



**HAL**  
open science

# Origine évolutive de la bio-minéralisation chez les vertébrés : contributions de la biologie squelettique et développementale des chondrichthyens

Nicolas Leurs

► **To cite this version:**

Nicolas Leurs. Origine évolutive de la bio-minéralisation chez les vertébrés : contributions de la biologie squelettique et développementale des chondrichthyens. Evolution [q-bio.PE]. Université de Montpellier, 2022. Français. NNT : 2022UMONG060 . tel-04135694

**HAL Id: tel-04135694**

**<https://theses.hal.science/tel-04135694v1>**

Submitted on 21 Jun 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

Spécialité : Écologie, Évolution, Ressources Génétiques, Paléobiologie

École doctorale GAIA

Institut des Sciences de l'Évolution de Montpellier

Origine évolutive de la bio-minéralisation chez les  
vertébrés : contributions de la biologie squelettique  
développementale des chondrichthyens

Présentée par Nicolas LEURS

Le 18 novembre 2022

Sous la direction de Mélanie Debiais-Thibaud et Camille Martinand-Mari

Devant le jury composé de

Paul Eckhard WITTEN, Professeur, Université de Ghent

Didier CASANE, Professeur, Université Paris-Saclay

Camille BERTHELOT, Chargée de recherche, Institut Pasteur

Donald DAVESNE, Chercheur, Museum für Naturkunde

Sophie PANTALACCI, Directrice de recherche, ENS Lyon

Frédéric DELSUC, Directeur de recherche, ISEM

Mélanie Debiais-Thibaud, Maître de conférences, ISEM

Camille Martinand-Mari, Chargée de recherche, ISEM

Rapporteur

Rapporteur

Examinatrice

Examineur

Examinatrice

Président du jury

Directrice de thèse

Directrice de thèse



UNIVERSITÉ  
DE MONTPELLIER



# Remerciements

J'aimerais commencer par remercier mes directrices de thèse Mélanie DEBIAIS-THIIBAUD et Camille MARTINAND-MARI pour un encadrement exceptionnel. Votre humanité, intégrité et bienveillance hors pair, m'ont permis de réaliser ma thèse dans un cadre où je me suis senti soutenu autant scientifiquement que moralement. Chanceux seront ceux, qui comme moi, auront la possibilité de travailler sous votre tutelle. Merci.

Je tiens à remercier sincèrement mes rapporteurs Paul Eckhard WITTEN et Didier CASANE ainsi que les autres participant à l'évaluation de cette thèse: Frédéric DELSUC, Camille BERTHELOT, Sophie PANTALACCI et Donald DAVESNE. J'espère que vous aurez tous trouvé dans ce manuscrit des éléments pour nourrir votre intérêt.

Les travaux menés au cours de ces trois ans n'auraient pas abouti sans l'aide de Stéphanie VENTEO et Marie SEMON, qui grâce à leur expertise et à leur patience m'ont permis de développer de nouvelles compétences pratiques, essentielles pour le bon déroulement de ma thèse. Merci également à Mari-ka TILAK pour son aide et sa patience au laboratoire.

Je voudrais aussi remercier en particulier Alexandre ASSEMAT. Depuis le début de nos thèses un concert de métal a basculé notre relation de collègues vers une vraie amitié à laquelle je tiens particulièrement. On a pu s'épauler, geeker et rigoler ensemble pendant trois bonnes années, et encore bien d'autres, je l'espère.

Je voudrais aussi remercier ma famille, ma mère Fabienne et mes deux grands frères Christophe et Alexandre. Malgré la distance, vous avez toujours su me soutenir, m'encourager et m'aimer, chacun à votre manière. Vous êtes ma source d'inspiration et de bonheur. Sachez que vous avez une part non négligeable de responsabilité sur le travail présenté dans ce manuscrit et je vous en serais toujours reconnaissant.

Kim, mi Amor, je te remercie du fond de mon cœur pour la stabilité que tu apportes à ma vie. Merci pour ta complicité, ta joyeuseté, et pour le soutien inconditionnel que tu me donnes sans le savoir. Surtout merci pour les fous rires quotidiens parce qu'on est, tu le sais, aussi bêtes l'un que l'autre.

¡¡Gracias a todos!!



*“Así es la vida de caprichosa, a veces negra y a veces color rosa.  
Así es la vida jacarandosa, te quita, te pone, te sube, te baja, y a  
veces te lo da... ¡Así es la vida!”*

*– Elefante*



# Contents

<b>Table of contents</b>	<b>i</b>
<b>List of figures</b>	<b>v</b>
<b>List of tables</b>	<b>vii</b>
<b>List of abbreviations</b>	<b>ix</b>
<b>Preamble: Epistemology of Evo-Devo</b>	<b>xi</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Vertebrate skeletal diversity . . . . .	1
1.1.1 Cyclostomes . . . . .	3
1.1.2 Armoured fishes . . . . .	5
1.1.3 Extant jawed vertebrates . . . . .	7
1.1.4 What about the exoskeleton? . . . . .	11
1.1.5 Evolution of skeletal mineralization . . . . .	12
1.2 Skeletal development dynamics . . . . .	12
1.2.1 Endochondral and intramembranous ossification . . . . .	12
1.2.2 Extracellular matrix composition . . . . .	14
1.2.3 Extracellular matrix mineralization . . . . .	15
1.2.4 Odontode development and matrix proteins . . . . .	16
1.2.5 Outside bony fishes . . . . .	16
1.3 Chondrichthyans unchained . . . . .	17
1.3.1 Development of the tessellated skeleton . . . . .	17
1.3.2 Developmental genetics in chondrichthyans . . . . .	21
1.3.3 How to identify mineralization genes in chondrichthyans . . . . .	23
1.4 Outline of the thesis . . . . .	24
<b>2 Bony fish mineralization genes in the catshark</b>	<b>27</b>
2.1 Introduction . . . . .	27

2.2	Journal article — Evolution of Matrix Gla and Bone Gla Protein Genes in Jawed Vertebrates . . . . .	29
2.3	Journal article — Divergent Expression of <i>Sparc</i> / <i>Sparc-L</i> / and <i>Scpp</i> Genes During Jawed Vertebrate Cartilage Mineralization . . . . .	47
2.4	Journal article — Parallel Evolution of Ameloblastic <i>scpp</i> Genes in Bony and Cartilaginous Vertebrates . . . . .	58
2.5	Discussion: Candidates genes and comparative analyses in Evo-Devo . . . . .	69
<b>3</b>	<b>The transcriptomic approach that was supposed to work better</b>	<b>71</b>
3.1	Introduction . . . . .	71
3.2	Materials and methods . . . . .	72
3.2.1	Transcriptomic data generated by the Sanger’s Institute . . . . .	72
3.2.2	Mineralization-focused RNAseq data . . . . .	73
3.2.3	mRNA <i>in situ</i> Hybridization . . . . .	74
3.2.4	Phylogenetic reconstructions . . . . .	74
3.2.5	Synteny analyses . . . . .	74
3.3	Results and Discussion . . . . .	74
3.3.1	Bulk vertebra RNAseq characterization . . . . .	74
3.3.2	Sorting RNAseq data to extract skeleton-relevant data . . . . .	78
3.3.3	Listing candidate genes . . . . .	79
3.3.4	Case studies: <i>clec3a</i> , <i>otos</i> , <i>spp2</i> . . . . .	84
3.4	Concluding remarks . . . . .	91
<b>4</b>	<b>The genes that do not want to fall into ranks</b>	<b>93</b>
4.1	Introduction . . . . .	93
4.2	Genomic and transcriptomic characterization of the <i>slrp</i> family in <i>S. canicula</i> . . . . .	94
4.2.1	Introduction . . . . .	94
4.2.2	Materials and Methods . . . . .	95
4.2.3	Results . . . . .	99
4.2.4	Discussion . . . . .	108
4.2.5	Concluding remarks . . . . .	113
4.3	Expansion of Mmp proteins in cartilaginous fishes . . . . .	114
4.3.1	Introduction . . . . .	114
4.3.2	Material and Methods . . . . .	115
4.3.3	Results . . . . .	116
4.3.4	Discussion . . . . .	122
4.4	Discussion: Evolutionary dynamics of clustered gene families . . . . .	126

<b>5 Conclusion: Evolution of the biomineralization toolkit in gnathostomes</b>	<b>129</b>
5.1 O tissue, who art thou . . . . .	129
5.2 Cell homology conundrum . . . . .	130
5.3 The biomineralization toolkit of gnathostomes . . . . .	132
<b>Résumé étendu en français</b>	<b>139</b>



# List of Figures

1.1	Vertebrate phylogeny . . . . .	2
1.2	Cyclostome skeletal elements . . . . .	4
1.3	Ostracoderm headshield . . . . .	6
1.4	Palcoderm and osteichthyan bone . . . . .	8
1.5	Elasmobranch endoskeleton . . . . .	11
1.6	Types of ossification in bony fishes . . . . .	13
1.7	Tooth morphogenesis . . . . .	17
1.8	<i>S. canicula</i> mineralized tissues . . . . .	19
1.9	Selachian centra . . . . .	20
2.1	Preliminary ISH assays for <i>alp</i> , <i>fam20c</i> and <i>scpp</i> . . . . .	70
3.1	RNAseq step-wise methods . . . . .	75
3.2	Enrichment analysis 1 . . . . .	76
3.3	Deconvolution results . . . . .	77
3.4	Enrichment analysis 2 . . . . .	79
3.5	TPM patterns of up-regulated putative skeletal markers . . . . .	83
3.6	Histology and ISH of selected DEGs . . . . .	84
3.7	<i>clec3</i> phylogeny . . . . .	85
3.8	<i>otos</i> phylogeny . . . . .	86
3.9	<i>spp2</i> phylogeny . . . . .	88
3.10	<i>spp2</i> synteny . . . . .	89
3.11	<i>spp2</i> TPMs in adult tissues . . . . .	90
4.1	Whole <i>slrp</i> phylogeny . . . . .	100
4.2	<i>slrp</i> synteny . . . . .	102
4.3	sub-sampled <i>slrp</i> phylogenies . . . . .	103
4.4	Histology and collagen ISH . . . . .	106
4.5	<i>slrp</i> ISH . . . . .	107
4.6	<i>slrp</i> evolutionary scenario . . . . .	110
4.7	Whole <i>mmp</i> phylogeny . . . . .	118
4.8	<i>mmp21</i> phylogeny . . . . .	119

4.9	Clustered mmp phylogeny . . . . .	120
4.10	Clustered <i>mmp</i> synteny . . . . .	121
4.11	Ancestral state of clustered <i>mmps</i> . . . . .	125
5.1	Skeletal cells and tissues evolution: null hypothesis . . . . .	134
5.2	Skeletal cells and tissues evolution: maximum parsimony . . . . .	135
5.3	Skeletal cells and tissues evolution: proposed scenario . . . . .	137

# List of Tables

3.1	List of screened DE genes . . . . .	80
3.2	List of screened DE genes bis . . . . .	81
4.1	TPM values of canonical <i>slrps</i> . . . . .	105
4.2	Recap of <i>slrp</i> expression in <i>S.canicula</i> tissues / cells . . . . .	108
4.3	TPM values of <i>S. canicula mmp</i> genes . . . . .	123
4.4	TPM values of <i>S. canicula mmp</i> genes bis . . . . .	124
5.1	Described expression of genes studied in this manuscript . . . . .	133
5.2	Ancestral character reconstruction along the maximum parsimony reconstruction	136



# List of Abbreviations

- BLAST: Basic Local Alignment Search Tool
- bp: base pair
- cDNA: complementary deoxyribonucleic acid
- DE: Differential expression
- DEG: Differentially expressed genes
- ECM: Extracellular matrix
- Evo-Devo: Evolutionary developmental biology
- GRN: Gene regulatory network
- ISH: *In situ* hybridization
- kb: kilo-bases
- LRR: Leucine-rich repeat
- LRRCE: LRR C-terminal ear repeat
- mRNA: messenger ribonucleic acid
- NCP: Non-collagenous proteins
- RT-qPCR: Real time quantitative polymerase chain reaction
- TL: Total length
- TPM: Transcripts per kilobase million
- WGD: Whole genome duplications



# Preamble: Epistemology of Evo-Devo

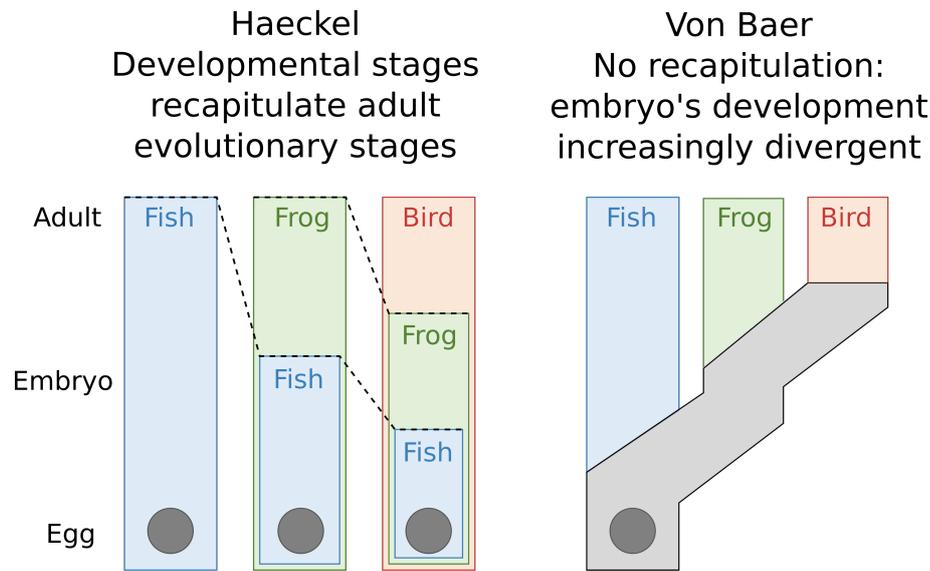
## Evolutionary developmental biology

For centuries researchers have been trying to understand how animals develop into very different forms, how they change during their development and what are the processes underlying these changes. This study has become the field of evolutionary developmental biology (Evo-Devo) and its history is full with debate, conundrums and ever changing theories, similar to the animal development they try to recapitulate.

## Recapitulation

Johann Friedrich Meckel was a German pathologist, anatomist, zoologist and professor at the University of Halle, Germany. Together with the French embryologist Étienne Serres, they defined the “Meckel-Serres Law”: a theory of parallelism between embryonic developmental stages and a hierarchical structure of life: the *scala naturae* (Ernst and Serres, 1828). This law proposed that “higher” individuals in the scale of nature had to go through developmental stages that represent “lower” individual in the scale, such that a frog, considered higher (or more “evolved”) in the scale than a fish, had to go through a “fish developmental phase” before actually becoming a frog. The law was further revised by Ernst Haeckel rejecting the proposed “replay” of the *scala naturae* during development. He proposed the term *heterochrony* (i.e. differences in the timing, rate or duration of developmental processes between organisms) and incorporated evolutionary principles into a “Biogenic Law”, often described using his famous phrase “ontogeny recapitulates phylogeny” (Haeckel, 1866). Haeckel believed that evolutionary stages were repeated in the embryonic development of an animal (figure below). This theory was largely debated and opposed mainly by Karl Ernst Von Baer who posited that embryos started from one basic form, similar in different animals, and then they would develop in a branching pattern towards increasingly different forms (figure below; Abzhanov, 2013; Richtsmeier, 2018). The early work of Haeckel and Von Baer largely stimulated the growth of evolutionary developmental biology and gave rise to two main hypotheses on the similarity of developmental stages between species: the Funnel model (most similar very early, then divergent) and the Hourglass model (divergent early, most similar in the middle and then divergent again later in

development), both of which have some degree of morphological and molecular support (Irie and Kuratani, 2011).



