Parallel Evolution of Ameloblastic scpp Genes in Bony and Cartilaginous Vertebrates

Nicolas Leurs, Camille Martinand-Mari, Sylvain Marcellini, Mélanie Debiais-Thibaud

Parallel Evolution of Ameloblastic *scpp* Genes in Bony and Cartilaginous Vertebrates

Nicolas Leurs,[1] Camille Martinand-Mari,[1] Sylvain Marcellini,*[,2] and Mélanie Debiais-Thibaud*[,1]

[1]Institut des Sciences de l'Evolution de Montpellier, ISEM, Univ Montpellier, CNRS, IRD, Montpellier, France
[2]Departamento de Biología Celular, Facultad de Ciencias Biológicas, Universidad de Concepción, Concepción, Chile

***Corresponding authors:** E-mails: melanie.debiais-thibaud@umontpellier.fr; smarcellini@udec.cl.

Associate editor: Fabia Ursula Battistuzzi

Abstract

In bony vertebrates, skeletal mineralization relies on the secretory calcium-binding phosphoproteins (Scpp) family whose members are acidic extracellular proteins posttranslationally regulated by the Fam20°C kinase. As *scpp* genes are absent from the elephant shark genome, they are currently thought to be specific to bony fishes (osteichthyans). Here, we report a *scpp* gene present in elasmobranchs (sharks and rays) that evolved from local tandem duplication of *sparc-L* 5′ exons and show that both genes experienced recent gene conversion in sharks. The elasmobranch *scpp* is remarkably similar to the osteichthyan *scpp* members as they share syntenic and gene structure features, code for a conserved signal peptide, tyrosine-rich and aspartate/glutamate-rich regions, and harbor putative Fam20°C phosphorylation sites. In addition, the catshark *scpp* is coexpressed with *sparc-L* and *fam20°C* in tooth and scale ameloblasts, similarly to some osteichthyan *scpp* genes. Despite these strong similarities, molecular clock and phylogenetic data demonstrate that the elasmobranch *scpp* gene originated independently from the osteichthyan *scpp* gene family. Our study reveals convergent events at the *sparc-L* locus in the two sister clades of jawed vertebrates, leading to parallel diversification of the skeletal biomineralization toolkit. The molecular evolution of *sparc-L* and its coexpression with *fam20°C* in catshark ameloblasts provides a unifying genetic basis that suggests that all convergent *scpp* duplicates inherited similar features from their *sparc-L* precursor. This conclusion supports a single origin for the hypermineralized outer odontode layer as produced by an ancestral developmental process performed by Sparc-L, implying the homology of the enamel and enameloid tissues in all vertebrates.

***Key words:* *fam20°C* , *sparc-L*, *scpp*, *Scyliorhinus canicula*, jawed vertebrate evolution, genomic convergence, gene conversion, ameloblasts, enamel, enameloid.**

Introduction

The secretory calcium-binding phosphoproteins (Scpp) family comprises extracellular proteins playing pivotal roles during bony vertebrate (osteichthyan) skeletal mineralization (Kawasaki and Weiss 2003; Kawasaki 2011). The *scpp* genes are organized into genomic clusters coding for acidic proteins involved in enamel, bone, and dentin matrix mineralization (Nikoloudaki 2021; Ustriyana et al. 2021). Bony fish *scpp* genes are grouped in a single family from their shared intron–exon structure rather than from shared sequence features as these have quickly diverged and are mostly nonalignable (Kawasaki and Weiss 2003). Notwithstanding, distinct Scpp members expressed in ameloblasts, osteoblasts, and odontoblasts share a common functional feature based on their posttranslational regulation by Fam20°C kinase phosphorylation (Tagliabracci et al. 2012; Wang et al. 2013; Liu et al. 2017; Zhang et al. 2018; Schytte et al. 2020; Shin et al. 2020).

The evolutionary history of the *scpp* gene family has been the focus of intense research with regard to the evolution of the vertebrate mineralized skeleton (Kawasaki and Weiss 2003; Kawasaki et al. 2005, 2017, 2021; Kawasaki 2009, 2011; Kawasaki and Amemiya 2014; Venkatesh, Lee, Ravi, Maurya, Korzh, et al. 2014; Venkatesh, Lee, Ravi, Maurya, Lian, et al. 2014; Qu et al. 2015; Braasch et al. 2016; Lin et al. 2016; Lv et al. 2017; Cheng et al. 2021). The two rounds of whole genome duplication in the vertebrate lineage produced the *sparc* and *sparc-L* paralogs (Kawasaki and Weiss 2003; Kawasaki et al. 2005; Kawasaki 2009; Bertrand et al. 2013; Enault et al. 2018). In the osteichthyan lineage, *sparc-L* locally duplicated into *sparc-L1* and *sparc-L2* (Bertrand et al. 2013; Enault et al. 2018), and multiple duplications and losses at this locus also produced a broad array of *scpp* members (Kawasaki et al. 2005; Kawasaki 2009, 2011). Some of these duplications occurred concomitantly with the origin of mammals and are associated with evolutionary innovation, as is the case for the milk *Caseins* and the saliva *Muc7* gene (Kawasaki et al. 2011; Xu et al. 2016; Luis Villanueva-Cañas et al. 2017). Genome projects have identified *scpp* genes at the *sparc-L1*/*sparc-L2* locus in all osteichthyan species examined to date, although with poor support of 1-to-1 orthology between teleosts and

Article

tetrapods, reflecting a high rate of sequence divergence and gene turnover (Kawasaki and Amemiya 2014; Qu et al. 2015; Braasch et al. 2016; Lin et al. 2016; Kawasaki et al. 2017; Lv et al. 2017; Cheng et al. 2021).

As careful examination of the elephant shark genome failed to identify any *scpp* gene around *sparc-L* (Venkatesh, Lee, Ravi, Maurya, Lian, et al. 2014), the *scpp* family was considered to be osteichthyan-specific (Venkatesh, Lee, Ravi, Maurya, Korzh, et al. 2014; Venkatesh, Lee, Ravi, Maurya, Lian, et al. 2014; Kawasaki et al. 2017, 2021), implying that chondrichthyans use a default mineralization genetic toolkit merely composed of Sparc and Sparc-L. This argument was also used as a support to the conclusion that chondrichthyan enameloid, the outer hypermineralized tissue of teeth and scales, is not homologous to the osteichthyan enamel (in tetrapods) nor ganoin (in actinopterygians, coelacanth, and lungfish) that critically rely on Scpp proteins for their mineralization (Qu et al. 2015; Kawasaki et al. 2017, 2021).

Chondrichthyans and stem jawed vertebrates, however, exhibit a rich diversity of mineralized skeletal tissues with similarity to those of bony fish (Ørvig 1951; Min and Janvier 1998; Donoghue and Sansom 2002; Donoghue et al. 2006; Johanson et al. 2012; Brazeau and Friedman 2015; Keating et al. 2018; Lemierre and Germain 2019; Brazeau et al. 2020; Berio et al. 2021), which seems at odds with the idea of a Sparc/Sparc-L default toolkit. To address this issue, it remains of utmost importance to better characterize the genetic basis of skeletal mineralization in chondrichthyans. By exploring high-quality genome assemblies, we show here for the first time that elasmobranchs (a chondrichthyan monophyletic group including sharks and rays) have independently evolved a *scpp* gene. Our data reveal that distinct mutational and evolutionary processes at the chondrichthyan *sparc-L* locus mimic many of the key features that are emblematic of osteichthyan *scpp* genes and illustrate how molecular parallelism has continuously enriched the vertebrate skeletal mineralization genetic repertoire. Taken together, our genomic and expression data provide a unifying genetic basis that supports the homology of developmental processes for enamel and enameloid in all jawed vertebrates.

