & Springer, T. A. (1995) *Immunity* **2,** 113–123.
69. Martinek, N., Zou, R., Berg, M., Sodek, J. & Ringuette, M. (2002) *Dev. Genes Evol.* **212,** 124–133.
70. Patty, L. (1996) in *Human Genome Evolution*, eds. Jackson, M., Strachan, T. & Dover, G. (BIOS Scientific, Oxford), pp. 35–71.

EVOLUTION