Haeckel vs Von Baer; redrawn diagram based on the original from Ian Alexander

## Evolutionary morphology

Through most of the 19th and 20th century, zoologists described how animals developed into adults displaying very different body plans but they knew little about the processes behind embryonic development or how they had evolved. As Sean B. Carroll said it in the *Origin of Forms* “Biologists could say, with confidence, that forms change, and that natural selection is an important force for change. Yet they could say nothing about how that change is accomplished. How bodies or body parts change, or how new structures arise, remained complete mysteries.” In the book *On the Origin of Species*, Charles Darwin argued that animals sharing similar embryonic structures would imply a shared common ancestor. An example of this is the finding of Alexander Kovalevskij (Kovalevskij, 1866) that tunicates were not molluscs, because in their larval stage they presented a notochord and pharyngeal slits. Therefore tunicates, sharing embryonic features with other chordates should be classified into said group. Today, chordates are still a valid group which includes vertebrates. Thus, embryology became an evolutionary science, classifying species based on homologies observed during development. However, without molecular evidence research stalled for several decades (Gilbert, 2003).

## Debut of developmental genetics

During the second half of the 20th century, with the arrival of recombinant DNA, bioinformatics and genomics, developmental biology experienced a revolution. In 1978, Edward B. Lewis

discovered homeotic genes that regulate the embryonic development in *Drosophila* flies. Quickly after, William McGinnis, together with Michael S. Levine, discovered similar gene sequences in other phyla such as vertebrates: the homeobox genes (McGinnis et al., 1984). This discovery was the first molecular evidence of strong similarities in genes that controlled development across all eukaryotes. Additional similarities were later found at deep phylogenetic scales: the *distal-less* gene was found to be involved in the development of the limbs in fruit flies (Cohen and Jürgens, 1989) and its vertebrate ortholog in the development of chicken wings and fish fins (Panganiban et al., 1997). Evo-Devo was finally starting to uncover how animal bodies were being built and changing during ontogeny at the molecular level.

## Evolutionary syntheses

In the early 20th century, before the advances in developmental genetics, the so-called modern synthesis was being formulated. The idea was to bring together Darwinian natural selection, population-level thinking and mendelian inheritance into a joint mathematical framework. This synthesis, widely accepted, was clear, neat and simple but it largely ignored development as a source of phenotypic variation. Although Gavin de Beer argued in his book *Embryos and Ancestors* (1930) that evolution could occur by heterochrony (Gould, 1977), the modern synthesis still did not take into account embryonic development to explain the forms of organisms. During the 1950s several biologists such as Conrad H. Waddington and Rupert Riedl called for an extended synthesis based on their work in epigenetics and evolvability, respectively. These subjects and many others, including Evo-Devo, have presented a growing number of challenges to the classical model of evolution. In response to these challenges, an extended evolutionary synthesis was proposed that takes into account the plurality of factors and causal relations in evolutionary processes. Notably, in this synthesis, development assumes a constructive role, and natural selection is not the only way through which variation in populations can be modified (Müller, 2017).

## The more the merrier

More recently, developmental genetics has boomed thanks to next-generation sequencing technologies, but within vertebrates, functional genetics still focus on few, “model species” including the mouse and the zebrafish. This has partially hindered the scale at which research in vertebrate Evo-devo is carried out, since the discipline relies on comparative analyses between species, and between these two species, one can only infer evolutionary events at the taxonomical scale of bony fishes. To understand homologies at the jawed vertebrate scale, studies on cartilaginous fish would be needed, and in a similar way, studies on cyclostomes are necessary for understanding Evo-Devo at the vertebrate scale. This trend has begun to change, since more and more researchers take an interest for the so-called non-model species. Evo-devo is

now being studied at the chordate or even at the bilaterian scale. We are at a stage where we can start to trace back homologies at very deep evolutionary scales, infer ancestral characters and understand how they came to be (Tarazona et al., 2016).

In this manuscript, I will be presenting my Evo-Devo work on cartilaginous fishes using the small-spotted catshark *Scyliorhinus canicula* as a model species. My work was aimed at better understanding how the skeleton of the catshark develops from a developmental genetics standpoint. Then through comparative analyses with bony fishes I tried to contribute to our understanding of the evolution of the skeleton and the mineralized tissues among jawed vertebrates.

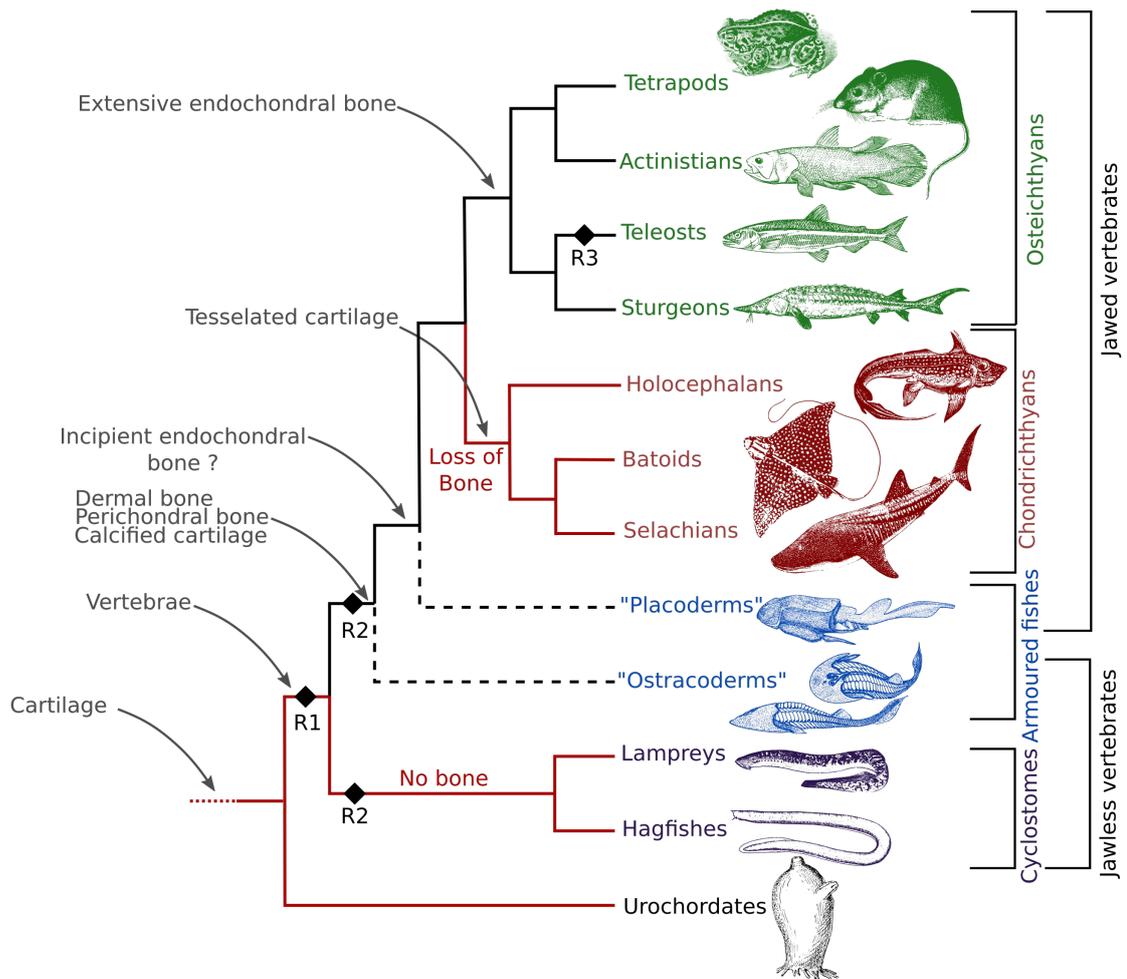
# Chapter 1

## Introduction

### 1.1 Vertebrate skeletal diversity

Vertebrates as a phylum was first proposed by Lamarck in 1794 (Lamarck, 1794), distinguishing them from invertebrates. The clade is partially defined by a major morphological feature: the presence of a vertebral column (Fig 1.1). This structure is made up of segmentally arranged cartilaginous elements (or bony nodules) associated with the notochord, called vertebrae (Arratia et al., 2001; Janvier, 1996; Kardong, 1997; Walker, Liem, et al., 1994). Vertebrae are what we call a “composite” organ. In general it is formed from several parts including the neural arches that surround the spinal chord, the centrum which develops around the notochord and, more posteriorly in the body, there can be hemal arches which will protect veins and arteries going to the tail and back (Fig 1.2e, g). These structures develop and fuse together throughout development to form the vertebrae (B. Hall, 2014b; Kardong, 1997). This internalized vertebral skeleton, being an evolutionary innovation of vertebrates (Shimeld and Holland, 2000) is key to understand the evolution of the clade.

Vertebrates are a rich and diverse group that diverged from other chordates at around 600 million years ago (S. Kumar et al., 2017) and underwent two rounds of whole genome duplications (Fig 1.1; Ohno, 1999). Since then, the skeleton has diversified within the clade giving rise to novel molecules, cell-types, tissues and even modes of mineralization. Therefore, understanding the evolution of the vertebrate skeleton allows us to understand vertebrate evolution as well. In the following I will be discussing the phylogeny of vertebrates in the light of their skeletal characteristics. I will be differentiating two types of skeletons: the internal skeleton (or endoskeleton) and the external skeleton (or exoskeleton). This distinction has been useful in paleontology and Evo-Devo although the definition of each term is not fully accepted by all authors. In particular, dermal bone (a type of bone that develops within the dermis and not from cartilage) has traditionally been regarded as part of the exoskeleton by paleontologists (Janvier, 1996), but from a developmental standpoint it is more accurate to include it as a part of the endoskeleton (see B. Hall, 2014a for a review on these terms). Therefore, for all



**Figure 1.1:** Vertebrate phylogeny and evolution of skeletal mineralization. Dotted branches correspond to extinct lineages. Groups in quotation marks are considered paraphyletic. Black diamonds represent events of whole genome duplication.

intents and purposes, I will consider the exoskeleton as all structures in direct contact with the environment, which include teeth and scales, while the endoskeleton will include all the other skeletal tissues such as various types of bone (e.g. dermal) and cartilage.

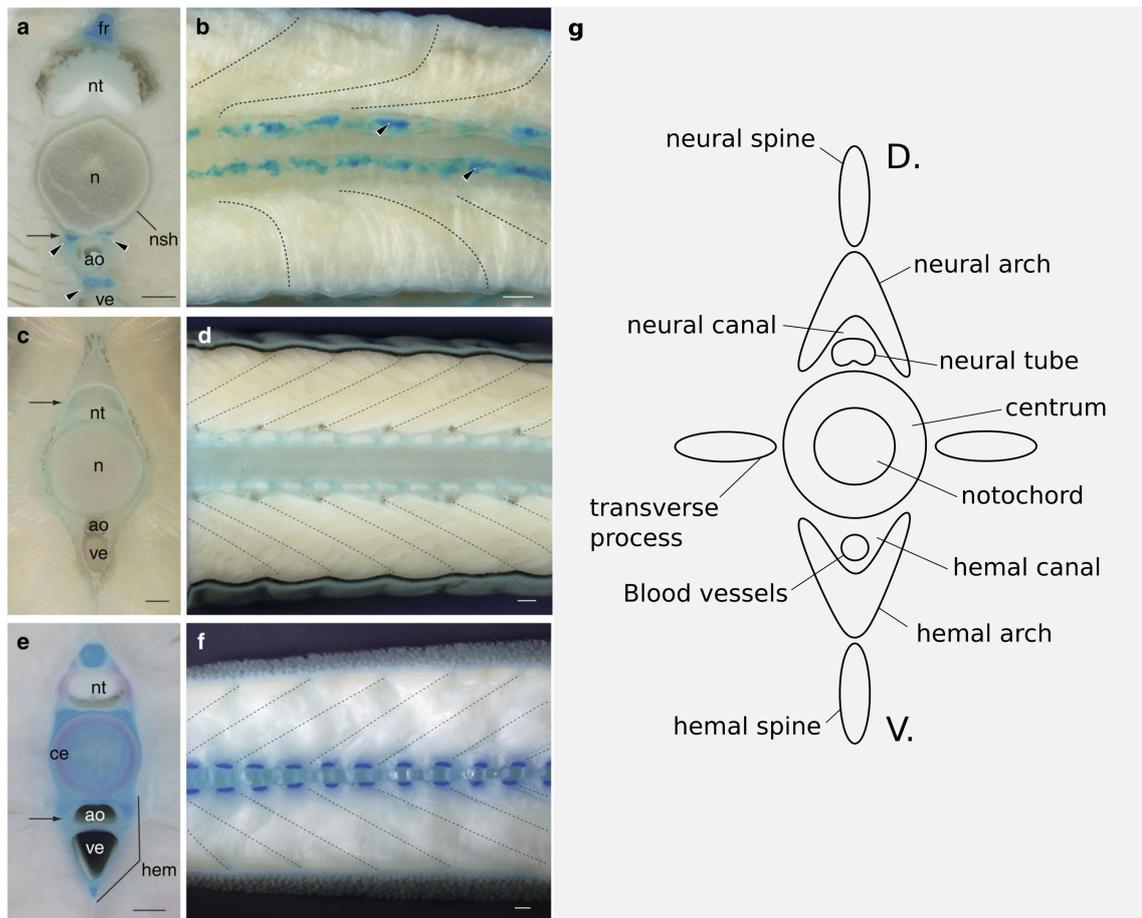
### 1.1.1 Cyclostomes

Cyclostomes are a vertebrate clade made up of both lampreys and hagfishes which represent the only lineages of extant jawless vertebrates (Fig 1.1). Historically, the monophyly of the cyclostomes was subject of long and intense debates because of discrepancies between morphological and molecular data (Haeckel, 1892; B. K. Hall, 1999; Janvier, 1996). However, it is currently accepted that both hagfishes and lampreys form a single clade sister to jawed vertebrates (the cyclostome hypothesis; see Miyashita et al., 2019).

Hagfishes have been the most controversial in terms of their belonging in the vertebrate clade. They have a very complex skull made up of cartilaginous bars with no proper braincase. The post-cranial skeleton consists of a notochord reaching the midbrain anteriorly and the caudal fin posteriorly (Janvier, 1996). Therefore, for a long time it was believed that hagfishes lacked the axial skeletal system in the form of vertebral elements (Kardong, 1997; Walker, Liem, et al., 1994). Assuming the monophyletic hypothesis of cyclostomes, it was unclear whether the hagfish skeletal morphology was an ancestral or a derived condition. However, vertebra-like elements have been reported in the hagfish *Eptatretus burgeri*, arising as small sclerotome-derived cartilaginous vertebral elements under the notochord (Fig 1.2a, b), i.e. the same anatomical positions as in jawed vertebrates (Fig 1.2e, f). These elements are evidenced by an alcian blue staining, that specifically marks cartilage. The authors suggested that these nodules were homologous to the gnathostome (jawed vertebrate) vertebrae, although they do not show a classical segmented pattern (Ota et al., 2011). Later, they were able to show that these nodules undergo a suppression of the chondrification process in the mid-trunk region during embryonic development (Ota et al., 2014), suggesting that the loss of clear segmented vertebral elements is a derived condition in hagfishes rather than an ancestral vertebrate state.