Results

A Novel Elasmobranch *scpp* Gene

We identified a nonannotated gene at the *sparc-L* locus in high-quality genome assemblies of two elasmobranch species, the lesser spotted catshark *Scyliorhinus canicula* and the skate *Amblyraja radiata* (fig. 1). We named this gene *scpp* because it is genomic, structural, and expression features are remarkably similar to the osteichthyan *scpp* genes. The syntenic location of the elasmobranch *scpp* gene and its tail-to-tail orientation with respect to *sparc-L* are equivalent to the osteichthyan acidic *scpp* gene cluster (fig. 1; see Kawasaki 2009; Braasch et al. 2016). In addition, the intron/exon structure of the catshark and skate *scpp* genes is identical to the 5′ intron/exon structure of the nearby *sparc-L* gene, supporting its origin from partial local duplication of the *sparc-L* locus (fig. 1). Hence, the first three introns located in the open reading frame of the elasmobranch *scpp* are in phase 0 (fig. 1), which is the phase most commonly observed in osteichthyan *scpp* genes (fig. 1; see Kawasaki and Weiss 2003). A fourth intron is located in phase 1 in *scpp* and *sparc-L* (fig. 1). We searched genomic and transcriptomic data and identified a similar *scpp* gene in other elasmobranch species belonging to selachimorphs (the sharks *Scyliorhinus torazame*, *Chiloscyllium punctatum*, and *Carcharodon carcharias*) and batoids (the ray *Raja clavata* and the skate *Okamejei kenojei*, supplementary material S1, Supplementary Material online). In all cases, the encoded elasmobranch Scpp protein displays a signal peptide conserved with Sparc, Sparc-L/L1/L2, and the osteichthyan Scpp sequences, together with an enrichment of aspartate/glutamate (D/E) residues and the presence of conspicuous putative Fam20°C phosphorylation SxE sites (fig. 1 and table 1; see Kawasaki et al. 2005, 2007; Kawasaki 2009; Qu et al. 2015). Furthermore, similarly to the osteichthyan Scpps, the elasmobranch Scpp protein lacks the Kazal/calcium-binding domain located in the C-terminal region of Sparc and Sparc-L/L1/L2 (fig. 1 and table 1; see Kawasaki et al. 2005, 2007; Kawasaki 2009; Bertrand et al. 2013; Qu et al. 2015). The elasmobranch Scpp protein also includes a N-terminal tyrosine-rich (Y-rich) region absent from Sparc and Sparc-L/L1/L2 but present in a wide array of osteichthyan Scpps (fig. 1 and table 1; see Kawasaki and Amemiya 2014; Kawasaki et al. 2021). The prediction for any novel *scpp* gene is that it should be expressed at sites that are relevant for biomineralization during skeletal development. We, therefore, performed in situ hybridizations on developing catshark embryos and found that *scpp* is expressed in ameloblasts from both teeth and scales, specifically at the secretory stage (fig. 2*A*, *B*, *E*, and *F*). By contrast, *sparc-L* and *fam20°C* are expressed in ameloblasts, both at the secretory and maturation stages, with an additional expression of *fam20°C* in odontoblasts (fig. 2*C*, *D*, *G*, and *H*). The spatially restricted *scpp* expression to the catshark secretory ameloblasts is strikingly similar to the situation reported for a variety of osteichthyan acidic and P/Q rich *scpp* gene members (table 1; see MacDougall et al. 1998; Kawasaki 2009; Gasse and Sire 2015; Kawasaki et al. 2021). Taken together, genomic, structural, and expression features support the classification of this novel elasmobranch gene as a *bona fide* member of the *scpp* gene family, with similarities to both the acidic and the P/Q rich osteichthyan members (table 1).

Scpp and Sparc-L Share Specific Sequence Features in Elasmobranchs

Although the signal peptide is the only similar domain between Sparc-L1/L2 and the Scpp proteins in osteichthyans (Kawasaki et al. 2005, 2007), the elasmobranch Sparc-L and

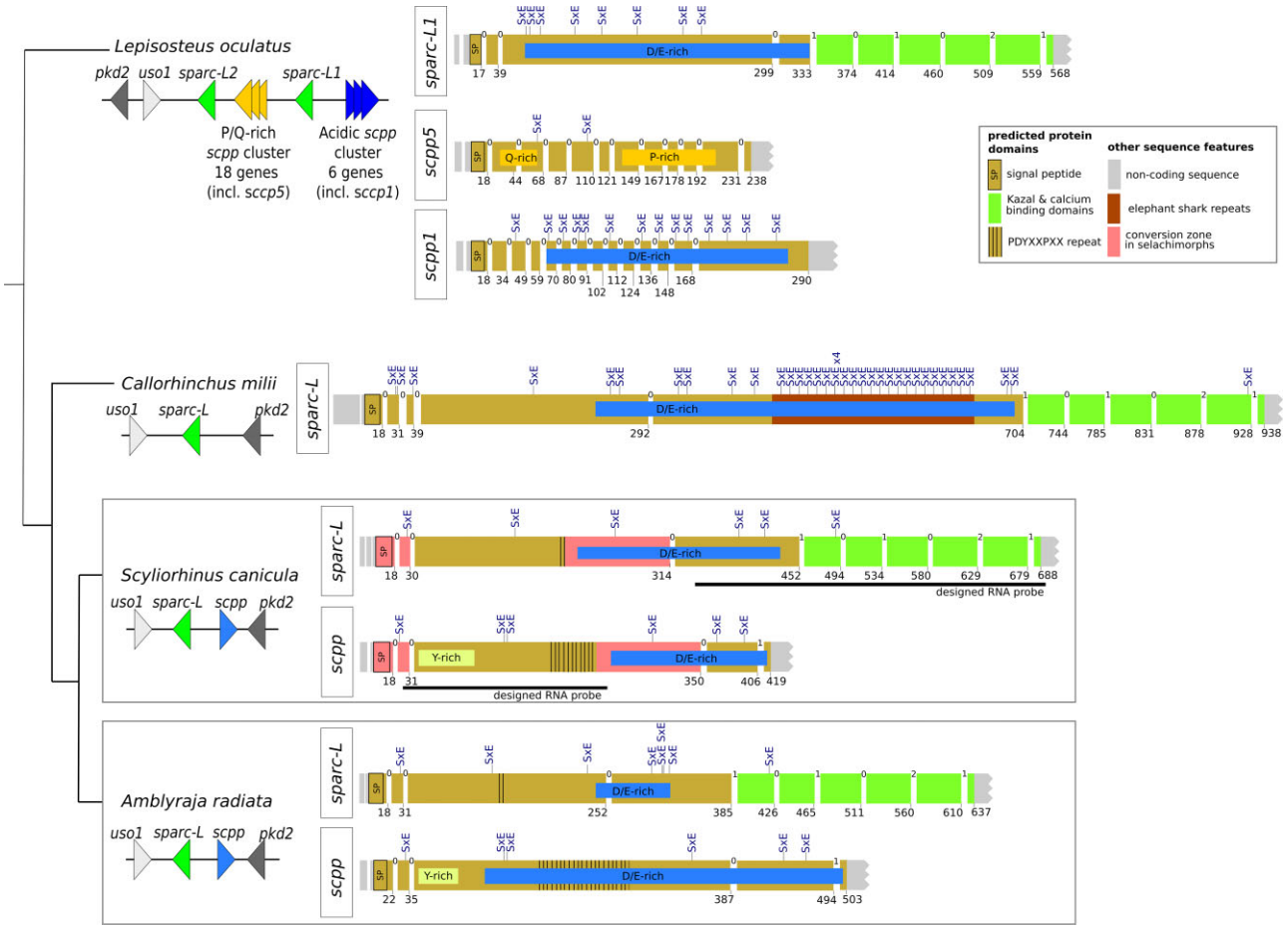

Fig. 1. Characterization of a novel *scpp* gene in shark and ray genomes. Synteny and intron–exon structure of the *sparc-L* and *scpp* loci for the indicated osteichthyan (*Lepisosteus oculatus*) and chondrichthyan species (the holocephalan *Callorhinchus milii*; the shark *S. canicula*; the ray *A. radiata*). For synteny, triangles show relative gene position and orientation. The osteichthyan *scpp* members are represented as whole P/Q-rich or acidic clusters and not individually. For each exon, a color code legend identifies the UTRs or the encoded protein domains. Numbers at the bottom right corner of each exon indicate the position of the last encoded amino-acid, and numbers between two consecutive coding exons indicate the translation phase. Putative Fam20°C phosphorylation sites (SxE) are shown along each sequence. All exons are drawn to scale, except for the last exon which is trimmed. The position of the *sparc-L* and *scpp* in situ hybridization probes is indicated.