Lampreys, the other cyclostome clade, share more similarities with the rest of the vertebrates. Since the lamprey larvae (ammocoetes) are quite similar to some cephalochordates (amphioxus), the life history of a lamprey was thought to recapitulate the origin of vertebrates (Miyashita et al., 2021). However, this larval stage, has recently been shown to be a derived feature of modern lampreys (Miyashita et al., 2021). Similar to hagfishes, lampreys possess a cartilaginous skull with a more developed braincase that includes lateral walls and transverse cartilaginous bars. They display a large notochord where, dorsally, a series of cartilaginous nodes, called arcualia, can be observed (Fig 1.2c, d; Janvier, 1996). These structures are considered to be rudimentary vertebral elements and have been suggested to be homologous to neural arches in extant jawed vertebrates (Romer, Parsons, et al., 1978).

Although these clades present very derived skeletal features, it is still possible to see the



**Figure 1.2:** (a–f) Alcian blue-stained transverse and horizontal sections of a whole-mount hagfish (*E. burgeri*; (a): transverse, (b): horizontal), a lamprey (*Lethenteron japonicum*; (c): transverse, (d): horizontal) and a catshark (*Galeus nipponesis*; (e): transverse, (f): horizontal). Cartilaginous nodules are indicated by black arrowheads in (a) and (b). The levels of the horizontal sections in each specimen are indicated by arrows in (a), (c) and (e). ao, dorsal aorta; ce, centrum; fr, fin ray; hem, hemal arch; n, notochord; nsh, notochordal sheath; nt, neural tube; ve, vein. Scale bars, 1 mm. Modified from Ota et al., 2011. (g) The ideal vertebra, redrawn from Criswell et al., 2017. D. Dorsal; V. Ventral

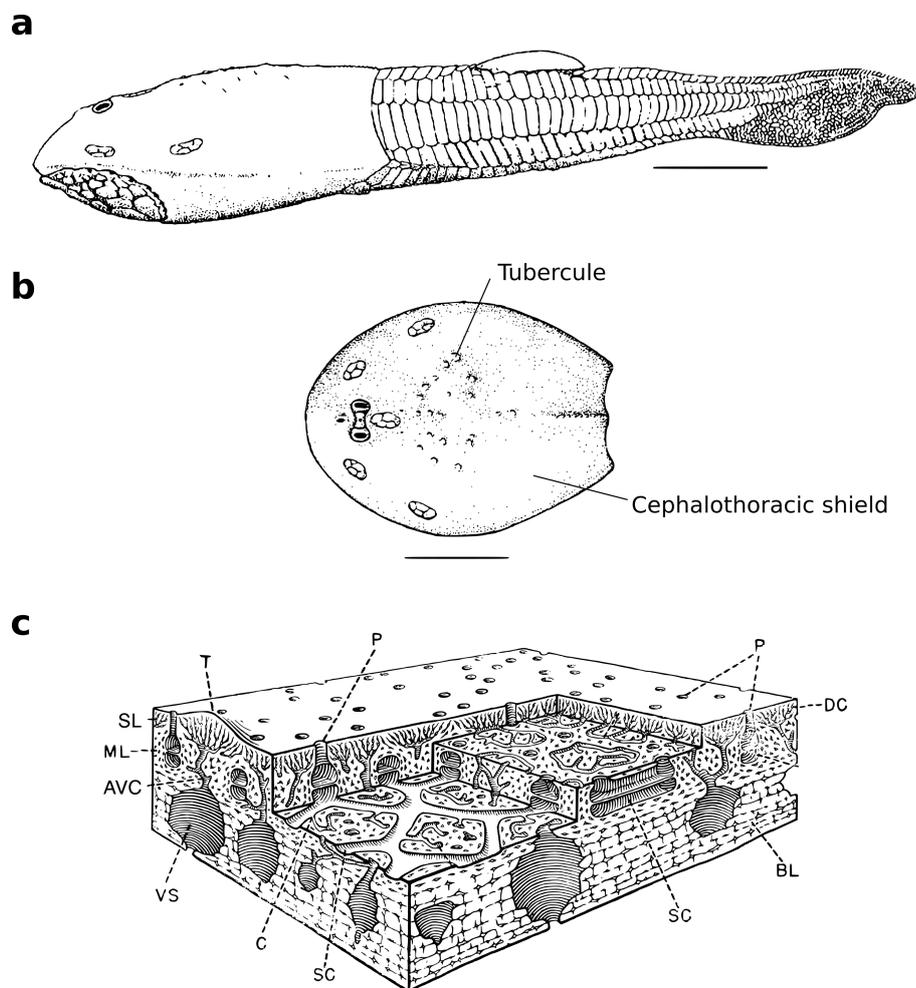
remnants of segmental cartilaginous elements in hagfish embryos, and clear rudimentary cartilaginous nodules in adult lampreys. Coupled with conserved cartilaginous crania, these features reflect an ancestral condition common to all vertebrates: a cartilaginous axial skeleton, made of anterior structures associated with a brain and posterior repeated structures associated with a notochord.

### 1.1.2 Armoured fishes

Fossil vertebrates denote a paraphyletic group of armoured fishes that appeared first during the early Silurian (Janvier, 1996), after the split of extant cyclostomes. They can be divided into two (again paraphyletic) groups depending on whether they possess a jaw or not (Fig 1.1).

The fossil jawless vertebrates are often loosely grouped as “ostracoderms” a paraphyletic group that does not represent a single evolutionary lineage (Fig 1.1; Benton, 2014). However, I will be using this term because the exact phylogenetic relationships of ostracoderms are not needed to describe the diversity of skeletal tissues they used to display. Ostracoderms are composed of anaspids, galeaspids, heterostracomorphs, osteostracans (e.g. Fig 1.3a, b) and thelodonts (Donoghue and Sansom, 2002). The lack of axial skeleton preservation in fossil jawless vertebrates is generally interpreted as being cartilaginous (Janvier, 1996; Stensiö et al., 1927) as in cyclostomes. However, in contrast to the “naked” cyclostomes, ostracoderms had an extensively mineralized cephalic skeleton made up of dermal bone (Fig 1.3b, c; Janvier, 1996). In addition to this dermal bone, other types of mineralized tissues have been described in ostracoderms. For instance, in an anaspid specimen from the Late Devonian, all the internal skeletal elements seem to be made up of calcified cartilage (Janvier and Arsenault, 2002). Mineralized plaques of calcified cartilage in the heads of other ostracoderms have also been described (Denison, 1967). Some groups have been shown to possess perichondral bone (i.e. bone that develops superficially onto cartilaginous elements) limited to the head (osteostracans and galeaspids) and the shoulder girdle (osteostracans) (Janvier, 1996). Additionally, in contrast to the typical cellular bone of other ostracoderms, the dermal skeleton of galeaspids and heterostracomorphs has been shown to be composed of acellular bone (bone lacking expected cell lacunae; Janvier, 1990; Min and Janvier, 1998). Ostracoderms represent the oldest known fossils displaying instances of mineralized skeletons with a high diversity of tissues including calcified cartilage, dermal and perichondral bone as well as acellular bone, suggesting a very quick and concomitant evolution of a variety of mineralized tissues in early vertebrates (Donoghue and Sansom, 2002).

The other class of armoured fishes are the placoderms (e.g. Fig 1.1 and Fig 1.4a), that distinguish themselves from ostracoderms by possessing both a dermal skeleton, covering the head and the anterior part of the trunk, and a jaw (Janvier, 1996) making them belong to the clade of gnathostomes. Placoderms are further characterized by an axial and appendicular skeleton (another gnathostome shared derived character) made up of cartilage lined with



**Figure 1.3:** Reconstruction of the cephalothoracic headshield and complete external anatomy in the osteostracan *Tremataspis mammillata*; (a) Lateral view of the whole body and (b) dorsal view of the headshield; modified from O'Shea et al., 2019. (c) Block diagram depicting the histology of the cephalic dermal skeleton, modified from Denison, 1947. Abbreviations: AVC, ascending vascular canal; BL, basal layer; C, connection between sensory and vascular canal systems; DC, dentine-like canals; ML, middle layer; P, pore of sensory canal system; SC, sensory canal; SL, superficial layer; T, tubercle; VS, vascular cavity for vascular sinus. Scale bars: a,b. 10mm.

perichondral bone (Denison, 1978; Donoghue and Sansom, 2002; Janvier, 1996). Although the monophyly of the clade is still debated today (King et al., 2017; Q. Li et al., 2021; Y.-a. Zhu et al., 2021), placoderms show the first instance of a mineralized vertebral skeleton (Fig 1.4a; but see Chevrinai et al., 2018). Placoderms do not seem to present novel ways of mineralizing their skeletons compared to fossil jawless vertebrates (Giles et al., 2013), even though research on placoderm dermal armour has been ongoing since the 19th century (e.g. Agassiz, 1844; Gürich, 1891). However, a new instance of mineralized tissue was only recently described in a Late Devonian placoderm (Brazeau et al., 2020): the endochondral bone (Fig 1.4b, c). This type of bone appears not only ossified perichondrally, but also endochondrally (i.e. internally) in the form of a “spongy” or porous bone, by replacement of the initial cartilaginous tissue. This finding, if confirmed, completely changes our understanding of skeletal tissue evolution, as endochondral bone is currently considered a derived feature of bony fishes only. Additionally, placoderms display the oldest instances of mineralized vertebral elements. These elements termed *arcocentra* are formed by ossification of cartilage extending from the arcualia around the notochord. However, it has been shown that mineralized centra are not homologous among vertebrates, and have been independently acquired in different lineages (Arratia et al., 2001). However, the placoderm condition, including a persistent notochord and ossified arcocentra, seems to be the ancestral condition of gnathostomes (Arratia et al., 2001).

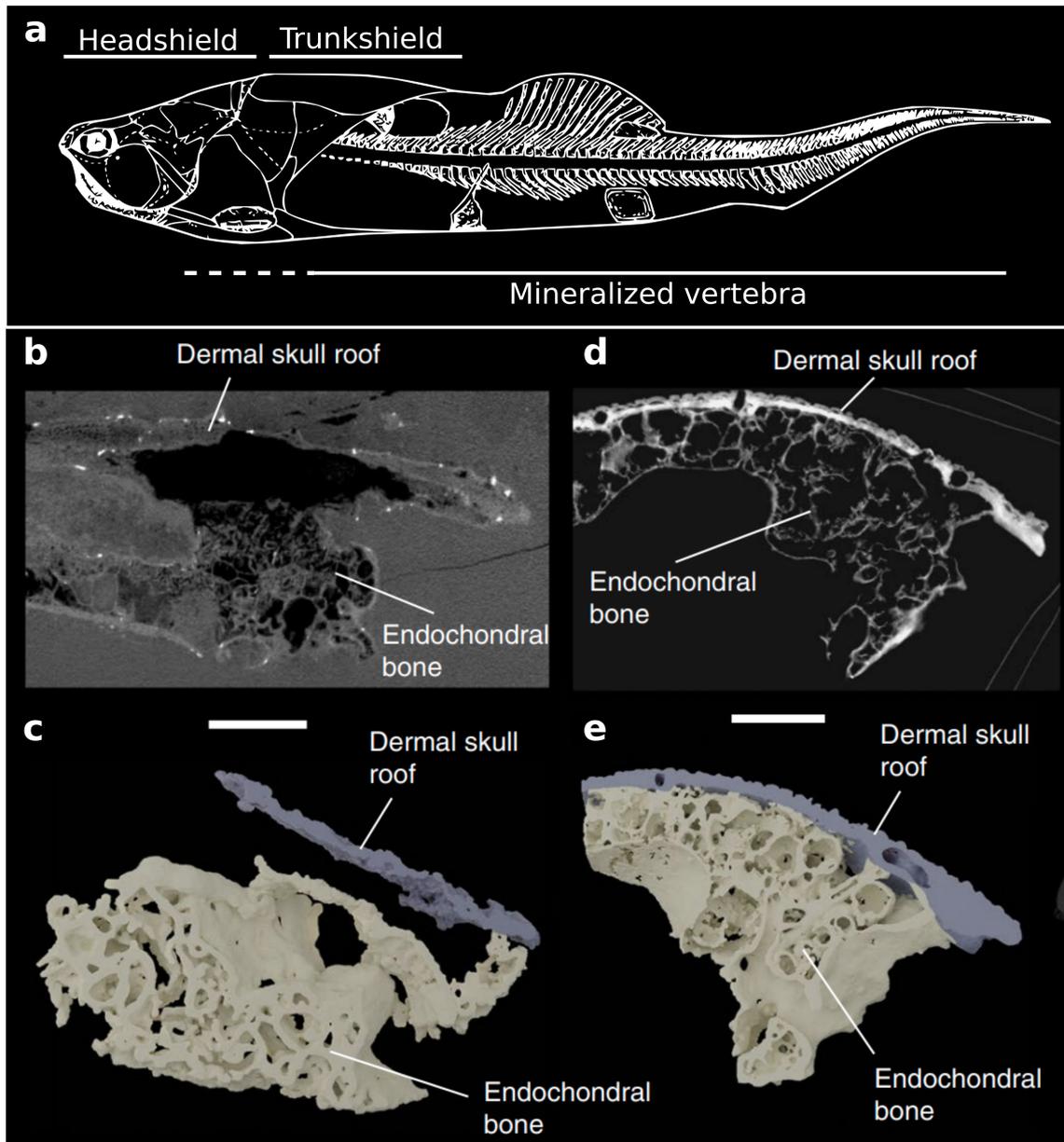
It is clear that centuries of vertebrate fossil descriptions account for an already diversified repertoire of mineralized tissues appearing in different anatomical structures. In the following I will focus on extant jawed vertebrates and their specific skeletal innovations.

### 1.1.3 Extant jawed vertebrates

Extant jawed vertebrates (gnathostomes) are divided into two monophyletic clades based on specific skeletal features. On the one hand bony fishes (osteichthyans) have a skeleton made up of mostly bone, and on the other hand cartilaginous fishes (chondrichthyans) have a skeleton made up of mostly mineralized and unmineralized cartilage devoid of bone (Fig 1.1).

#### Osteichthyans

Osteichthyans are the most diverse clade of vertebrates and have long been defined by a shared derived character (a synapomorphy): the endochondral bone (Fig 1.1 and Fig 1.4d,e). It is important to note that only one occurrence of endochondral bone in the headshield of an early vertebrate fossil has been reported, questioning the origin of endochondral bone (Brazeau et al., 2020). Although most of the skeleton of bony fishes is made out of endochondral bone, several parts are essentially made out of dermal bones such as the cranium (Fig 1.4d, e) and part of the pectoral girdle (Janvier, 1996). Osteichthyans are further divided into two lineages made up of actinopterygians (ray-finned fishes) on the one hand, and sarcopterigians (lobe-finned fishes) on the other.



**Figure 1.4:** (a) Reconstruction of the head, thoracic armour and post thoracic endoskeleton of a placoderm *Coccoosteus cuspidatus*; modified from Miles and Westoll, 1968. (b) A transverse tomographic slice through an Early Devonian placoderm (*Minjinia turgenensis*) and (c) Three-dimensional rendering of its trabecular bone structure. (d) A transverse tomographic section through the braincase of the extinct osteichthyan *Ligulalepis* (e) Three-dimensional rendering of the trabecular bone in *Ligulalepis*. Scale bars, 10mm (b,c); 1mm (d,e). Modified from Brazeau et al., 2020

Within actinopterygians there are several skeletal features that are important to mention when discussing their diversity. In particular this lineage displays considerable phenotypic variance among their vertebral elements (Arratia et al., 2001). Mainly, for this manuscript, some species display arcocentra such as in placoderms, but also chordocentra: a type of vertebral centrum that forms by mineralization of the middle fibrous part of the notochordal sheath. A major clade of actinopterygian is called teleosteans, and are characterized by a movable premaxilla, allowing them to protrude their jaws outwards from the mouth. In this clade, many species present a kind of acellular bone (anosteocytic bone), considered as a secondarily derived character different from ostracoderm acellular bone (Davesne et al., 2019; Kolliker, 1859). Additionally, it has been shown in the zebrafish that “classical” endochondral bones are quite reduced in number, and that a new type of bone called tubular bone has emerged as a derived feature in this species (Weigele and Franz-Odenaal, 2016). In the other clades of actinopterygians fishes, referred to as the non-teleostean fish group (Dornburg and Near, 2021), other kinds of specific skeletal features can be observed. For example, acipenseriformes (sturgeons and paddlefishes) display an extensively cartilaginous axial skeleton. Their cartilaginous skeleton displays a persistent notochord throughout life and cartilaginous arcualia (reviewed in Leprevost and Sire, 2014). Additionally, this clade presents extensively mineralized scutes made up of dermal bone.

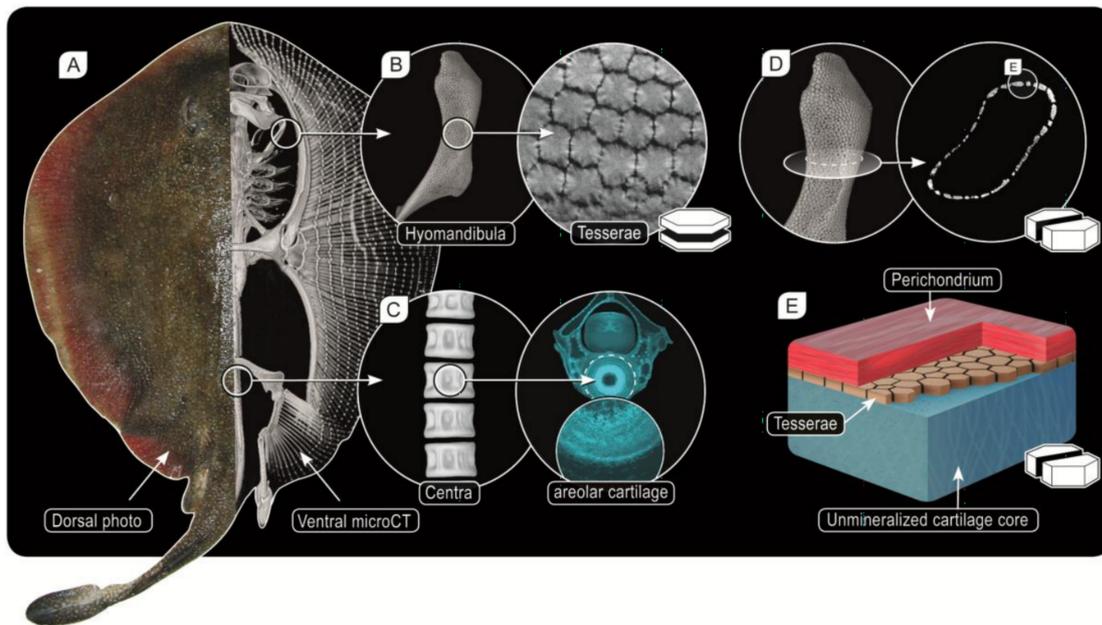
Sarcopterygians, the other clade of osteichthyans (Fig 1.1), are characterized by the structure of their paired fins, in which the appendicular skeleton articulates to the girdle by means of a single element (i.e. monobasal articulation; Janvier, 1996). Within sarcopterygians, tetrapods are a major clade, and are characterized by a very specific appendicular skeleton, no longer considered as a fin and made up of mostly long endochondral bones. In a non-tetrapod sarcopterygian lineage, the actinistians, represented today by *Latimeria chalumnae* (the coelacanth), the vertebral column consists of a large notochord surrounded by its fibrous sheaths and paired rows of cartilaginous elements (remnants of arcualia; Arratia et al., 2001). Within sarcopterygians, several clades display bony plaques covering most of the body called osteoderms. They can be found in lizards, crocodillians, frogs and even some mammals such as armadillos (Hill, 2006; Vickaryous et al., 2016; Vickaryous and Hall, 2008). Additionally, a unique bony structure arose in the order testudines: the turtle shell, made out of several endoskeletal bones, both dermal and endochondral, fused together and covered with keratin (Burke et al., 2015).

## Chondrichthyans

Chondrichthyans are the other major clade of vertebrates, sister group to osteichthyans and are divided into holocephalans (chimeras), batoids (rays and skates) and selachians (sharks), with the last two forming a clade called the elasmobranchs (Fig 1.1). Cartilaginous fishes, as their name indicates, are characterized by their cartilaginous skeleton devoid of bone (Janvier, 1996). Nonetheless, the skeleton of these organisms is stiffened, albeit only by a thin layer of calcified cartilage. The layer is made of a mosaic of small ( $\approx 100 \mu\text{m}$ ) polygonal tiles called tesserae, a

character stated to be “the critical defining character” for all chondrichthyans (Fig 1.5a, b, d and e; Grogan et al., 2012; Seidel et al., 2020). Most chondrichthyans develop chordacentra but in a different manner than actinopterygians. It is formed by cartilaginous cells that migrate from the neural arches into the fibrous sheath of the notochord that grow into a typical biconcave cartilage (Arratia et al., 2001). Therefore, the mineralized centra of chondrichthyans is considered not homologous to the one of bony fishes. The skeleton of chondrichthyans displays several types of derived calcification that have been classified based on their microstructure: areolar, globular and prismatic (Dean and Summers, 2006). Areolar mineralization (or alveolar; Moss, 1977) occurs in the vertebral chordacentra, and is laid down in concentric rings (Fig 1.5c; although it is not observed in holocephalans, Debais-Thibaud, 2019). Globular calcification (or spherulitic e.g. Janvier et al., 2004; Ørving, 1951) is a mineralized cartilaginous matrix formed by  $\approx 40\text{-}55$  nm spherules of mineral fused together. Prismatic calcification is always superficial vis-a-vis the cartilaginous elements and was defined by its capacity to specularly reflect light (i.e. in a mirror-like manner). Prismatic and globular calcification usually co-occur in the tesseræ while areolar calcification only occurs in the chordacentra (Dean and Summers, 2006). When considering chondrichthyan mineralized tissues, there is a clear bias in studies towards tesseræ and therefore towards globular and prismatic types of mineralization. This however is expected since most cartilaginous endoskeletal elements of selachians and batoids are covered by tesseræ (e.g., the jaws, fins, and most vertebral elements; Chaumel et al., 2020; Dean et al., 2009).

Tesseræ develop on the surface of cartilaginous elements but underneath the perichondrium (i.e. defined as the connective tissue surrounding cartilaginous elements; Fig 1.5e). They can be divided into two zones, the body (cartilage side) and the cap (perichondrium side), each of different composition and presenting cells of different shapes (Chaumel et al., 2020; Kemp and Westrin, 1979; Seidel, Blumer, Pechriggl, et al., 2017). Because of a lack of bone forming cells (but see Peignoux-Deville et al., 1982) and very derived forms of mineralization, chondrichthyans are thought to have lost the ability to make bone (Janvier, 1996). This hypothesis might be reinforced with the recent discovery of endochondral bone in a placoderm which would confirm that their mode of skeletal mineralization is a derived character instead of an ancestral one (Brazeau et al., 2020). However, a fourth type of mineralization, lamellar in histological sections, occurs in the neural arches of some chondrichthyans. It has been described and repeatedly compared to bone tissue, because of the presence of elongated cells enclosed in a collagen type I matrix similar to osteoblasts in bone, challenging the idea that this lineage has lost the ability to make bone (Atake et al., 2019; Eames et al., 2007; Peignoux-Deville et al., 1982).



**Figure 1.5:** Mineralized cartilage in the endoskeleton of elasmobranchs (sharks and rays). A) Photograph of round stingray *Urobatis halleri* (left), and CT image (right) showing most of the skeleton is mineralized. Two major types of calcified cartilage are described in elasmobranchs: B) tessellated calcified cartilage consisting of individually mineralized tiles called tesserae (shown here in planar view), covering each skeletal elements' surface and C) areolar calcified cartilage located in the centre of the vertebrae. D) In cross sections of skeletal elements (here the hyomandibula), tesserae can be seen in vertical view. E) General organization of the tissues forming elasmobranchs endoskeletons showing a mineralized, tessellated layer that is sandwiched between a hyaline cartilage core and an outer fibrous, connective tissue called perichondrium. Modified from Seidel, Blumer, Zaslansky, et al., 2017

### 1.1.4 What about the exoskeleton?

As explained above, the exoskeleton comprises mineralized structures that are in direct contact with the external environment, and includes teeth and scales. These structures are formed by different processes, cell and tissues compared to endoskeletal structures, therefore a separation should be made between these structures. It is important to note that teeth and especially scales are generic terms for several structures for which the homology relationships are not completely resolved (reviewed in Dhouailly et al., 2019). For instance, cyclostome “teeth” are not really teeth because they are made out of three keratinous cones one on top of each other, that emerge from the base of their oral epithelium (Trott and Lucow, 1964). Similarly, what is commonly referred to as scales in tetrapods are very different in terms of tissue composition (most often keratin) to the dermal denticles of chondrichthyans or to teleostean scales. In the following, I will be focusing on odontodes defined by Ørvig as mineralizing units that develop in the same manner as teeth (Ørvig, 1967), although I will consider teeth themselves as odontodes.

Odontodes are essentially characterized as a dentine crown surrounding a pulp cavity and covered by a hypermineralized tissue (Huysseune and Sire, 1998). In fossil fishes, odontodes are abundant and present slight differences in the histology of their tissues (reviewed in Sire

et al., 2009). However they can be found in isolated units such as in thelodonts or fixed on a bone layer as in heterostracomorphs, osteostracans and placoderms (Donoghue and Sansom, 2002). In chondrichthyans, odontodes (e.g. placoid scales) cover the totality of the body (except in holocephalans and some batoids) and are thought to be homologous to the aforementioned fossil odontodes, but also to the dermal elements among actinopterygians, such as ganoid scales (bichirs, lepisosteids) and the cosmoid to elasmoid scale of the coelacanth (sarcopterygian; Huysseune and Sire, 1998). All these structures can be considered odontodes due to a shared developmental mode of morphogenesis.

### 1.1.5 Evolution of skeletal mineralization

In the previous sections I have illustrated the diversity of skeletal structures among vertebrates. Fig 1.1 recapitulates the evolution of these structures along the vertebrate phylogeny, and highlights the dynamics of gains and losses along several branches. Although many more events have happened within each lineage, it is clear that anatomical characterization, histological analyses and, most importantly, comparative analyses between extant and fossil vertebrates are excellent tools to understand the evolution of the vertebrate skeleton. However, we can do better and further characterize the processes that underlie the structures observed in extant species and their evolutionary history. Understanding the processes through which, i.e. endochondral bone is formed, or how prismatic cartilage is synthesized are important steps to improve comparative analyses between vertebrates at the molecular, cellular and tissular levels. This will help us elucidate how these structures arose and not just when. In the next sections, I will be describing current knowledge on the developmental genetics of skeletal biology in extant vertebrates.

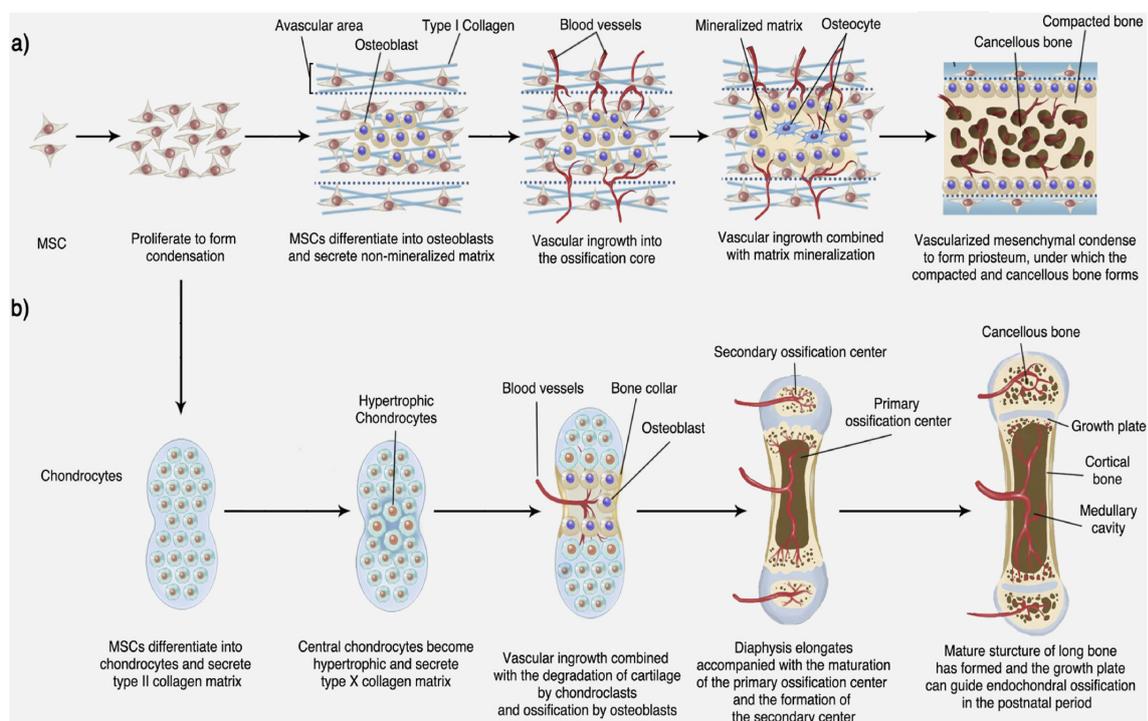
## 1.2 Skeletal development dynamics

It should first be noted that most of the current developmental studies concerning skeletal biology are biased towards model species belonging to osteichthyans, in particular the mouse model. Because of this bias, most of our understanding concerns ossification processes specific to bony fishes, and therefore conclusions are limited to this taxonomical level. However, more and more recent studies are starting to look at other vertebrates such as cyclostomes or chondrichthyans, increasing our understanding of the evolution of the skeleton. I will start by describing the skeletal development as a general overview of what is known in bony fishes and will later focus on its sister clades. In the following, descriptions will mostly concern tetrapods unless another organism is explicitly mentioned.

### 1.2.1 Endochondral and intramembranous ossification

Endochondral ossification is the process by which endochondral bones (or long bones) are built from cells of mesodermic origin. The first step is for these cells to make mesenchymal

condensations and differentiate into cartilage producing cells: the chondrocytes. They will then synthesize a cartilaginous anlagen to serve as a “blueprint” for the future bone (Mackie et al., 2008). Importantly, the chondrocytes will change morphology becoming hypertrophic while the cartilage is invaded by blood vessels, bone-synthesizing cells: osteoblasts, and bone-degrading cells: osteoclasts. This event will create several ossification centers in the long bone (first in the center then at the extremities) that will progressively degrade the cartilage matrix and replace it with bone (Maes et al., 2010). The interplay, between cartilage degradation and bone synthesis in this step is responsible for the typical spongy or porous structure of endochondral bone. In parallel, osteoblast progenitors in the perichondrium differentiate into osteoblasts and deposit cortical bone (perichondral bone) around the cartilage anlagen (Fig 1.6b; Berendsen and Olsen, 2015).



**Figure 1.6:** Types of ossification, modified from G. Zhu et al., 2021. (a) Intramembranous ossification. (b) Endochondral ossification. MSC: Mesenchymal Stem Cells. cancellous=spongy

Compared to long bone ossification, intramembranous (or dermal) ossification is less understood (Abzhanov et al., 2007; Ai-Aql et al., 2008). It is through this process that dermal bones (or flat bones) are constructed in bony fishes (and probably in armoured fishes). In general, dermal bones are considered to be derived from neural-crest cells, a bundle of cells that separate from the neural tube (ectoderm), migrate and differentiate into a variety of specialized cells (B. K. Hall, 2008). In intramembranous ossification, neural-crest cells first become mesenchymal cells (through an epithelial-mesenchymal transition; Simões-Costa and Bronner, 2015) that condense and then differentiate into osteoblasts (Berendsen and Olsen, 2015; Karsenty, 2003; Fig 1.6a). Although some differences exist in terms of developmental processes in intramem-

branous *versus* endochondral ossification (Abzhanov et al., 2007), the specialized cells serve the same purpose: mineralize an extracellular matrix.