Scpp sequences can be readily aligned and compared (fig. 1, supplementary materials S2 and S3, Supplementary Material online). Notably, elasmobranch Sparc-L and Scpp proteins both contain a variable 8 amino acid-long periodic motif whose consensus is PDYXXPXX and that is absent from the elephant shark Sparc-L (fig. 1 and table 1, supplementary materials S1, S3, and S4, Supplementary Material online) and from all the osteichthyan Sparc-L1/L2 proteins we examined (not shown). The elasmobranch Sparc-L and Scpp proteins also share a well-conserved internal region which is absent from the elephant shark Sparc-L and from the osteichthyan Sparc-L1/L2 homologues (fig. 1, supplementary material S3, Supplementary Material online). We envision three possible evolutionary scenarios for the evolution of the jawed vertebrate *scpp* genes, each of which is associated with a specific phylogenetic tree topology (fig. 3A). According to Hypothesis 1, *scpp* genes in bony and cartilaginous fishes are orthologous and evolved from a single *scpp* duplication in all jawed vertebrates and a subsequent loss in holocephalans. Hypothesis 2 states that parallel duplication events generated *scpp* sequences in bony and cartilaginous fishes, followed by a *scpp* loss in holocephalans. Finally, according to Hypothesis 3, parallel events of duplication generated *scpp* sequences in bony and elasmobranch fishes. Alignments and phylogenetic analyses were performed with full-length Scpp and Sparc-L sequences from cartilaginous fishes and osteichthyan Sparc-L1/L2 sequences (supplementary materials S2 and S3, Supplementary Material online). The *scpp* sequences from two bony fishes (*Lepisosteus oculatus* and *Erpetoichthys calabaricus*) were too divergent to generate an informative alignment (supplementary material S5, Supplementary Material online). Our phylogenetic reconstruction shows that the elasmobranch Sparc-L and Scpp group together in a well-supported node with the elephant shark Sparc-L sequence as an outgroup (fig. 3B). This result is consistent with an independent origin of the elasmobranch *scpp* gene by duplication of the *sparc-L* 5′ exons after the split from holocephalans, thereby supporting Hypothesis 3 (fig. 3A).

Table 1. Comparison of the Structure of the Scpp and Sparc-L/L1/L2 Proteins and of the Expression Patterns of Their Corresponding Genes.

	Osteichthyan acidic Scpps	Osteichthyan P/Q-rich Scpps	Elasmobranch Scpp	Sparc-L/-L1/-L2	
				N-terminal Region	C-terminal Region
Signal peptide (A)	++	++	++	++	−
Presence of SxE sites (B)	++	++	++	++	−
Kazal and calcium-binding domains (C)	−	−	−	−	++
Y-rich domain (D)	+	+	++	−	−
D/E-rich domain (E)	++	−	++	++	−
P/Q-rich domain (F)	−	++	−	−	−
PDYXXPXX repeated motif and conserved internal domain (G)	−	−	++	+ (only in elasmobranch Sparc-L)	−
Ameloblastic expression (H)	+ (mouse *Dpm1*, zebrafish *spp1*)	++	++	+ (only for the elasmobranch *sparc-L*)	
Odontoblastic expression (I)	++	+	−	−	

Note.—The table recapitulates the presence (++, in all or most examined members; and +, at least in some members, shown in green) or the absence (−) of structural and expression features for the indicated proteins and their corresponding genes. Sources: (A) this study and see Kawasaki et al. (2005, 2007); (B) this study and see Kawasaki et al. (2005, 2007); (C) this study and see Kawasaki et al. (2005, 2007), Kawasaki (2009), Bertrand et al. (2013); (D) this study and see Kawasaki and Amemiya (2014), Kawasaki et al. (2021); (E) this study and see Kawasaki et al. (2005, 2007), Kawasaki (2009), Qu et al. (2015); (F) Kawasaki and Amemiya (2014), Kawasaki et al. (2017); (G) this study; (H) In osteichthyans, P/Q-rich *scpp* members are typically expressed in ameloblasts (Kawasaki 2009). Similarly to the elasmobranch *scpp* gene, a specific expression at the early secretory stage has been reported for a variety of P/Q-rich *scpps* such as the lizard and salamander *amtn* gene (Gasse and Sire 2015) and the gar *ambn*, *enam*, and *scpp5* genes (Kawasaki et al. 2021). Osteichthyan acidic *scpps* are typically associated with dentine and bone (Kawasaki 2009), but a specific expression in secretory ameloblasts has also been reported for zebrafish *spp1* (Kawasaki 2009) and mouse *Dmp1* (MacDougall et al. 1998). Although the mouse *Sparc-L1* gene is not expressed in teeth, nor in any other skeletal tissue (Soderling et al. 1997), the elasmobranch *sparc-L* gene harbors a pan-ameloblastic expression (this study and see Enault et al. 2018); (I) this study and see Enault et al. (2018), Kawasaki et al. (2005, 2021), Kawasaki (2009), Yonekura et al. (2013).

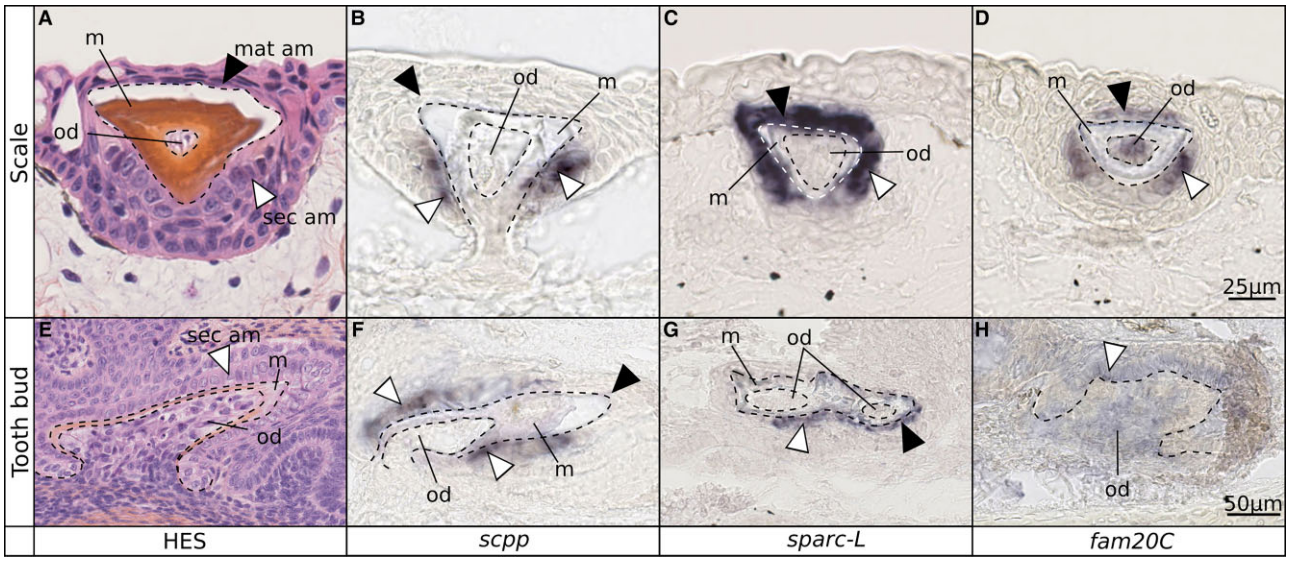

Fig. 2. The lesser spotted catshark *scpp* gene is coexpressed with *sparc-L* and *fam20°C* in tooth and scale ameloblasts. Sections were performed at the level of developing scales (*A*–*D*, 8 cm-long embryos) and teeth (*E*–*H*, 9.5 cm-long embryos). Hematoxylin-Eosin-Safran histological staining (*A* and *E*) shows cells in the epithelium (ameloblasts, am) and the mesenchyme (odontoblasts, od), as well as the mineralized extracellular matrix (m). White arrowheads indicate secretory ameloblasts (sec am) and black arrowheads indicate maturation-stage ameloblasts (mat am). In situ hybridization signal identifies cells expressing *scpp* (*B* and *F*), *sparc-L* (*C* and *G*), or *fam20°C* (*D* and *H*).

The Elasmobranch *scpp* Gene Experienced Nonallelic Gene Conversion with *sparc-L* Specifically in Selachimorphs

The phylogenetic analysis using full-length proteins shows that the Sparc-L and Scpp sequences consistently group together as species pairs in the shark clade (fig. 3*B*), a phenomenon reminiscent of concerted evolution due to nonallelic gene conversion between paralogs (Zhou et al. 2019). From this tree topology, it is impossible to deduce if the *scpp* gene evolved once at the base of the elasmobranchs, or if it results from independent duplications in selachimorphs and batoids. Performing a stringent genomic alignment reveals that gene conversion has occurred in four discrete genomic regions shared between the catshark *sparc-L* and *scpp* genes that encompass the coding exons 1 and 2, and overlaps with the end of the coding exon 3 (fig. 4*A*, supplementary material S6, Supplementary Material online). Accordingly, a phylogenetic reconstruction

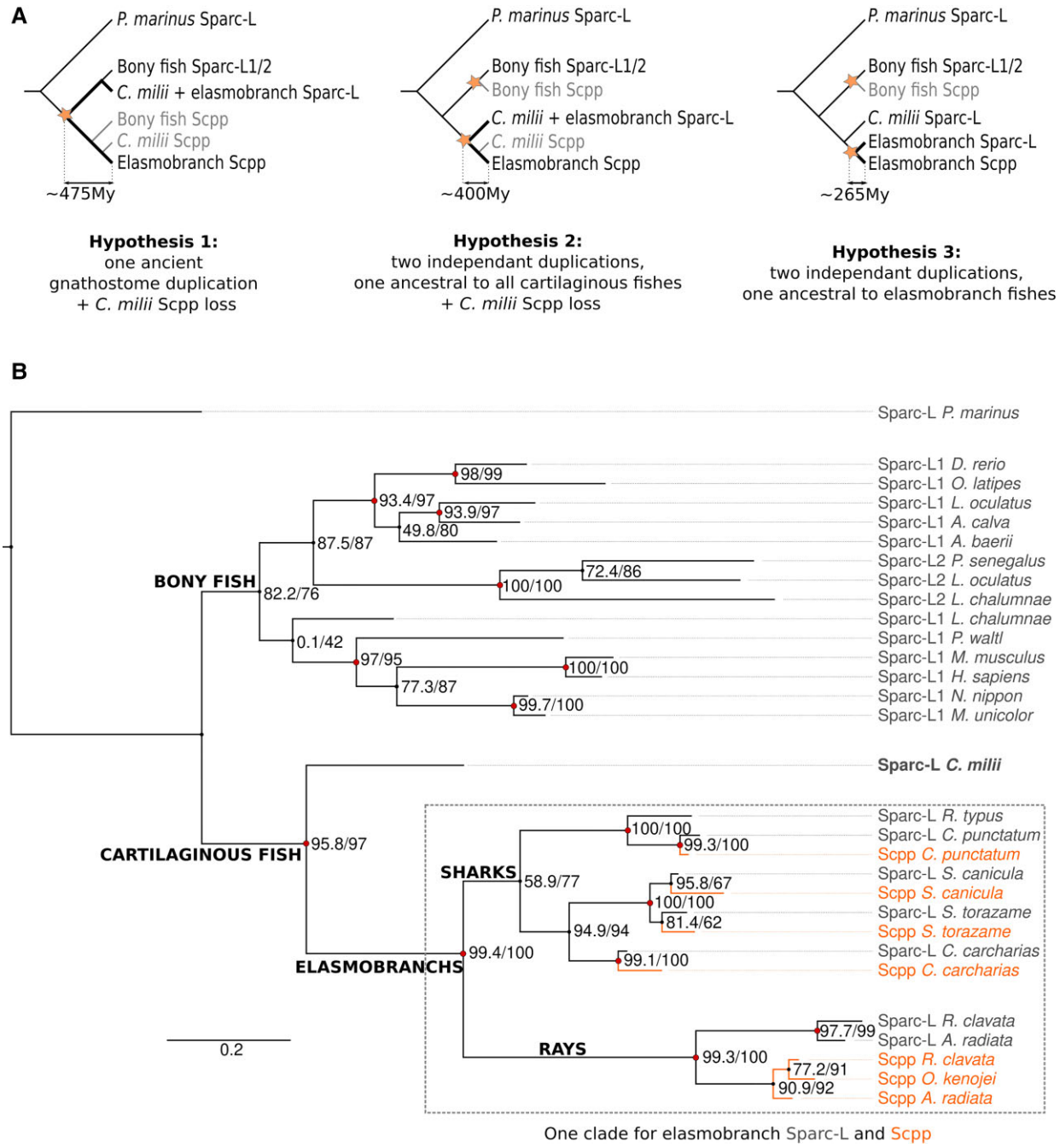

Fig. 3. Evolution of *scpp* from the *sparc-L* locus. (*A*) The cladograms represent three possible evolutionary scenarios: *scpp* genes in bony and cartilaginous fishes are orthologous and evolved from a single *scpp* duplication in all jawed vertebrates, followed by a *scpp* loss in holocephalans (Hypothesis 1); parallel duplication events generated *scpp* sequences in bony fishes and in cartilaginous fishes, followed by a *scpp* loss in holocephalans (Hypothesis 2); parallel duplication events generated *scpp* sequences in bony and elasmobranch fishes (Hypothesis 3). (*B*) Phylogenetic reconstruction of chondrichthyan Sparc-L and Scpp proteins. The phylogeny was performed with full-length Scpp and Sparc-L sequences and was rooted with the sea lamprey Sparc-L sequence (refer to supplementary material S14, Supplementary Material online for species names and accession numbers). The evolutionary model used was JTT + I + G4 and the alignment was 538 amino acid long (supplementary material S3, Supplementary Material online). Node support was evaluated with sh-lrt and 5000 ultra-fast bootstrap replicates, values are indicated on each node. Scpp proteins are shown in orange and red dots indicate bootstraps values superior to 95.

performed solely with the regions subjected to gene conversion (i.e., amino acid sequences encoded by coding exons 1 and 2 and the end of coding exon 3) groups sequences by shark species pairs, whereas the ray sequences form two distinct Sparc-L and Scpp clades (fig. 4*B*, supplementary material S7, Supplementary Material online). By contrast, a phylogenetic analysis excluding the conversion regions the variable PDYXXPXX repeat together with the Kazal and calcium-binding domains separate the elasmobranch sequences into two well-supported