In mature bone tissue, the most abundant cell-type are osteocytes that differentiate from osteoblasts, embedded in the mineral matrix as a result of their own activity (Bellido, 2014). Osteocytes are not isolated in the bone matrix since they communicate through a network of canaliculi, called the lacuno-canalicular system, which together with osteocytes are critical for bone homeostasis and remodelling (Burra et al., 2010).

### 1.2.2 Extracellular matrix composition

To understand the cellular and molecular processes that allow for the mineralization process, we need to understand the biological material that is going to mineralize: the extracellular matrix (ECM). An abundant ECM is the main characteristic of connective tissues such as cartilage or bone. ECMs are defined as non-cellular three-dimensional macro-molecular networks, consisting of a complex assembly of fibrous structural proteins (e.g. collagens and elastin), proteoglycans and other specialized proteins that can vary from one tissue to another. In the following, I will be describing some components of the extracellular matrix of bone and cartilage with a slight focus on osteoblast-secreted proteins and their function in mineralization (Clause and Barker, 2013).

Collagens are the most ubiquitous proteins in metazoans and the most common template for the formation of skeletal elements (Weiner and Addadi, 1997). They can form different macro-molecular structures such as fibrils (fibrillar collagens e.g. collagen I or II) or networks (e.g. collagen VI or X; reviewed in Gordon and Hahn, 2010), giving the ECM different mechanical and biological properties. For example, in cartilage, networks of collagen type VI form a pericellular matrix that immediately surrounds chondrocytes forming a unique structure: the chondron. This creates a micro-environment allowing the cells to be more resilient to external stimuli (e.g. mechanical) or regulate specific signals (e.g. apoptotic signal; reviewed in Z. Zhang, 2015). Complementing the pericellular matrix of chondrocytes, is the interstitial matrix (i.e. everything in-between chondrons), which in hyaline cartilage is mainly composed of water sequestered in a mesh made of type II collagen, hyaluronan and aggrecan (two proteoglycans). This interstitial matrix is highly resistant to tension due to the collagen fibrils, but also to compression due to the accumulation of negatively charged aggrecan on hyaluronan chains. The anionic groups on aggrecan carry with them mobile counter ions such as  $\text{Na}^+$ , which in turn draw in water swelling and expanding the cartilage (Hardingham, 2006). Furthermore, the main protein in bone, Collagen I, is more closely linked to the mineralization process. It has been shown that this particular fibril forming protein serves as a scaffold for direct mineral deposition in bone (Robinson and Watson, 1955; Weiner and Traub, 1992). Another example is Collagen X, a network forming collagen, which is crucial for endochondral ossification. Its expression in hypertrophic chondrocytes contributes to the initiation of calcium deposition

in the cartilaginous matrix (Kirsch and Wuthier, 1994). Expression of this gene has been observed in osteoblasts of amphibians (Aldea et al., 2013) and actinopterygians (Albertson et al., 2010; Eames et al., 2012; Laue et al., 2008), which suggests that it was originally involved in mineralization of both cartilage and bone.

Non-collagenous proteins (NCPs) are also abundant in ECMs. In particular, a subclass of proteoglycans called the Small Leucine-Rich Proteoglycans (SLRPs) possess multivalent binding abilities that can modulate the assembly of collagens in ECMs (reviewed in S. Chen and Birk, 2013). Although not all of them seem to be associated with mineralization (Ustriyana et al., 2021), some, such as Decorin (Dcn) have been proposed to play an inhibitory role of mineralization (Hoshi et al., 1999) by occupying the space between collagen fibers destined to be filled with mineral (Scott and Orford, 1981; Scott et al., 1981).

Two other proteins have been of particular interest in mineralization studies: the vitamin K-dependant or Gla proteins (matrix Gla protein: Mgp and bone Gla protein: Bgp or osteocalcin). Both of these proteins possess a “Gla” domain characterized by the ability to undergo  $\gamma$ -carboxylation of glutamate residues, resulting in a putative ability to bind calcium (Yáñez et al., 2012). Mgp has been shown to be an inhibitor of several processes including precipitation of calcium in cartilage and blood vessels (reviewed in Wen et al., 2018). On the other hand, large amounts of Bgp are deposited in the bone matrix (endochondral and dermal), and seems to incorporate calcium ions in the mineral crystals of the bone matrix (reviewed in Wen et al., 2018).

Another family of NCPs is involved in mineralization: Sparc (Secreted Protein Acidic and Rich in Cysteine, formerly called osteonectin), Sparc-L1 (Sparc-like1) and the Scpps (Secretory Calcium-binding Phosphoproteins). In short, *Sparc* and *Sparc-L1* arose from a whole genome duplication event, while the *Scpps* would be the result of tandem duplications from *Sparc-L* (Bertrand et al., 2013; Kawasaki et al., 2007). The Sparc protein is secreted by osteoblasts and functions by binding both collagen fibrils and calcium (Kessels et al., 2014; Schreiweis et al., 2007). However, it was later shown that Sparc modulates cell–matrix interactions without having structural roles in the matrix (Bornstein and Sage, 2002). Similarly, *Sparc-L1* has been shown to be spatially co-expressed with *Sparc* in tissues that undergo ossification (Weigele et al., 2015). Lastly, the Scpp subfamily includes a large number of members with most of them having a role in mineralizing tissues. For example, two Scpps osteopontin (Opn) and bone sialoprotein (Bsp) are thought to be respectively an inhibitor and a nucleator of mineral formation in bone (reviewed in Qin et al., 2004).

### 1.2.3 Extracellular matrix mineralization

The mineral in bone is composed of apatite crystals ( $\text{Ca}_{10}(\text{PO}_4)_6\text{X}_2$ , X = mostly F or OH; (Lowenstam, Weiner, et al., 1989; and see Eliaz and Metoki, 2017 for a comprehensive review), one of the most common forms of calcium phosphate minerals. However, the majority of min-

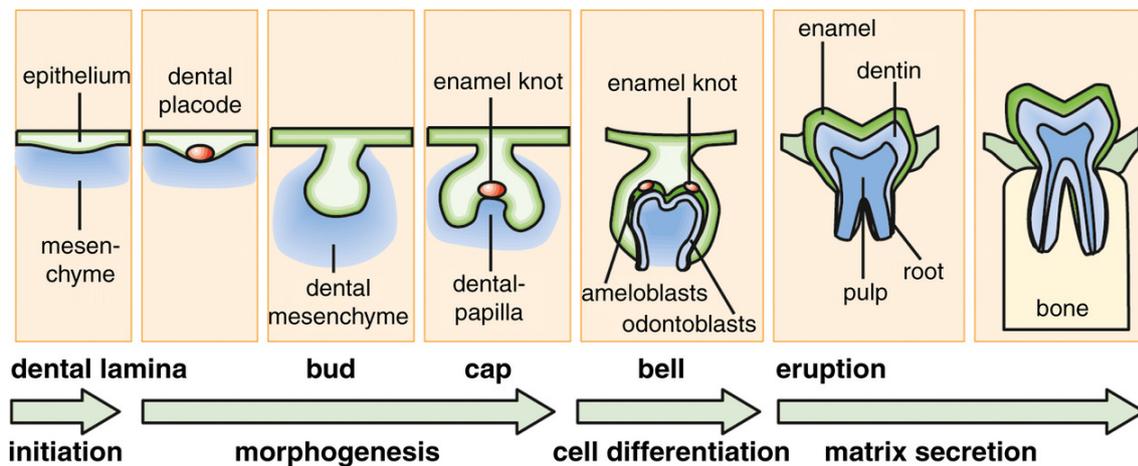
eral structures in metazoans is calcium carbonate. Therefore, why apatite is the main mineral component of bone is still an open question, but it would seem that the properties of apatite being a “sparingly soluble salt” would make it a safe reservoir for calcium and phosphorus (see Pasteris et al., 2008). Additionally, the spatial and chemical mechanisms that precipitate apatite on collagen fibrils are not fully understood (Grynopas and Omelon, 2007; Murshed, 2018; Weiner, 2006). In 2009, Omelon and collaborators proposed an interesting hypothesis on the control of skeletal mineralization by polyphosphates. These polyphosphates are the result of the condensation (or merging) of orthophosphates ( $\text{PO}_4^{3-}$ ) into linear ions (Wazer and Holst, 1950) and were identified in several active ossification sites. Thanks to polyphosphates, high concentrations of phosphate and calcium can be transported as calcium-polyphosphate complexes while remaining below the saturation level of apatite and therefore preventing its precipitation. When mineralization is initiated, an osteoblast-secreted enzyme called alkaline phosphatase (Alp) hydrolytically degrades the polyphosphates. This action releases orthophosphates and sequestered calcium, increasing their concentrations, raising apatite saturation levels and supplying the driving force for apatite formation (Omelon et al., 2009). This mechanism was later shown to happen in chondrichthyan tesserae, suggesting that this may be a general regulating feature in the mineralization of vertebrate skeletons (Omelon et al., 2014).

#### 1.2.4 Odontode development and matrix proteins

Contrary to other skeletal tissues, odontodes possess their very own developmental process and mineralization patterns. Odontode formation in mammals starts with a thickening of the epithelium, followed by a condensation of the underlying mesenchyme, made of neural crest cells. Subsequent morphogenesis is driven by cell proliferation leading to the formation of an enamel knot which regulates the growth and folding of the epithelium. Finally, cell differentiation directs ameloblast formation in the epithelial compartment and odontoblast formation in the mesenchymal compartment. These cells will respectively secrete enamel and dentin matrices, the odontode specific tissues (Fig 1.7; reviewed in Catón and Tucker, 2009). In mammals, dentin and enamel are very special tissues because the matrix proteins they display are almost always part of the Scpp family. In dentin, the bulk of the odontode tissues, the secreted matrix proteins are collagen type I but also dentin sialoprotein (Dsp) and dentin phosphoprotein (Dpp), two Scpp proteins. However, in enamel the very hard outer layer of the odontode, the ECM is composed of 2 other Scpps, enamelin (Enam) and amelogenin (Amel) but not collagen (B. Hall, 2014b).

#### 1.2.5 Outside bony fishes

The questions that immediately comes to mind from an evolutionary perspective is how generalized are these processes and structures, how can we compare them in vertebrate organisms and how can we identify the ancestral features that evolved at the origin of vertebrates? Cy-



**Figure 1.7:** The process of tooth development. The dental placode and enamel knots are signalling centres regulating tooth morphogenesis; Thesleff, 2014

clostomes do not present mineralized skeletons, but another important aspect of their skeletal biology must be taken into account. Contrary to all extant gnathostomes that present a type of cartilage that is predominantly composed of collagen type II, lampreys and hagfishes possess a non-collagenous type of cartilage. In lampreys, the ECM lacks recognizable collagen fibrils and form a dense network of randomly arranged fibrils. Their cartilages also react to Verhoeff's stain, an elastin specific stain (Wright et al., 2001). These elements along with some others indicate that lamprey cartilages are based on an elastin-like protein that was named Lamprin. Similar observations were made in the hagfish, with the main cartilage protein being named Myxinin. Additionally, in hagfishes a special type of cartilage (denoted type II) is unique in that it is apparently made up of collagen I (reviewed in Wright et al., 2001). Hagfish and lamprey cartilages probably represent derived features among vertebrates. Although still little is known about cyclostome skeletal biology, another lineage can be useful to understand the evolution of the skeleton and of mineralization: the cartilaginous fishes. This lineage has lost the ability to make bone, but still kept the capacity to mineralize some tissues. Chondrichthyans are therefore an important taxa to study for the understanding of the evolution of mineralization. In the following sections, I will be focusing on the latest work on chondrichthyan development, skeletal biology and mineralization.

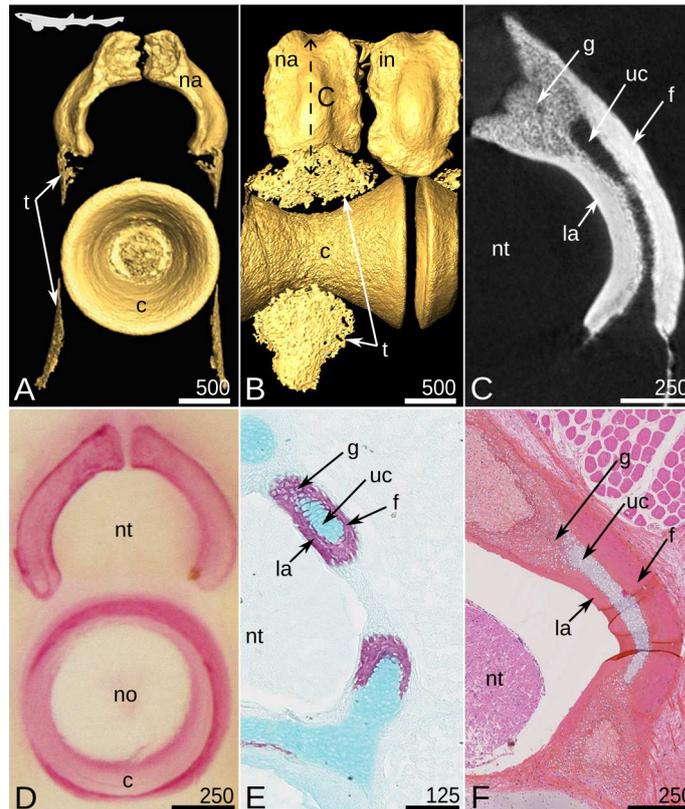
## 1.3 Chondrichthyans unchained

### 1.3.1 Development of the tessellated skeleton

As mentioned above, the majority of the skeleton of elasmobranch fishes (sharks, rays and relatives) is tessellated. The onset of calcium phosphate mineralization might be similar to bony fishes as suggested by localized alkaline phosphatase activity that precedes tesserae formation in the swell shark, *Cephaloscyllium ventriosum* (Eames et al., 2007). Additionally, similar to

bony fishes, chondrocytes seem to play an important role in mineralization since most of the profound changes in chondrocyte morphology coincide with the start of tesseral formation in the round sting ray *Urobatis halleri*. However, in contrast to bony fishes, cartilage mineralization is associated with flattened peripheral chondrocytes rather than deep, hypertrophied cells (Dean et al., 2009). Once mineralization has begun, tesserae grow both in depth and width and by accretion of mineral on all surfaces, such that they increase in size and mineral content throughout ontogeny (Dean et al., 2009; Doyle, 1968). As a result, chondrocytes find themselves embedded alive in the tesserae, engulfed into gaps (lacunae) in the mineralized tissue, making the tesserae a cellular tissue (Dean et al., 2009; Seidel et al., 2020). Chondrocytes within tesserae maintain an unmineralized pericellular matrix (Dean et al., 2009; Kemp and Westrin, 1979; Moss, 1977; Seidel, Blumer, Pechriggl, et al., 2017; Seidel et al., 2016), suggesting local inhibition of mineralization by chondrocytes (Seidel et al., 2016). These intra-tesseral cells however are not completely isolated, since it has been shown that the lacunae where they reside are connected to each other through small canaliculi forming a continuous lacuno-canalicular network (Chaumel et al., 2020; Dean et al., 2010), similar to what has been observed in cellular bone tissue. As tesserae start to grow, they can end up touching each other. In some places, tesserae are in direct contact, while in others they are separated by gaps filled with unmineralized ECM including cells and dense fiber bundles. These zones are called inter-tesseral joints and are composed of at least three types of collagens (Type I, II and X) in *Urobatis halleri* (Seidel, Blumer, Pechriggl, et al., 2017). Therefore, tessellated cartilage growth follows early organization of isolated, surface mineralization centers that grow appositionally to maintain contact as the underlying uncalcified matrix expands in volume (Dean et al., 2009). Tesserae morphogenesis groups together two main types of mineralization: (i) globular mineralization at the subperichondral level in many chondrichthyan skeletal units, involving standard chondrocytes and (ii) prismatic mineralization occurring at the junction between cartilage units and perichondrium probably involving cells of both tissues (chondrocytes and fibroblasts; Debiais-Thibaud, 2019).