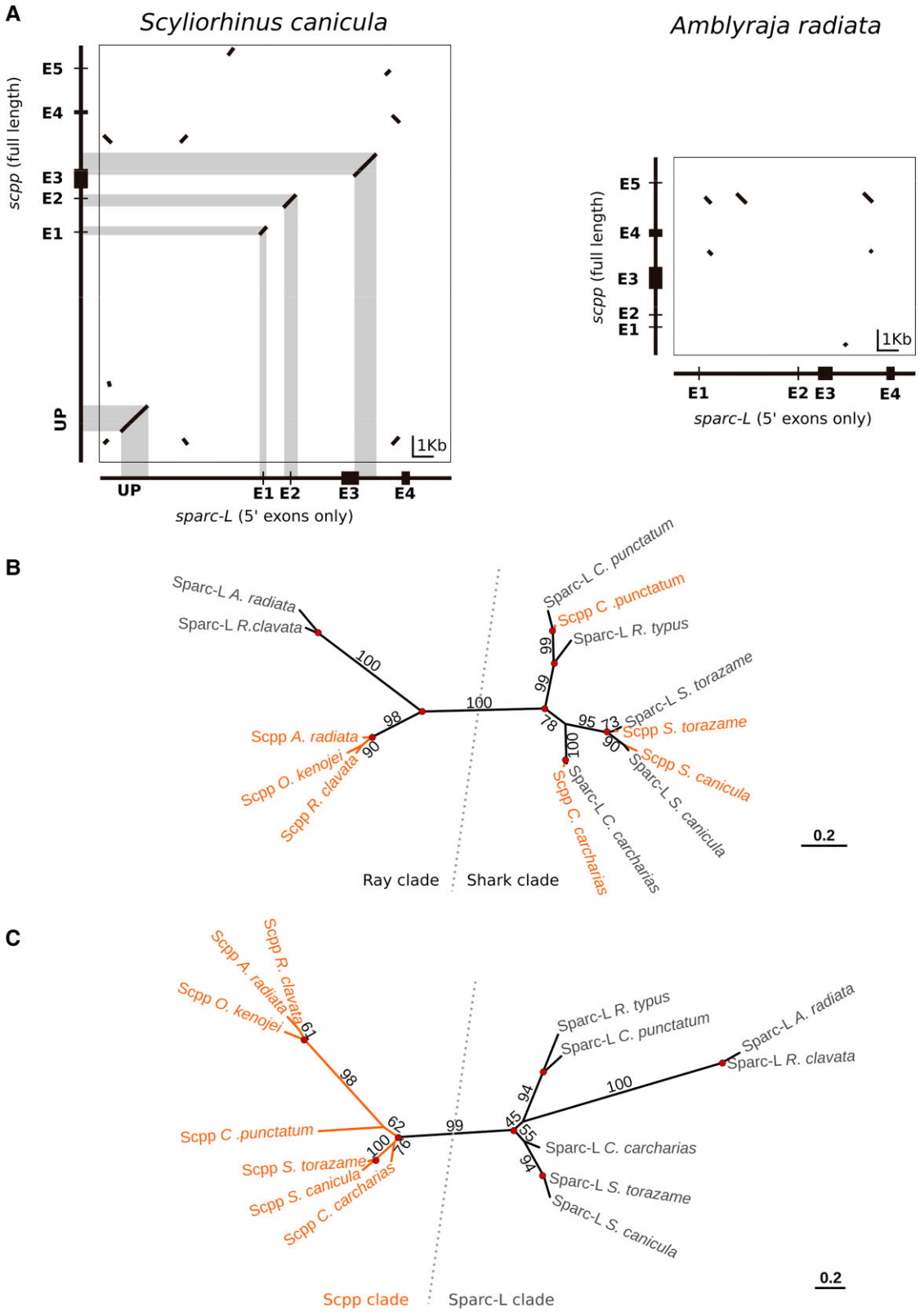

FIG. 4. The elasmobranch *scpp* gene experienced nonallelic gene conversion with *sparc-L* specifically in selachimorphs. (*A*) Dot blots representing the position of highly conserved sequences between genomic regions containing *sparc-L* and *scpp* of *S. canicula* (left panel) and *A. radiata* (right panel). The shaded areas correspond to the conversion zones (aligned in supplementary material S7, Supplementary Material online). The 1 kb scale applies horizontally and vertically. (*B*) Unrooted phylogenetic tree performed with protein regions under gene conversion, under the JTT model (alignment from supplementary material S7, Supplementary Material online). (*C*) Unrooted phylogenetic tree performed with sequences excluding the conversion regions, the variable repeat and the Kazal/calcium-binding domains, under the FLU + F + G4 model (alignment from supplementary material S8, Supplementary Material online). In (*B*) and (*C*), Sparc-L and Scpp proteins are shown in grey and orange, respectively, bootstrap values are indicated on branches and red nodes indicate values superior to 95.

Sparc-L and Scpp clades (fig. 4C, supplementary material S8, Supplementary Material online), which is also recovered in a phylogenetic reconstruction with the full species sampling (supplementary materials S9 and S10, Supplementary Material online). These analyses support both: (1) recent gene conversion in the shark lineage between *sparc-L* and *scpp* and (2) a single evolutionary origin for a *scpp* gene in all elasmobranchs. We reasoned that using the Ks value as a proxy to the timing of divergence between the *sparc-L* and *scpp* duplicates would represent an unbiased way to further test Hypothesis 3 (fig. 3A). Hence, we calculated a synonymous mutation rate (Ks) in the nucleotide sequence located outside of the conversion zone (supplementary materials S11 and S12, Supplementary Material online), and used the available median value for selachimorph neutral mutation rate (Ks $\approx$ 0.2/100 My, according to Hara et al. 2018). Ks calculations between pairs of paralogs from the same selachimorph species had mean values of $\sim$0.7 (see branch lengths in supplementary material S12, Supplementary Material online) corresponding to a timing of divergence between the *sparc-L* and *scpp* genes of about 350 My. These values, therefore, confirm an origin of *scpp* from a local duplication of *sparc-L* in an ancestor of all elasmobranch fishes (265–400 My; third hypothesis in fig. 3A).

An Acidic N-terminal Domain and Fam20°C Phosphorylation Sites are Ancestral Features of Chondrichthyan Sparc-L Proteins

In elasmobranchs, exon–intron structure, protein alignment, and phylogeny show that the Scpp proteins evolved from the Sparc-L N-terminal region (i.e., excluding the Kazal/calcium-binding domain). As the Sparc-L N-terminal region from sharks and rays is enriched in acidic residues and contains SxE consensus sites (fig. 1 and table 1, supplementary material S1, Supplementary Material online), we speculate that both structural features were already present in the last common ancestor of all elasmobranchs and were inherited by Scpp upon its emergence in this lineage. Similarly, previous work had noted that the elephant shark Sparc-L N-terminal region is highly acidic in nature (Kawasaki et al. 2017). Here, we further show that elephant shark Sparc-L acidity is dramatically exacerbated by the amplification of a 24 amino acid-long motif (containing 39.3% of D/E residues, fig. 1, see supplementary materials S1 and S4, Supplementary Material online) that suffices to decrease the theoretical isoelectric point of the protein (4.14 with 11 repeats vs. 4.52 with one repeat). In addition, the region encompassed by the duplicated motif contains 26 putative Fam20°C phosphorylation sites (fig. 1, supplementary material S1 and S4, Supplementary Material online) generating in the N-terminal region of elephant shark Sparc-L the highest abundance of SxE motifs of all sequences examined in this study (5.1 sites per 100 a.a. vs. 1.7 and 1.8 sites per 100 a.a. on average for the elasmobranch Sparc-L N-terminal regions and Scpp proteins, respectively). Despite being poorly alignable, the elasmobranch and holocephalan Sparc-L N-terminal regions display both an enrichment in acidic residues and the conspicuous presence of putative Fam20°C phosphorylation sites, which can be inferred to represent ancestral characters of all chondrichthyan Sparc-L proteins, and to have been inherited by Scpp upon its emergence in elasmobranchs.

Discussion

We demonstrate that a *scpp* gene evolved by duplication of the *sparc-L* 5′exons of in elasmobranchs, similarly to their evolution through the duplication of the *sparc-L1* and/or *L2* in osteichthyans, revealing at least two independent origins of *scpp* genes in jawed vertebrates. Independent evolution of the bony fish and elasmobranch *scpp* genes is supported both by our phylogenetic reconstructions and by molecular clock analysis. Despite convergent origins of the elasmobranch and the osteichthyan *scpp* genes, we chose to name this duplicate accordingly, because the duplication events that generated *scpp* genes in osteichthyans are largely unknown (discussed in Kawasaki et al. 2017) and may also include parallel duplication events. Furthermore, we reveal a history of gene conversion between *scpp* and *sparc-L* in selachimorphs, which is comparable to other cases of lineage-specific gene conversion between tandem duplicates in vertebrate genomes, as reported for the *protocadherin* and *beta-globin* genes (Noonan et al. 2004; Hoffmann et al. 2018). Remarkably, the elasmobranch Scpp convergently evolved emblematic characteristics of the osteichthyan Scpp proteins, as it lacks the Sparc-L Kazal/calcium-binding domain (Kawasaki et al. 2004) and acquired an N-terminal tyrosine-rich domain (Kawasaki and Amemiya 2014; Kawasaki et al. 2021). An additional level of convergence results from the highly specific expression patterns of *scpp* genes in presecretory ameloblasts (fig. 2). Indeed, expression data in shark and ray showed that *sparc-L* was already a pan-ameloblastic gene in the last common ancestor of all elasmobranchs (Enault et al. 2018), implying that, upon duplication, *scpp* diverged to restrict its expression pattern to the secretory phase of amelogenesis, as has been observed for several osteichthyan *scpp* genes (table 1; see Gasse et al. 2015; Kawasaki et al. 2021). In this respect, the parallel evolution of osteichthyan and elasmobranch Scpp proteins involved coordinated changes at: (1) the genomic level (i.e., a lineage-specific duplication of 5′ exons of the ancestral *sparc-L or sparc-L1 or -L2* that excludes the exons coding for the C-terminal Kazal/calcium-binding domain); (2) the structural level (i.e., the emergence of a Y-rich domain, the maintenance of acidic residues and SxE sites in spite of significant sequence turnover); and (3) the regulatory level (i.e., the restriction of expression to the secretory ameloblasts). *Scpp* parallel evolution, therefore, represents a unique scenario that improves our understanding of independently evolved genes. Indeed, our proposed model (fig. 5) differs from other cases of independently evolved proteins that have

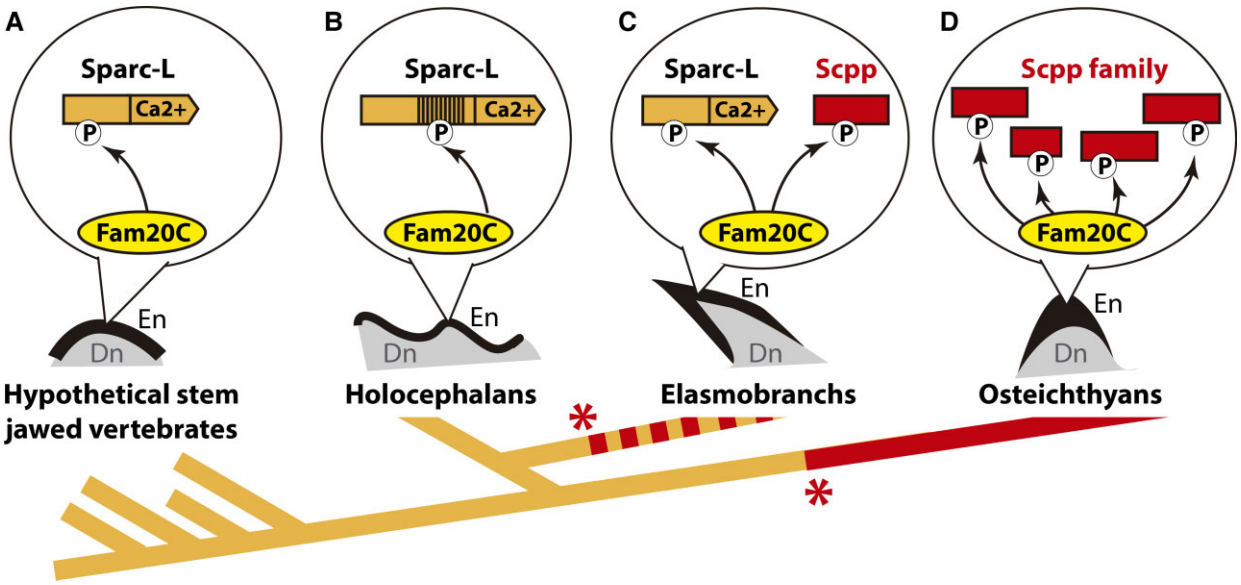

Fig. 5. Evolutionary scenario for the turnover of the jawed vertebrate ameloblastic genetic toolkit. The composition of the enamel/enameloid mineralization toolkit is shown for the hypothetical stem jawed vertebrates (*A*), holocephalans (*B*), elasmobranchs (*C*), and osteichthyans (*D*), whose odontodes contain a dentine core (Dn) covered by an enamel/enameloid layer (En). Arrows indicate the phosphorylation (P) of Sparc-L (orange) and/or Scpp (red) proteins by Fam20°C (yellow). Sparc-L harbors a C-terminal Kazal/calcium-binding domain (Ca2+) absent from Scpp proteins. The holocephalan Sparc-L N-terminal domain contains 11 tandem duplications of an acidic motif (vertical lines) enriched in SxE sites. A cladogram shows for each jawed vertebrate lineage the ameloblastic expression of *sparc-L* and *scpp* genes (orange and red branches, respectively). Asterisks indicate parallel *scpp* duplication events. According to this model the ancestral enamel/enameloid mineralization toolkit was based on Sparc-L, a situation maintained in holocephalans. The elasmobranch toolkit includes Sparc-L and Scpp, whereas Sparc-L1 and Sparc-L2 were functionally replaced by the Scpp family in osteichthyans.

originated from parallel fixation of mutations in different paralogs belonging to relatively large preexisting gene families (Christin et al. 2007; Flores-Aldama et al. 2020; Barua et al. 2021), from the progressive accumulation of mutations to produce a similar motif (Kriener et al. 2000), or from the emergence of biased amino-acid domains in proteins of completely unrelated origins (Calatayud et al. 2021).

The specific expression of *scpp* in secretory ameloblasts during catshark tooth and scale development strongly suggests that the elasmobranch *scpp* is a functional gene involved in enameloid matrix secretion and/or maturation. This idea is further supported by the presence of acidic domains and conspicuous putative Fam20°C phosphorylation sites in the elasmobranch Scpp proteins, as well as by the expression of *fam20°C* in catshark ameloblasts. Hence, in elasmobranchs, sequence and expression data support the hypothesis that Scpp binds calcium ions and is phosphorylated by Fam20°C during matrix mineralization, as has been reported for osteichthyan Scpp proteins (Tagliabracci et al. 2012; Wang et al. 2013; Klaning et al. 2014; Liu et al. 2017; Zhang et al. 2018; Schytte et al. 2020). Intriguingly, in the elephant shark, the Sparc-L protein is the most acidic and most enriched in putative Fam20°C phosphorylation sites, whereas the enameloid cap is relatively thin (Gillis and Donoghue 2007; Jerve et al. 2014). This apparent paradox might be resolved by considering that enameloid reduction in *Callorhinchus milii* is a derived character, and that holocephalan stem group specimen displays a well-mineralized outer odontode layer (Gillis and Donoghue 2007; Jerve et al. 2014). In addition to the chondrichthyan Sparc-L, the osteichthyan Sparc-L1/L2 proteins also harbor an N-terminal domain enriched in acidic residues (Kawasaki and Weiss 2006; Kawasaki et al. 2017) and in SxE sites (1.9 sites per 100 a.a. on average, data not shown). As a result, we conclude that, in the jawed vertebrate lineage, an acidic Sparc-L protein was ancestrally phosphorylated by Fam20°C and was also binding calcium during enameloid formation, a situation that probably already existed in osteostracans and placoderms (fig. 5*A*), and that remained largely unchanged in the holocephalan lineage (fig. 5*B*). Elasmobranchs and osteichthyans convergently evolved Scpp proteins that respectively function redundantly with Sparc-L (fig. 5*C*) or eventually took over the role ancestrally played by Sparc-L1/L2 (fig. 5*D*).

Previously, the absence of *scpp* genes in holocephalans was thought to be representative of the whole chondrichthyan group, and has been used as an argument to propose that osteichthyan enamel and chondrichthyan enameloid are not homologous tissues (Qu et al. 2015; Kawasaki et al. 2017, 2021). Here, we reinforce arguments supporting the homology of some developmental processes that generate an outer hypermineralized layer in vertebrate odontodes (also discussed in Kawasaki et al. 2017). We consolidate the idea that an acidic Sparc-L subjected to Fam20°C posttranslational regulation represents a crucial molecular element of the ancestral ameloblastic genetic toolkit. Although Sparc-L and Scpp are probably still partially redundant in elasmobranchs, the dramatic

expansion of the *Scpp* gene family in osteichthyans has clearly minimized, or even abolished, the skeletal expression and function of Sparc-L1/L2 in this group (Bertrand et al. 2013; Enault et al. 2018), explaining why enamel was previously thought to be fundamentally different from chondrichthyan enameloid. One of the earliest works on the *scpp* genes employed the term "phenogenetic drift" to illustrate how homologous enamel and ganoin extracellular matrices from distinct osteichthyan lineages have progressively modified the nature and composition of the Scpp proteins they contain (Kawasaki et al. 2005). Our study expands the concept of phenogenetic drift by showing that the ameloblastic cassette is subjected to a high rate of turnover in jawed vertebrates, leading to different sets of proteins being secreted to the mineralizing front of the outer odontode layer of holocephalans, elasmobranchs, and osteichthyans. In this context, the hypothesis of an ancestral regulation of Sparc-L by Fam20°C represents a unifying genetic basis that, if demonstrated, would prove the homology of developmental processes that evolved in the remarkably diverse enamel/ganoin/enameloid cap covering the teeth and scales of extant and extinct jawed vertebrates.

Materials and Methods

Animal care, staging, histology, and in situ hybridization procedures of embryonic lesser spotted catshark *S. canicula* were performed as previously described (Enault et al. 2018). The *scpp* and *fam20°C* sequences were amplified and cloned using lesser spotted catshark vertebrae cDNA and the following primers (5′–3′): *scpp*-Forward GATTTGGG CAGCAACAGTCA, *scpp*-Reverse TGTCTAACCCCGGTG TGAAA; *fam20°C*-Forward GGCTGCTGATCATCATGG TG, *fam20°C*-Reverse GGAAAGCAGCAATCTCCGAG, to be used for RNA probe synthesis. Refer to Enault et al. (2018) for *sparc-L* probe synthesis. All in situ hybridizations included two negative control slides (no probe and sense probe), which were devoid of signal (supplementary material S13, Supplementary Material online).

All accession numbers and genomic coordinates are available in supplementary material S14, Supplementary Material online. Amino acid enrichments were detected with CAST (Promponas et al. 2000), internal repeats were searched with XSTREAM (Newman and Cooper 2007), and repeat logos were generated with the GGSEQLOGO R package (Wagih 2017). Isoelectric point was computed using Expasy online tool (Gasteiger et al. 2003). Highly similar sequences between the *Sparc-L* and *Scpp* loci were searched using BLAST 2.2.19 with default parameters (Altschul et al. 1997).