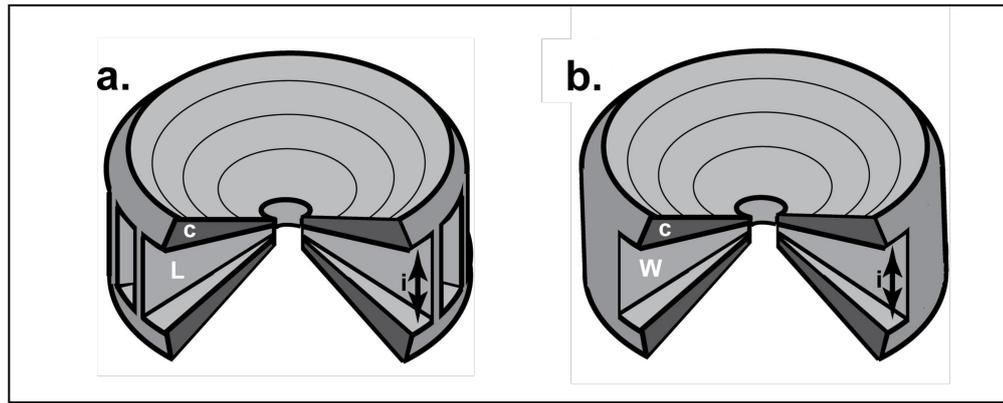
Tesserae however are not the only mineralized structures in chondrichthyan skeletons. In a recent article by Berio et al., 2021 a thorough description of the diversity of mineralized tissues in a series of chondrichthyans species is presented. These descriptions are restricted to the neural arches of the vertebrae and therefore do not include the diversity of areolar mineralization specific to chondrichthyan chordacentra. One of the scales of organization described focuses on cells and ECMs involved in mineralization (microscale; Dean and Summers, 2006) and allows for the distinction of three histotypes, while taking into account previous descriptions : globular, fibrous and lamellar mineralizations (respectively “g”, “f” and “la” in Fig 1.8). Prismatic mineralization (typical of the cap zone of the tesserae) was not included in these histotypes since this mineralization type is identified through polarized light microscopy, incompatible with histological techniques. However, interesting data is presented in link with the “bone-like” tissues previously reported (Atake et al., 2019; Eames et al., 2007; Peignoux-Deville et al., 1982). The authors suggest that this tissue might actually be two different types



**Figure 1.8:** Mineralized tissues in the anterior vertebrae of the small-spotted catshark *Scyliorhinus canicula* at juvenile stages [20 cm TL in (A–C) and 31 cm TL in (F)] and embryonic stages [8.4 cm TL in (D) and 8.0 cm TL in (E)]. (A–C) Micro-computed tomography imaging, frontal view (A) and lateral view (B) of 3D isosurfaces and virtual section (C, as located by the double sided dashed arrow on B). (D) Alizarin Red S staining of a thick section. (E) Alizarin Red S and Alcian Blue double staining on a cryostat section. (F) HES, histological staining; c, centrum; f, fibrous perichondrial mineralization; g, globular mineralized cartilage; in, interneural element; la, lamellar perichondrial mineralization; na, neural arch; no, notochord; nt, neural tube; t, tesserae; uc, unmineralized cartilage. Scale bars are in µm. Berio et al., 2021

of mineralization (fibrous and lamellar) with different growth processes and cell behaviours. Fibrous mineralization is defined by the presence of fibrocyte-shaped cells enclosed in a fibrous matrix, and occurs in the distal part of the neural arches (muscle side). It is shown to be of perichondrial nature and therefore must grow by external apposition (as also suggested by Atake et al., 2019). In contrast, lamellar mineralization displays a more linearly organized ECM but with an extremely low cell density, compared to fibrous mineralization. It occurs in the proximal part of the neural arches (neural tube side) and the growth is suggested to be appositional from cells (here: fibro-chondrocytes) facing the neural tube, where no thick unmineralized perichondrium is detected (Fig 1.8).

Vertebral centra studies in chondrichthyans are scarce. Some studies focus on the timing of mineralization showing that, typically, centra are among the first element to mineralize in chondrichthyans (Eames et al., 2007 but see Enault et al., 2016). In Atake and Eames, 2021 a review on morphological and molecular features of mineralized tissues in chondrichthyans, a paragraph focuses on centra evolution but with no comments on areolar mineralization. Ride-



**Figure 1.9:** Schematics of lamniform and carcharhiniform abdominal vertebrae. (a) 3D representation of a lamniform vertebral centra with some material shown transparent so that the cross-section of the structure can be seen. c - hourglass-shaped corpus calcarea, i – intermedialia, L – lamella. (b) 3D representation of a carcharhiniform vertebra with material cut away to reveal the cross-sectional structure. Symbols as in panel a, W – wedge. Modified from Morse et al., 2022

wood and MacBride, 1921 is the only study where a thorough description of chondrichthyan centra can be found. Centra are composed of three main components: an hourglass-shaped (or double cone) wall, an internal tissue, sandwiched between the hourglass walls, called “intermedialia” and several mineralized structures embedded in the intermedialia, either lamellae (e.g. in lamniforms, Fig 1.9a), or more bulky wedges (e.g. in carchariniiformes, Fig 1.9b; Morse et al., 2022; Ridewood and MacBride, 1921). The term areolar mineralization refers to the mesostructure of the mineralized matrix which forms calcified layers arranged in a concentric manner in the vertebral bodies (Ridewood and MacBride, 1921). However, when looking more in detail, it would seem that two different tissues are involved. The innermost zone of mineralization in the fibrous sheath of the notochord has been described as fibrous mineralization (or fibro-cartilage) while in the periphery of this mineralized layer, the initially hyaline cartilage mineralizes as globular mineralization (Debiais-Thibaud, 2019; Ridewood and MacBride, 1921). More recently Morse et al., 2022 scanned some centra in 3D for six shark species in an effort to characterise these structures from a biomineralization standpoint, i.e. describing the structures following Dean and Summers, 2006 scales of organization. However, the authors discuss their results in terms of mechanical constraints and differences in mineral density but no specifics on areolar mineralization are mentioned. Nonetheless, it is worth mentioning the observation of cartilage canals running radially in the intermedialia of all species with somewhat higher levels of mineralization around them (Morse et al., 2022).

Concerning odontodes it has been suggested that the modes of development from a gene regulation standpoint are ancestral to all vertebrates who display these structures (albeit with slight variations Debiais-Thibaud et al., 2015; Debiais-Thibaud et al., 2011). This seems to be true for morphogenesis (except the enamel knots) but is it also true for later development (discussed below). It is generally accepted that chondrichthyans teeth and scales are mostly made up of typical dentine while the outermost capping hypermineralized tissue is not enamel but

a similar structure called enameloid. Mature enameloid resembles enamel both topologically and functionally, but the crystalline arrangement is less ordered and not prismatic because of the presence of collagen fibrils prior to maturation. Additionally, its matrix includes collagen fibers considered to be produced by odontoblasts (Sire et al., 2009). However, little is known about the actual genes related to the mineralization of these structures in chondrichthyans.

### 1.3.2 Developmental genetics in chondrichthyans

In the previous section I reviewed current knowledge on structure and development of skeletal elements in chondrichthyans, with a strong focus on the endoskeleton. Gene expression in these structures can give us essential information on the regulation of mineralization and therefore insights into the vertebrate ancestral mineralization toolkit. Chondrichthyans are indeed extremely interesting from this point of view because: (i) they lost the ability to make bone, (ii) they have developed a derived form of mineralization, (iii) they display at least four types of mineralization, some of them resembling bone structures and (iv) they possess both oral and dermal odontodes that are also understudied in terms of mineralization. Additionally, chondrichthyans occupy an ideal phylogenetic position which combined with the aforementioned features, makes them excellent models for mineralization studies in vertebrates through comparative analyses. In the following I will be discussing the latest work on developmental genetics in link with mineralization in chondrichthyans.

First, although not in direct link with mineralization, Aleksandra et al., 2020 characterized how the cartilage of a skate *Leucoraja erinacea* develops, grows and heals. Importantly they show that the chondrocytes in this species express the markers of classic bony fish cartilage, i.e. collagen II and aggrecan. Showing that this tissue has kept the general features of a “classic” cartilage. Similarly, Eames et al., 2007 showed that in the swell shark, as described in other vertebrates, alkaline phosphatase activity predicts mineralization patterns and that Collagen type I shows a more perichondral-associated distribution which is mutually exclusive with collagen type II (cartilage associated; also see Enault et al., 2015).

In another study, this time in direct link with mineralization in cartilaginous fishes, Debiais-Thibaud et al., 2019 looked at a specific gene involved in endochondral ossification: Collagen type X. The authors explain that the Collagen X gene is a great candidate gene for mineralization because (i) it contributes to the initiation of calcium deposition in the cartilaginous matrix (Kirsch and Wuthier, 1994), (ii) it has been seen expressed in bony fish osteoblasts (Albertson et al., 2010; Eames et al., 2012; Laue et al., 2008) and (iii) it emerged through the two rounds of whole genome duplication before vertebrate diversification (Aldea et al., 2013) and should therefore be present in cartilaginous fishes. The authors offer the first description of *Col10a1* in cartilaginous fishes, with an unexpected expansion of this gene family, accounting for a total of six *Col10a1* chondrichthyan-specific genes. All the six paralogs showed a mineralization-associated expression pattern in the small-spotted catshark *Scyliorhinus canicula*. *Col10a1.1*

and *Col10a1.2* were transcribed in ameloblasts (secretory and maturation stages), whereas *Col10a1.4*, *Col10a1.5*, and *Col10a1.6* were transcribed in the mesenchymal cells of odontodes (future odontoblasts). The *Col10a1.3* duplicate showed expression in both the epithelium (ameloblasts) and mesenchyme (odontoblasts) and *Col10a1.4* expression was detected in cells going through globular, fibrous and lamellar mineralization types. These results show remarkable differences between bony and cartilaginous fishes both in gene repertoires but also in the way they are used. Importantly, this suggests a previously unknown involvement of a non-fibrillar collagen in enameloid formation. The authors propose, based on their results, that *Col10a1* was part of an ancestral biomineralization toolkit used in all mineralizing cells of early vertebrates. Then, the duplicates in chondrichthyans underwent gene subfunctionalization in different mineralizing tissues, following a duplication-degeneration-complementation model (DDC; Force et al., 1999) which explains that duplicates are intrinsically redundant and therefore will either be purged from the genome (degeneration), or must divide the ancestral function between paralogs (complementation). Alternatively, if the original function is an essential one, the new duplicate can acquire a new function altogether insuring the preservation of a non mutated copy.

Additional studies highlight the importance of the *Sparc/Sparc-L1/Scpp* gene family in cartilaginous fishes. As mentioned earlier, the classical scenario for the evolution of this family proposes that *Sparc* and *Sparc-L1* arose through two rounds of vertebrate-specific genome duplications (Bertrand et al., 2013; Kawasaki et al., 2007), while the *Scpp* gene family is thought to have originated from a series of tandem duplications of the *Sparc-L1* gene (Kawasaki et al., 2017; Ryll et al., 2014; Venkatesh, Lee, Ravi, Maurya, Korzh, et al., 2014). However, the evolutionary history of this family is more complicated. The gar and the coelacanth genomes to contain a *Sparc-L2*, which was shown to be an ancestral duplication common to osteichthyan and to be subsequently lost in tetrapods and teleosts (Enault et al., 2018). Additionally, *Scpps* are thought to be duplicates of *Sparc-L1* but the high rate of evolution of these genes do not allow for a correct phylogenetic reconstruction. The paralogy relationships with *Sparc-L1* are therefore only based on intron-exon structure, and the *Scpp* orthology relationships between osteichthyans lineages are still unclear (Kawasaki and Amemiya, 2014; Kawasaki et al., 2017). Moreover *Scpps* are thought to be specific to osteichthyans because none have been detected in the elephant shark genome, which was one of the only available chondrichthyans genome at the time (Venkatesh, Lee, Ravi, Maurya, Lian, et al., 2014), and might not be representative of the whole clade. The study that identified the loss of *Sparc-L2* in osteichthyan lineage (Enault et al., 2018), used *de novo* assembled chondrichthyan transcriptomes to identify *Sparc* and *Sparc-L* and show that they are well conserved in this clade. The authors focus on teeth and scale comparisons between two elasmobranch species as well as the western clawed frog. Their results show a conserved module made up of three genes co-expressed in the odontoblasts of the studied species: *Sparc*, *Col1a1* and *Col1a2* (both needed for collagen I fibril formation), which coupled with similar results in teleosts and mammals, confirm the hypothesis of dentin being homologous across jawed vertebrates. Moreover, the authors found a specific expression of

*Sparc-L* in chondrichthyan ameloblasts, which is not observed in bony fishes, but is very similar to their expression patterns of some *Scpps*. Therefore, they suggest that these differences in expression could result from the inheritance of an ameloblast-specific expression from the ancestral *Sparc-L* gene (before the chondrichthyan / osteichthyan split). What this hypothesis implies, is that the characteristic matrix proteins of enamel (i.e. *Scpps* such as enamelin and amelogenin) might reflect osteichthyan-specific innovations, deriving from a common odontode mineralization toolkit ancestral to vertebrates and making all outer odontode layers homologous to each other (ganoine, enameloid, enamel, etc...).

These studies highlight how different mineralization-associated genes underwent parallel evolutionary events in chondrichthyans and osteichthyans but are derived from an ancestral biomineralization toolkit that may have evolved independently in different cell types and organism lineages. However, the number of studies looking into these subjects are scarce, and our understanding of chondrichthyans developmental genetics in link with mineralization is still very poor. An important element that these studies (Collagen X and *Sparc/Sparc-L*) have in common is the use of novel transcriptomes for representatives of different chondrichthyans lineages. This data is essential for the identification and study of the candidate genes, which will only be further improved with well-sequenced genomes and better assembled transcriptomes. Additionally, these studies do fall in the domain of genetics although it is important to note that they do not concern functional genetics. Chondrichthyan mutants are logistically impossible to generate for several reasons so the real function of the studied genes and proteins cannot be evaluated in these models. However, with fresh tissues and *in situ* hybridizations, the cells expressing genes of interest can be identified, allowing for robust functional hypotheses.

### 1.3.3 How to identify mineralization genes in chondrichthyans

The previously cited articles have in common their general approach which is the candidate gene approach. This approach is quite intuitive, candidates genes would be chosen based on known functions in bony fishes and then studied in cartilaginous fishes. A candidate gene approach can be very proficient at detecting commonalities or differences between lineages, but it has a major flaw. By approaching chondrichthyan mineralization through a candidate gene approach, we are basing our studies on possibly derived features proper to osteichthyans. Therefore, we could be missing important chondrichthyan-specific innovations that have no direct link to the mode of ossification in other lineages or ancestral features that were lost secondarily in bony fishes. There is, however, another approach that can be used to avoid this issue. A more general approach using “omic” data, such as genomic, transcriptomic or proteomic data-sets can provide intrinsic chondrichthyan elements linked to mineralization, which could later be compared to existing literature on ossification. For example, transcriptomic analyses of only mineralized tissues cartilaginous fishes can help find out the most highly expressed genes in these tissues, or proteomic analyses of the mineralized and unmineralized matrices of these

tissues could point toward the most abundant proteins present in them. These very appealing techniques however, have not been used until now because of a very poor repertoire of well-assembled genomes and transcriptomes in chondrichthyans. Nonetheless, in recent years more and more chromosome-level assemblies of these genomes have been published granting us access to these techniques. This way we can potentially continue elucidating the chondrichthyan biomineralization toolkit, comparing it to the bony fish toolkit, and inferring the ancestral state of jawed vertebrates mineralization genes.