Sequences used for phylogenetic reconstructions were aligned using MAFFT (Katoh et al. 2002; Katoh and Standley 2013), using standard parameters. Alignments were cleaned using HmmCleaner with standard parameters (Di Franco et al. 2019). Phylogenetic analyses were performed using Maximum Likelihood under IQ-TREE 1.6.12 (Nguyen et al. 2015). Node support was estimated by performing five thousand ultra-fast (UF) bootstrap replicates (Hoang et al. 2018) and single branch tests (SH-aLRT, see Guindon et al. 2010). The minimum correlation coefficient for UF bootstrap convergence criterion was set to 0.90 and evolutionary models were chosen by ModelFinder (Kalyaanamoorthy et al. 2017). CDS nucleotide sequences for chondrichthyan *scpp* and *sparc-L* were aligned using MACSE (Ranwez et al. 2018). The resulting alignments were trimmed to exclude the conversion zones and Kazal domains. The trimmed alignment (supplementary material S11, Supplementary Material online) and the topology of the concatenated phylogeny (fig 4C) with *C. milii* as an outgroup were used to compute Ks values for each branch using PAML v4.9c (Yang 2007).

Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

Acknowledgments

We acknowledge the MRI platform (ANR-10-INBS-04), the GenSeq platform, Labex CEMEB (ANR-10-LABX-0004), and NUMEV (ANR-10-LABX-0020), and the RHEM facility (INCa_Inserm_DGOS_12553, FEDER-FSE 2014–2020 Languedoc Roussillon) for histology technical expertise, the IRP (LIA) MAST (CNRS) and the ANID (FONDECT 1190926 to S.M.). We thank Felipe Aguilera for help with the GGSEQLOGO R package, Benoit Nabholz for discussion on Ks, and Jake Leyhr, Tatjana Haitina, and Didier Casane for critical reading of the manuscript. We are grateful to the anonymous reviewers for their constructive comments that improved the quality of the present study.

Author Contributions

N.L. mined and analyzed the sequence data, N.L. and C.M.-M. performed the in situ hybridization experiments, S.M. and M.D.-T. prepared the figures and wrote the manuscript, all authors analyzed and interpreted the data.

Data Availability

The data underlying this article are available in the article and in its online supplementary material. Any complementary data will be shared on reasonable request to the corresponding author.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**:3389–3402.

Barua A, Koludarov I, Mikheyev AS. 2021. Co-option of the same ancestral gene family gave rise to mammalian and reptilian toxins. *BMC Biol.* **19**:268.

Berio F, Broyon M, Enault S NP, Lopez-Romero FA, Debiais-Thibaud M. 2021. Diversity and evolution of mineralized skeletal tissues in chondrichthyans. *Front Ecol Evol*. **9**:1–19.

Bertrand S, Fuentealba J, Aze A, Hudson C, Yasuo H, Torrejon M, Escriva H, Marcellini S. 2013. A dynamic history of gene duplications and losses characterizes the evolution of the SPARC family in eumetazoans. *Proc Biol Sci*. **280**:20122963.

Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, Amores A, Desvignes T, Batzel P, Catchen J, *et al*. 2016. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat Genet*. **48**:427–437.

Brazeau MD, Friedman M. 2015. The origin and early phylogenetic history of jawed vertebrates. *Nature* **520**:490–497.

Brazeau MD, Giles S, Dearden RP, Jerve A, Ariunchimeg Y, Zorig E, Sansom R, Guillerme T, Castiello M. 2020. Endochondral bone in an Early Devonian 'placoderm' from Mongolia. *Nat Ecol Evol*. **4**:1477–1484.

Calatayud S, Garcia-Risco M, Palacios O, Capdevila M, Canestro C, Albalat R. 2021. Tunicates illuminate the enigmatic evolution of chordate metallothioneins by gene gains and losses, independent modular expansions, and functional convergences. *Mol Biol Evol*. **38**:4435–4448.

Cheng P, Huang Y, Lv Y, Du H, Ruan Z, Li C, Ye H, Zhang H, Wu J, Wang C, *et al*. 2021. The American Paddlefish genome provides novel insights into chromosomal evolution and bone mineralization in early vertebrates. *Mol Biol Evol*. **38**: 1595–1607.

Christin PA, Salamin N, Savolainen V, Duvall MR, Besnard G. 2007. C4 photosynthesis evolved in grasses via parallel adaptive genetic changes. *Curr Biol*. **17**:1241–1247.

Di Franco A, Poujol R, Baurain D, Philippe H. 2019. Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evol Biol*. **19**:21.

Donoghue PC, Sansom IJ. 2002. Origin and early evolution of vertebrate skeletonization. *Microsc Res Tech*. **59**:352–372.

Donoghue PC, Sansom IJ, Downs JP. 2006. Early evolution of vertebrate skeletal tissues and cellular interactions, and the canalization of skeletal development. *J Exp Zool B Mol Dev Evol*. **306**:278–294.

Enault S, Munoz D, Simion P, Venteo S, Sire JY, Marcellini S, Debiais-Thibaud M. 2018. Evolution of dental tissue mineralization: an analysis of the jawed vertebrate SPARC and SPARC-L families. *BMC Evol Biol*. **18**:127.

Flores-Aldama L, Vandewege MW, Zavala K, Colenso CK, Gonzalez W, Brauchi SE, Opazo JC. 2020. Evolutionary analyses reveal independent origins of gene repertoires and structural motifs associated to fast inactivation in calcium-selective TRPV channels. *Sci Rep*. **10**:8684.

Gasse B, Chiari Y, Silvent J, Davit-Beal T, Sire JY. 2015. Amelotin: an enamel matrix protein that experienced distinct evolutionary histories in amphibians, sauropsids and mammals. *BMC Evol Biol*. **15**:47.

Gasse B, Sire JY. 2015. Comparative expression of the four enamel matrix protein genes, *amelogenin*, *ameloblastin*, *enamelin* and *amelotin* during amelogenesis in the lizard *Anolis carolinensis*. *Evodevo* **6**:29.

Gasteiger E, Gattiker A, Hoogland C, Ivanyi I, Appel RD, Bairoch A. 2003. ExPASy:The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*. **31**(13):3784–3788.

Gillis JA, Donoghue PC. 2007. The homology and phylogeny of chondrichthyan tooth enameloid. *J Morphol*. **268**:33–49.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*. **59**:307–321.

Hara Y, Yamaguchi K, Onimaru K, Kadota M, Koyanagi M, Keeley SD, Tatsumi K, Tanaka K, Motone F, Kageyama Y, *et al*. 2018. Shark genomes provide insights into elasmobranch evolution and the origin of vertebrates. *Nat Ecol Evol*. **2**:1761–1771.

Hoang DT, Vinh LS, Flouri T, Stamatakis A, von Haeseler A, Minh BQ. 2018. MPBoot: fast phylogenetic maximum parsimony tree inference and bootstrap approximation. *BMC Evol Biol*. **18**:11.

Hoffmann FG, Vandewege MW, Storz JF, Opazo JC. 2018. Gene turnover and diversification of the *alpha-* and *beta-Globin* gene families in sauropsid vertebrates. *Genome Biol Evol*. **10**:344–358.

Jerve A, Johanson Z, Ahlberg P, Boisvert C. 2014. Embryonic development of fin spines in *Callorhinchus milii* (Holocephali); implications for chondrichthyan fin spine evolution. *Evol Dev*. **16**: 339–353.

Johanson Z, Kearsley A, den Blaauwen J, Newman M, Smith MM. 2012. Ontogenetic development of an exceptionally preserved Devonian cartilaginous skeleton. *J Exp Zool B Mol Dev Evol*. **318**:50–58.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **14**:587–589.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. **30**:3059–3066.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. **30**:772–780.

Kawasaki K. 2009. The *SCPP* gene repertoire in bony vertebrates and graded differences in mineralized tissues. *Dev Genes Evol*. **219**: 147–157.

Kawasaki K. 2011. The *SCPP* gene family and the complexity of hard tissues in vertebrates. *Cells Tissues Organs* **194**:108–112.

Kawasaki K, Amemiya CT. 2014. *SCPP* genes in the coelacanth: tissue mineralization genes shared by sarcopterygians. *J Exp Zool B Mol Dev Evol*. **322**:390–402.

Kawasaki K, Buchanan AV, Weiss KM. 2007. Gene duplication and the evolution of vertebrate skeletal mineralization. *Cells Tissues Organs* **186**:7–24.