Several chondrichthyan species have now fully assembled genomes, but only one is a species that has been long used by researchers in laboratories : the small-spotted catshark (*Scyliorhinus canicula*). This catshark of the family Scyliorhinidae is the species on which I worked during my PhD. It is one of the most abundant elasmobranch species of the northeast Atlantic and the Mediterranean, with stable populations in most areas (Ballard et al., 1993). Its reproductive biology is well known: it is an oviparous species with embryos that take about 6 months to develop depending on sea temperature, the female possesses a spermatheca which allows for continuous fertilization throughout the year after the mating season and, at least in the Mediterranean sea, there is a seasonal pattern to the laying of the eggs from December to June (Ellis and Shackley, 1997). Additionally, a series of developmental stages have been thoroughly described for this species by Ballard et al., 1993. This means that we can acquire females right after the mating season, keep them in the lab during the laying period, pick up the eggs and then release the females while we wait for the embryos to reach the appropriate stages. Therefore, the knowledge of its development, life cycle and reproductive biology, its easy maintenance in aquariums and its abundance in European coasts makes *Scyliorhinus canicula* a perfect non-conventional model to study the development of chondrichthyans, known to European marine biology stations since the XIXth century (Chevrel, 1889).

## 1.4 Outline of the thesis

In this thesis I focus on the developmental genetics of the skeletal mineralization in a cartilaginous fish: the small-spotted catshark *Scyliorhinus canicula*. I describe in detail putative genetic factors underlying the mineralization process, and I do this through two approaches, by candidate gene study and transcriptomics. This allows me to better characterize the mineralization toolkit in a chondrichthyan species, so that through comparative analysis an ancestral state for jawed vertebrates can be inferred, at least partially. First, with the candidate gene approach, I test the level of conservation (or non-conservation) concerning several known mineralization-associated genes. In addition, this is the default approach when omic data is not available, which was my case when I started this project. Next, I describe how I used recently generated genomic and transcriptomic data from the small-spotted catshark to study the intrinsic genes expressed specifically in mineralizing tissues. Both of these approaches allow me to detect interesting genes which are then studied through phylogenetic analyses, quantification and lo-

calization of expression as well as comparative analyses with bony fishes (whenever possible). In the last part, I present two articles in preparation that result from the work we realized with the help of two master students. These articles come from a candidate gene approach but concern ample gene families and in an effort to include all paralogs, the articles shift towards a molecular evolution framework rather than an Evo-Devo one.



# Chapter 2

## Bony fish mineralization genes in the catshark

### 2.1 Introduction

In this first chapter I will be presenting my work on several candidate genes, studied in the small-spotted catshark. This candidate gene approach was the first and main one used during my thesis, since very little omic data was available in 2019. I have already mentioned the pros and contras of this approach, but I will now focus on the general procedure common to all the following articles. First when considering a candidate gene, known to have a specific function in mineralization in bony fishes, the first step is to verify whether this gene exists in chondrichthyans and more particularly in my model species. This is where the little omic data I had access to, comes into play: one bulk jaw transcriptome and one bulk vertebra transcriptome of *Scyliorhinus canicula* that were available in the lab and were used as an assembled reference transcriptome. These transcriptomes however, might not contain the transcript of a candidate gene, simply because of the tissular specificity or because of assembly errors. However, as the jaw and the vertebrae both contain tissues that mineralize I would expect to find my candidate genes most of the time. Further screening of other available chondrichthyan transcriptomes can further validate the presence/absence of the candidate gene in chondrichthyans as a whole. By this time, the analysis is already interesting as I can (i) not find my candidate gene which could relate to bone loss in chondrichthyans, (ii) find the gene and potentially have a conserved feature of the vertebrate mineralization toolkit or (iii) even more thrilling, find an expansion of the candidate genes in chondrichthyans (i.e. a 1:2 or 1:3, etc... orthology relationship between bony and cartilaginous fishes). Other specific cases are possible such as pseudogenization, but are harder to identify. Every one of these cases calls for a phylogenetic analysis since it represents a robust way of making sure the candidate gene identified in chondrichthyans is indeed the true ortholog of the known (or targeted) bony fish gene. This is also important since many genes are members of a bigger functional class where conserved domains can be source of miss-

identification. Such phylogenetic reconstruction, can be coupled with synteny analyses, when good genomic data is available, to further retrace the evolutionary history of the genes.

Identifying candidate genes, and reconstructing their evolutionary history is not enough since this gives us no evidence of functional conservation. In fact, functional studies in non-model species are very challenging if not impossible. In particular the techniques used for these experiments must be perfectly adapted to the species life cycle which can present a few obstacles. For instance, *Scyliorhinus canicula* is an oviparous species and when the eggs are laid the embryos are already at a far too advanced stage for mutagenesis/knockout experiments to be carried out. How to get around these biological limitations to figure out the functional similarities between bony/cartilaginous fishes orthologs? What I did is try and find lines of evidence that point toward the putative function of the studied gene (and protein) and compare them between species. One way to do this is to characterize the intron-exon structure of the protein as well as functional domains, such as catalytic, phosphorylation, acetylation or calcium binding domains (among many others). Possible non-functional mutations can be identified in these domains, i.e. deviations from the consensus motifs defined for a given domain, but also good conservation can be identified and therefore it may be extrapolated as conservation of the putative function. Another way of doing this is to more precisely quantify and localize the expression of the candidate genes through RT-qPCR and *in situ* hybridizations. Studying of the ontogenic trajectories of gene expression during embryonic development can also hint toward a specific putative function. For example, a rise of gene expression at a given developmental stage can be linked to the onset of mineralization if that time frame is known. For localization, mRNA *in situ* hybridization can localize the expression of the candidate genes to specific cells in the tissues and show whether these are linked to mineralization processes.

The identification of candidate genes in chondrichthyans, reconstruction of their evolutionary history at the gnathostome/vertebrate scale, protein domain conservation and both quantification and localization of expression in tissues and through development are the main tools I have used to elucidate part of the chondrichthyan mineralization toolkit and the ancestral state of this toolkit in gnathostomes. In the following, I will describe and discuss three articles related to several mineralization-associated genes previously mentioned. First I will be focusing on Mgp and Bgp two calcium binding vitamin K-dependent proteins, identified as positive (Bgp) and negative (Bgp and Mgp) regulators of biomineralization in bony fishes. Then, I will be presenting an article on *sparc* and *sparc-L* two paralogs known respectively for being extremely abundant in bone and for being the origin of all the *scpp* genes in bony fishes. Lastly, I will be focusing on said *scpps*, since they are expressed in many mineralized structures in bony fishes and were thought to be absent from chondrichthyan genomes, until now. I collaborated with Tatjana Haitina and Sylvain Marcellini which were of great help for comparative analyses of the catshark with the zebrafish and the tropical clawed frog and respectively.



# Evolution of Matrix Gla and Bone Gla Protein Genes in Jawed Vertebrates

Nicolas Leurs<sup>1</sup>, Camille Martinand-Mari<sup>1</sup>, Stéphanie Ventéo<sup>2</sup>, Tatjana Haitina<sup>3\*</sup> and Mélanie Debiais-Thibaud<sup>1\*</sup>

<sup>1</sup> ISEM, CNRS, IRD, EPHE, Univ. Montpellier, Montpellier, France, <sup>2</sup> Institute for Neurosciences of Montpellier, Saint Eloi Hospital, Inserm UMR 1051, Univ. Montpellier, Montpellier, France, <sup>3</sup> Department of Organismal Biology, Uppsala University, Uppsala, Sweden

## OPEN ACCESS

### Edited by:

Susana Seixas,  
University of Porto, Portugal

### Reviewed by:

Shigehiro Kuraku,  
RIKEN Center for Biosystems  
Dynamics Research (BDR), Japan  
Vincent Laizé,  
University of Algarve, Portugal

### \*Correspondence:

Mélanie Debiais-Thibaud  
melanie.debiais-  
thibaud@umontpellier.fr  
Tatjana Haitina  
tatjana.haitina@ebc.uu.se

### Specialty section:

This article was submitted to  
Evolutionary and Population Genetics,  
a section of the journal  
Frontiers in Genetics

**Received:** 23 October 2020

**Accepted:** 08 February 2021

**Published:** 10 March 2021

### Citation:

Leurs N, Martinand-Mari C,  
Ventéo S, Haitina T and  
Debiais-Thibaud M (2021) Evolution  
of Matrix Gla and Bone Gla Protein  
Genes in Jawed Vertebrates.  
*Front. Genet.* 12:620659.  
doi: 10.3389/fgene.2021.620659

Matrix Gla protein (Mgp) and bone Gla protein (Bgp) are vitamin-K dependent proteins that bind calcium in their  $\gamma$ -carboxylated versions in mammals. They are recognized as positive (Bgp) or negative (Mgp and Bgp) regulators of biomineralization in a number of tissues, including skeletal tissues of bony vertebrates. The Mgp/Bgp gene family is poorly known in cartilaginous fishes, which precludes the understanding of the evolution of the biomineralization toolkit at the emergence of jawed vertebrates. Here we took advantage of recently released genomic and transcriptomic data in cartilaginous fishes and described the genomic loci and gene expression patterns of the Mgp/Bgp gene family. We identified three genes, *Mgp1*, *Mgp2*, and *Bgp*, in cartilaginous fishes instead of the single previously reported *Mgp* gene. We describe their genomic loci, resulting in a dynamic evolutionary scenario for this gene family including several events of local (tandem) duplications, but also of translocation events, along jawed vertebrate evolution. We describe the expression patterns of *Mgp1*, *Mgp2*, and *Bgp* in embryonic stages covering organogenesis in the small-spotted catshark *Scyliorhinus canicula* and present a comparative analysis with Mgp/Bgp family members previously described in bony vertebrates, highlighting ancestral features such as early embryonic, soft tissues, and neuronal expressions, but also derived features of cartilaginous fishes such as expression in fin supporting fibers. Our results support an ancestral function of Mgp in skeletal mineralization and a later derived function of Bgp in skeletal development that may be related to the divergence of bony vertebrates.

**Keywords:** Gla protein, osteocalcin, shark, skeleton, evo-devo, biomineralization, bglap

## INTRODUCTION

Vertebrates display a range of skeletal tissues that are biomineralized through the regulation of calcium phosphate crystal deposition (Janvier, 1996; Donoghue and Sansom, 2002; Omelon et al., 2009), except in the extant cyclostome group (agnathan fishes: lampreys and hagfishes) where the skeletal units are made of cartilage with no detection of calcium precipitates (Yao et al., 2011; Ota et al., 2013). Several vitamin K-dependent (VKD) proteins were shown to be involved in skeletal

tissue mineralization in jawed vertebrates (reviewed in Bordoloi et al., 2018; Wen et al., 2018). Of these, *Mgp* (matrix Gla protein) and *Bgp* (bone Gla protein, *bglap*, and osteocalcin) display consistent similarities in their sequences and were considered to belong to the same gene family (Laizé et al., 2005; Cancela et al., 2014). Both these proteins display a Gla domain characterized by the ability to undergo  $\gamma$ -carboxylation of several glutamate residues, resulting in a putative ability of the protein to bind calcium (reviewed in Yáñez et al., 2012).

Expression of the *Mgp* and *Bgp* genes in mouse first appeared spatially exclusive, with *Bgp* expressed uniquely in osteoblasts or osteocytes but also in odontoblasts, while *Mgp* expression was restricted to hypertrophic chondrocytes (Ikeda et al., 1992; D'Errico et al., 1997). More recent data support the expression of *Mgp* in other skeletal cells, including osteoblasts and osteoclasts (Coen et al., 2009). *Mgp* was also shown to be largely expressed in many soft tissues such as kidney, lung, heart, and spleen (Fraser and Price, 1988). *Mgp* was shown to act as an inhibitor of several processes both in skeletal and soft tissues: calcium precipitation in hyaline cartilage and human vascular smooth muscle cells (Luo et al., 1997; Schurgers et al., 2007; reviewed by Wen et al., 2018), and also dentin or bone matrices mineralization and osteoclast differentiation (Kaipatur et al., 2008; Zhang et al., 2019). The *Bgp* protein, on the other hand, seems to function in two ways, either in its carboxylated form by regulating hydroxyapatite crystal growth in skeletal tissues or in its non-carboxylated form by potentially acting as a circulating hormone that may be involved in energy metabolism and other functions (Diegel et al., 2020). Focusing on their skeletal functions, the data gathered from mammals indicate that *Bgp* is involved in the regulation, both positive and negative, of biomineralization processes in bone tissues, while *Mgp* is an inhibitory protein for these biomineralization processes (reviewed in Wen et al., 2018).

The evolutionary history of the *Mgp/Bgp* gene family has been discussed for more than two decades, particularly in relation to the evolution of a mineralized skeleton in vertebrates (Rice et al., 1994; Cancela et al., 2001, 2014; Pinto et al., 2001; Simes et al., 2003; Laizé et al., 2005; Gavaia et al., 2006; Viegas et al., 2013). The search for *Mgp* and *Bgp* genes in a variety of bony vertebrates led to the identification of two *Bgp* copies in several teleost fishes [the most recently identified being named *OC2* (Laizé et al., 2005; Cancela et al., 2014; Cavaco et al., 2014)] and also in some tetrapods [amphibians and sauropsids, where the recently identified duplicate was named *OC3* (Cancela et al., 2014)] while *Mgp* was, until now, only found to be present as a single gene (Cancela et al., 2014). The hypothesis was raised that *Mgp* and *Bgp* genes originated from an ancestral gene after the two whole-genome duplications in vertebrates (Laizé et al., 2005) and that more recent events of duplication of *Bgp* occurred more recently and independently in the bony fish and the tetrapod lineages (Cancela et al., 2014). Cartilaginous fishes, e.g., sharks (selachians), skates and rays (batoids), and holocephalans, are crucial in this issue as their lineage diverged from bony fishes more than 450 million years ago and they display a skeleton devoid of bone tissue but made of hyaline and mineralized cartilage (Janvier, 1996).

Several authors have previously described the presence of a *Mgp* gene in two shark species, the school shark *Galeorhinus galeus* (Rice et al., 1994) and the blue shark *Prionace glauca* (Ortiz-Delgado et al., 2006) for which they showed high conservation with tetrapod for (i) the *Mgp* amino-acid motifs which are critical for post-translational modifications [serine phosphorylation and glutamate  $\gamma$ -carboxylation (Price et al., 1994; Ortiz-Delgado et al., 2006)]; (ii) *Mgp* expression pattern and *Mgp* sites of accumulation [vertebral cartilage, endothelium, kidney, heart (vascular endothelia and smooth muscle), and dentinal matrix (Ortiz-Delgado et al., 2006)]. Previous studies have not identified any sequence that would be homologous to *Bgp* in cartilaginous fish genomes (Cancela et al., 2014).

The current explosion of genomic data, including in the cartilaginous fish lineage, allows the better description of gene complement and gene expression in this *Mgp/Bgp* family. Here we collect transcriptomic and genomic data from different jawed vertebrates, including several cartilaginous fishes where we identify an unknown diversity of *Mgp/Bgp* sequences and their genomic loci. We describe their gene expression patterns in embryonic stages of the small-spotted catshark *Scyliorhinus canicula* and uncover highly conserved but also previously unknown sites of expression.