Kawasaki K, Keating JN, Nakatomi M, Welten M, Mikami M, Sasagawa I, Puttick MN, Donoghue PCJ, Ishiyama M. 2021. Coevolution of enamel, ganoin, enameloid, and their matrix *SCPP* genes in osteichthyans. *iScience* **24**:102023.

Kawasaki K, Lafont AG, Sire JY. 2011. The evolution of milk casein genes from tooth genes before the origin of mammals. *Mol Biol Evol*. **28**:2053–2061.

Kawasaki K, Mikami M, Nakatomi M, Braasch I, Batzel P, Postlethwait JH, Sato A, Sasagawa I, Ishiyama M. 2017. *SCPP* genes and their relatives in gar: rapid expansion of mineralization genes in osteichthyans. *J Exp Zool B Mol Dev Evol*. **328**:645–665.

Kawasaki K, Suzuki T, Weiss KM. 2004. Genetic basis for the evolution of vertebrate mineralized tissue. *Proc Natl Acad Sci U S A* **101**:11356–11361.

Kawasaki K, Suzuki T, Weiss KM. 2005. Phenogenetic drift in evolution: the changing genetic basis of vertebrate teeth. *Proc Natl Acad Sci U S A*. **102**:18063–18068.

Kawasaki K, Weiss KM. 2003. Mineralized tissue and vertebrate evolution: the secretory calcium-binding phosphoprotein gene cluster. *Proc Natl Acad Sci U S A*. **100**:4060–4065.

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

Keating JN, Marquart CL, Marone F, Donoghue PCJ. 2018. The nature of aspidin and the evolutionary origin of bone. *Nat Ecol Evol*. **2**: 1501–1506.

Klaning E, Christensen B, Sorensen ES, Vorup-Jensen T, Jensen JK. 2014. Osteopontin binds multiple calcium ions with high affinity and independently of phosphorylation status. *Bone* **66**:90–95.

Kriener K, O'HUigin C, Tichy H, Klein J. 2000. Convergent evolution of major histocompatibility complex molecules in humans and New World monkeys. *Immunogenetics* **51**:169–178.

Lemierre A, Germain D. 2019. A new mineralized tissue in the early vertebrate Astraspis. *J Anat*. **235**:1105–1113.

Lin Q, Fan S, Zhang Y, Xu M, Zhang H, Yang Y, Lee AP, Woltering JM, Ravi V, Gunter HM, *et al.* 2016. The seahorse genome and the evolution of its specialized morphology. *Nature* **540**:395–399.

Liu P, Ma S, Zhang H, Liu C, Lu Y, Chen L, Qin C. 2017. Specific ablation of mouse Fam20°C in cells expressing type I collagen leads to skeletal defects and hypophosphatemia. *Sci Rep*. **7**:3590.

Luis Villanueva-Cañas J, Ruiz-Orera J, Agea MI, Gallo M, Andreu D, Albà MM. 2017. New genes and functional innovation in mammals. *Genome Biol Evol*. **9**:1886–1900.

Lv Y, Kawasaki K, Li J, Li Y, Bian C, Huang Y, You X, Shi Q. 2017. A genomic survey of *SCPP* family genes in fishes provides novel insights into the evolution of fish scales. *Int J Mol Sci*. **18**:2432.

MacDougall M, Nydegger J, Gu TT, Simmons D, Luan X, Cavender A, D'Souza RN. 1998. Developmental regulation of dentin sialophosphoprotein during ameloblast differentiation: a potential enamel matrix nucleator. *Connect Tissue Res*. **39**:25–37; discussion 63–27.

Min Z, Janvier P. 1998. The histological structure of the endoskeleton in galeaspids (Galeaspida, Vertebrata). *J Vertebrate Paleontol*. **18**: 650–654.

Newman AM, Cooper JB. 2007. XSTREAM: a practical algorithm for identification and architecture modeling of tandem repeats in protein sequences. *BMC Bioinform*. **8**:382.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol*. **32**:268–274.

Nikoloudaki G. 2021. Functions of matricellular proteins in dental tissues and their emerging roles in orofacial tissue development, maintenance, and disease. *Int J Mol Sci*. **22**:6626.

Noonan JP, Grimwood J, Danke J, Schmutz J, Dickson M, Amemiya CT, Myers RM. 2004. Coelacanth genome sequence reveals the evolutionary history of vertebrate genes. *Genome Res*. **14**: 2397–2405.

Ørvig T. 1951. Histologic studies of placoderms and fossil elasmobranchs. I: the endoskeleton, with remarks on the hard tissues of lower vertebrates in general. *Arkiv Zool*. **2**:321–454.

Promponas VJ, Enright AJ, Tsoka S, Kreil DP, Leroy C, Hamodrakas S, Sander C, Ouzounis CA. 2000. CAST: an iterative algorithm for the complexity analysis of sequence tracts. Complexity analysis of sequence tracts. *Bioinformatics* **16**:915–922.

Qu Q, Haitina T, Zhu M, Ahlberg PE. 2015. New genomic and fossil data illuminate the origin of enamel. *Nature* **526**: 108–111.

Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. 2018. MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. *Mol Biol Evol*. **35**: 2582–2584.

Schytte GN, Christensen B, Bregenov I, Kjøge K, Scavenius C, Petersen SV, Enghild JJ, Sørensen ES. 2020. FAM20°C phosphorylation of the RGDSVVYGLR motif in osteopontin inhibits interaction with the alphavbeta3 integrin. *J Cell Biochem*. **121**:4809–4818.

Shin NY, Yamazaki H, Beniash E, Yang X, Margolis SS, Pugach MK, Simmer JP, Margolis HC. 2020. Amelogenin phosphorylation regulates tooth enamel formation by stabilizing a transient amorphous mineral precursor. *J Biol Chem*. **295**:1943–1959.

Soderling JA, Reed MJ, Corsa A, Sage EH. 1997. Cloning and expression of murine SC1, a gene product homologous to SPARC. *J Histochem Cytochem*. **45**:823–835.

Tagliabracci VS, Engel JL, Wen J, Wiley SE, Worby CA, Kinch LN, Xiao J, Grishin NV, Dixon JE. 2012. Secreted kinase phosphorylates extracellular proteins that regulate biomineralization. *Science* **336**:1150–1153.

Ustriyana P, Schulte F, Gombedza F, Gil-Bona A, Paruchuri S, Bidlack FB, Hardt M, Landis WJ, Sahai N. 2021. Spatial survey of non-collagenous proteins in mineralizing and non-mineralizing vertebrate tissues *ex vivo*. *Bone Rep*. **14**:100754.

Venkatesh B, Lee AP, Ravi V, Maurya AK, Korzh V, Lim ZW, Ingham PW, Boehm T, Brenner S, Warren WC. 2014. On the origin of *SCPP* genes. *Evol Dev*. **16**:125–126.

Venkatesh B, Lee AP, Ravi V, Maurya AK, Lian MM, Swann JB, Ohta Y, Flajnik MF, Sutoh Y, Kasahara M, *et al.* 2014. Elephant shark genome provides unique insights into gnathostome evolution. *Nature* **505**:174–179.

Wagih O. 2017. ggseqlogo: a versatile R package for drawing sequence logos. *Bioinformatics* **33**:3645–3647.

Wang SK, Samann AC, Hu JC, Simmer JP. 2013. FAM20°C functions intracellularly within both ameloblasts and odontoblasts in vivo. *J Bone Miner Res*. **28**:2508–2511.

Xu D, Pavlidis P, Thamadilok S, Redwood E, Fox S, Blekhman R, Ruhl S, Gokcumen O. 2016. Recent evolution of the salivary mucin MUC7. *Sci Rep*. **6**:31791.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. **24**:1586–1591.

Yonekura T, Homma H, Sakurai A, Moriguchi M, Miake Y, Toyosawa S, Shintani S. 2013. Identification, characterization, and expression of *dentin matrix protein 1* gene in *Xenopus laevis*. *J Exp Zool B Mol Dev Evol*. **320**:525–537.

Zhang H, Zhu Q, Cui J, Wang Y, Chen MJ, Guo X, Tagliabracci VS, Dixon JE, Xiao J. 2018. Structure and evolution of the Fam20 kinases. *Nat Commun*. **9**:1218.

Zhou H, Visnovska T, Gong H, Schmeier S, Hickford J, Ganley ARD. 2019. Contrasting patterns of coding and flanking region evolution in mammalian keratin associated protein-1 genes. *Mol Phylogenet Evol*. **133**:352–361.