## MATERIALS AND METHODS

### Collection of *Mgp/Bgp* Sequences in the Genomes and Transcriptomes of Chondrichthyans Synteny Analyses

Matrix Gla protein and bone Gla protein sequences for human, mouse, and zebrafish were collected from GenBank and were used to screen locally assembled small-spotted catshark (*Scyliorhinus canicula*) and thornback ray (*Raja clavata*) transcriptomic data (Debiais-Thibaud et al., 2019) as well as the most recently assembled genome for *S. canicula* (sScyCan1.1, GCA\_902713615.1), using TBLASTN. Additional cDNA sequences were obtained by screening accessible transcriptomic data collected by the SkateBase project<sup>1</sup> [little skate *Leucoraja erinacea* transcriptome (Contig Build-2, GEO:GSM643957) and small-spotted catshark transcriptome (GEO:GSM643958)] using TBLASTN. Small-spotted catshark, little skate, and thornback ray sequences were then used to screen other databases for elephant shark genome assembly (GCA\_000165045.2 *Callorhynchus milii*-6.1.3) (Venkatesh et al., 2014) and whale shark genome *Rhincodon typus* (GCA\_001642345.2 ASM164234v2) (Tan et al., 2019). Thornback ray, small-spotted catshark, and little skate cDNA sequences were used to map synteny on the thorny skate *Amblyraja radiata* and the smalltooth sawfish *Pristis pectinata* draft assembled genomes using TBLASTN [data accessed from, and analyzed in agreement with, Vertebrate Genome Project (Rhie et al., 2020), PriPec2.pri, GCA\_009764475.1]. Syntenic genes in chondrichthyans are

<sup>1</sup><http://skatebase.org/>

*Ddx47* (elephant shark XM\_007909802.1 in GenBank; thorny skate ENSARAT00005031107 in Ensembl Rapid Release) and *Erp27* (elephant shark XM\_007909813.1 in GenBank; thorny skate ENSARAT00005031079 in Ensembl Rapid Release). Synteny data in bony fish genomes were extracted from the Ensembl database for selected genomes [human genome assembly: (GRCh38.p10); mouse *Mus musculus* (GRCm38.p6); chicken *Gallus gallus* (GRCg6a); tropical clawed frog *Xenopus tropicalis* (Xenopus\_tropicalis\_v9.1); elephant shark *Callorhynchus milii* (Callorhynchus\_milii-6.1.3); gar *Lepisosteus oculatus* (LepOcu1); zebrafish *Danio rerio* (GRCz10); Chinese softshell turtle *Pelodiscus sinensis* (PelSin\_1.0); central bearded dragon *Pogona vitticeps* (pvi1.1); reedfish *Erpetoichthys calabaricus* (fErpCal1.1); Asian bonytongue *Scleropages formosus* (fSclFor1.1)] and from NCBI for the caecilian *Microcaecilia unicolor* (aMicUnil.1).

### Phylogenetic Reconstruction

Protein sequences for all identified Mgp and Bgp genes from the different chondrichthyan species together with sequences from osteichthyan species were used for phylogenetic tree reconstruction. These protein sequences are preproteins as they are obtained from the translation of either the cDNA sequence or of a predicted gene from available genomes. All sequences used in this study are detailed with IDs and origin in the **Supplementary Material 1**. Sequences were aligned using MAFFT (Katoh et al., 2002; Katoh and Standley, 2013) using standard parameters (**Supplementary Material 2**). Because a large proportion of the sequences is predicted from genomes and may include false exons, this alignment was then cleaned using HmmCleaner with standard parameters (option “-large”) to remove low similarity segments (Di Franco et al., 2019). Our final alignment used for subsequent phylogenetic reconstruction was 129 amino-acid long and is available in the **Supplementary Material 3**. Phylogenetic analyses were performed on the amino-acid alignment to infer the evolutionary history of these genes. This data set was used to reconstruct gene phylogenies in Maximum Likelihood using IQ-TREE 1.6.1 (Nguyen et al., 2015) under the JTT + I + G4 evolution model for amino-acid data. Node support was estimated by performing a thousand ultra-fast (UF) bootstrap replicates (Hoang et al., 2017) and single branch tests (SH-aLRT; Guindon et al., 2010).

### Protein Domain Description

Conservation of protein domains was evaluated by mapping previously identified functional regions (Laizé et al., 2005) onto the aligned sequences of human, mouse, chicken, zebrafish, elephant shark, and small-spotted catshark Mgp or Bgp proteins. Additional motif recognition was validated on the small-spotted catshark and elephant shark protein sequences with InterPro (Finn et al., 2017), SMART (Letunic et al., 2021), and FIMO version 5.3.0 (Grant et al., 2011).

### Reconciliation Between the Gene Phylogeny and Species Phylogeny

Evolutionary scenario for gene duplication/loss was built minimizing the duplication and loss score with standard

parameters in Treerecs (Comte et al., 2020), using contracted versions of the gene and species trees (**Supplementary Material 4**).

### In situ Hybridization and Histology

Identified small-spotted catshark *Mgp1*, *Mgp2*, and *Bgp* cDNA sequences were used to design the following primers (sequences are given in the 5'-3' orientation): Fw TCACAGATTCACACTCGCTG and Rv GGCCGAACCAGAGC TGCTG amplifying 702 bp for *Mgp1*; Fw CCGATCTCAC AAAGTGGCT and Rv CACAGACTGCAGCAAATAGT amplifying 817 bp for *Mgp2*; Fw CCAGAGAAGATGATGG TCCT and Rv GGGGAATTAACAGAGTCGTC amplifying 675 bp for *Bgp*. Sequences were amplified from cDNA reverse-transcribed from total RNA extractions of a mix of embryonic stages. These PCR products were ligated into the pGEM-T easy vector using the TA cloning kit (Promega). Inserts with flanking T7 and SP6 sites were amplified using M13F/M13R primers and sequenced to verify the amplicon sequence and orientation, and these PCR products were then used as templates for the synthesis of antisense DIG riboprobes [3  $\mu$ l reaction, 100–200 ng PCR product, DIG RNA labeling mix (Roche) with either T7 or SP6 (depending on the amplicon orientation) RNA polymerase (Promega), following manufacturer's instructions]. Before *in situ* hybridization, all DIG-labeled riboprobes were purified on MicroSpin G50 column (GE Healthcare). The obtained expression patterns were different for each probe, excluding detectable cross-hybridization between *Mgp1*/*Mgp2*/*Bgp* probes, so we did not use sense probes as negative control.

Whole embryos of either 6 cm total length, 7.7 cm total length, or 9 cm long hatchlings, were euthanized in buffered tricaine, eviscerated and fixed for 48 h in 4% paraformaldehyde in 1 $\times$  phosphate-buffered saline (PBS) solution at 4  $^{\circ}$ C, rinsed in PBS 1 $\times$  for an hour, and then transferred in 50% ethanol (EtOH)-PBS 1 $\times$ , 75% EtOH-PBS 1 $\times$ , and three successive bathes of 100% EtOH before storage at  $-20^{\circ}$ C in EtOH 100%. We then sampled (i) the lower jaw (hatchling) and (ii) transversal slices in the posterior zone of the branchial arches to allow visualization of gene expression in, respectively, (i) teeth and the Meckel's cartilage; and (ii) abdominal vertebrae, pectoral fin, or branchial rays. Experiments of *in situ* hybridization were performed on 14  $\mu$ m thick cryosections of the chosen samples that had been progressively transferred back to PBS 1 $\times$ , then equilibrated in sucrose 30% for 24 h before being transferred and frozen in Tissue-Tek<sup>®</sup> O.C.T.<sup>™</sup> (Sakura Finetek France SAS). Consecutive cryosections were distributed on 10 successive slides to a maximum of 6–8 sections per slide and were stored at  $-20^{\circ}$ C. *In situ* hybridization on sections was performed as described previously (Enault et al., 2015) with stringent conditions of hybridization at 70 $^{\circ}$ C. *In situ* hybridization results were taken with Hamamatsu NanoZoomer 2.0-HT Slide Scanner (Montpellier RIO Imaging facility, INM Optique) with a 40 $\times$  objective.

Histological staining (Hematoxylin-Eosine-Saffran) was performed at the local histology platform (RHEM platform

at IRCM, Montpellier) on 7  $\mu\text{m}$  paraffin sections of non-demineralized samples on a histology automaton.

## RNA Isolation and qPCR Analysis

Early embryos (three or four for each stage) were collected from embryonic stages 18–32 (Ballard et al., 1993), with stage 32 embryos <3.5 cm total length. Total RNA was isolated with ReliaPrep RNA tissue Miniprep system according to the supplier's instructions (Promega), and their quality was verified on a Bioanalyzer 2100 instrument (Agilent): 500 ng of total RNA were used for cDNA preparation performed by Superscript II reverse transcriptase (Invitrogen) with an oligodT primer.

For quantitative PCR, 1:20 dilution of each cDNA was run in triplicate on a 384-well plate for each primer pair by using thermal cycling parameters: 95°C for 2 min, 95°C for 10 s, 68°C for 10 s, 72°C for 10 s (45 cycles), and an additional step 72°C for 10 min performed on a Light Cycler 480 with the SensiFAST SYBR No-ROX kit (Meridian Bioscience) (qPHD UM2/GenomiX Platform, Montpellier – France). Results were normalized with the expression of two reference genes *Eef1a* and *Rpl8* [previously used in elasmobranch fishes (O'Shaughnessy et al., 2015; Onimaru et al., 2016)] by geometric mean, and data were further analyzed with the Light Cycler 480 software 1.5.1.

We used Primer 3.0 to design all the sets of forward and reverse primers to amplify selected genes (sequences are given in the 5'-3' orientation): Fw TCGGGAGGAGAGATGCACAT and Rv TGCCACCAAAGTATCTGCCA amplifying 183 bp for *Mgp1*; Fw CCTGATTCTGCTGTGCCTGT and Rv TTTTCCATAGGC CGCCATGT amplifying 277 bp for *Mgp2*; Fw TGATGGT CCTTCCTCGGGA and Rv TGGTATCCAATCCTGTTTGC CA amplifying 180 bp for *Bgp*; Fw GGTGTGGGTGAATTT GAAGC and Rv TTGTACCATGCCAACCAGA amplifying 245 bp for *Eef1a*; Fw TTCATTGCAGCGGAGGGAAT and Rv TCAATACGACCACCACCAGC amplifying 302 bp for *Rpl8*.

The expression data obtained were compared over time to test if any gene was differentially expressed in time with a one-way ANOVA. A Shapiro–Wilk normality test was applied on the log transformed data, and for each gene the null hypothesis of normality was kept ( $P > 0.05$ ). We tested for heteroscedasticity of variance between developmental stages, and the null hypothesis had to be rejected only for the *Bgp* gene ( $P < 0.05$ ), even after log transformation. Note that we are very constrained by an unbalanced protocol (different number of observations in each developmental stage) and small sample size, which limits statistical power.

## Embryo Collection and Ethics Statement

Embryos of the small-spotted catshark *S. canicula* originated from a Mediterranean population of adult females housed at Station Méditerranéenne de l'Environnement Littoral, Sète, France. Handling of small-spotted catshark embryos followed all institutional, national, and international guidelines [European Communities Council Directive of September 22, 2010 (2010/63/UE)]: no further approval by an ethics committee was necessary as the biological material is embryonic and no live experimental procedures were carried out. Embryos were raised in seawater tanks at 16–18 °C and euthanized by overdose

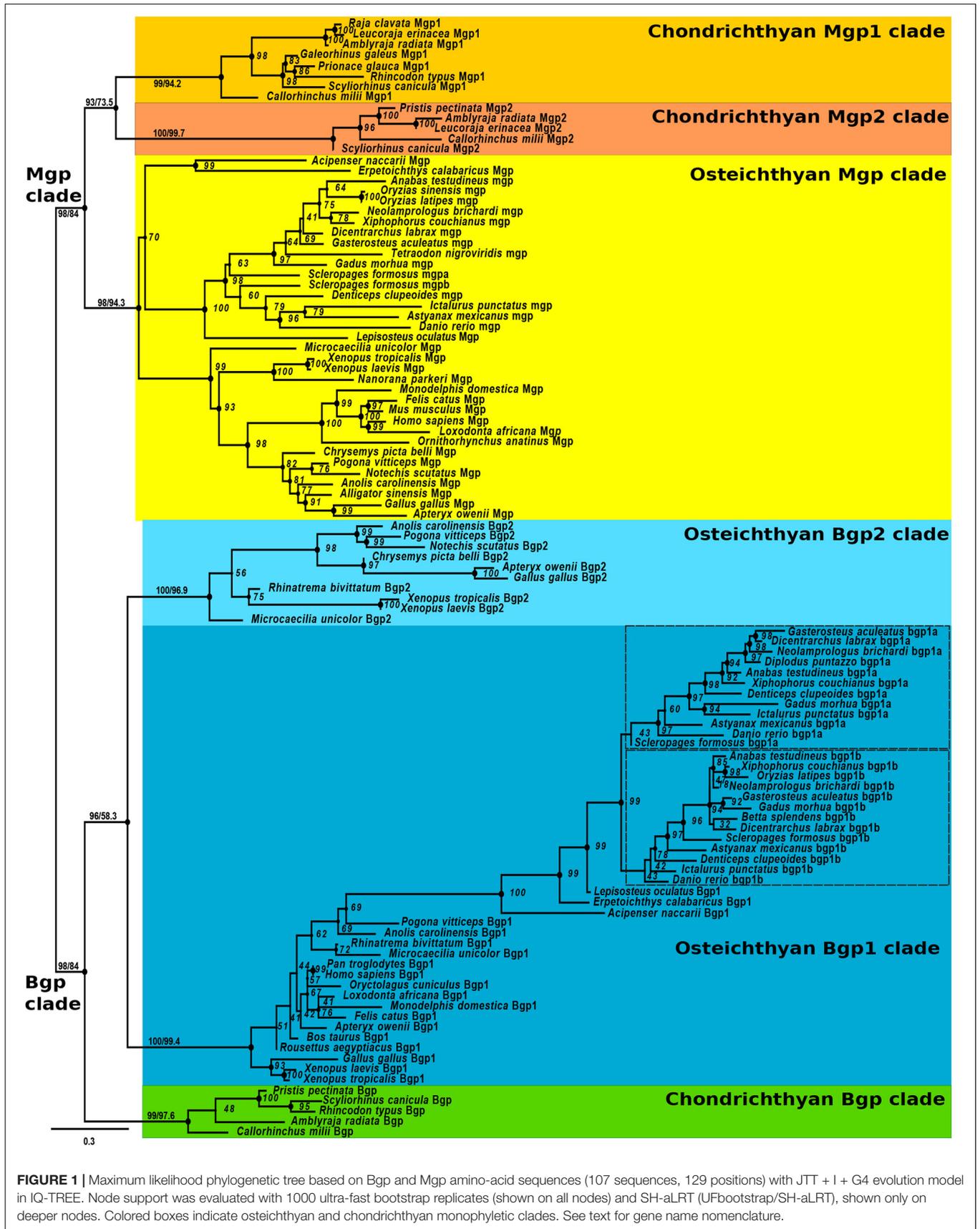
of tricaine (MS222, Sigma) at appropriate stages (Ballard et al., 1993; Enault et al., 2016).

## RESULTS

### Evolution of the Mgp/Bgp Gene Complement in Jawed Vertebrates

Three transcripts were identified as Mgp or Bgp genes in the small-spotted catshark transcriptome and named after their position in the phylogenetic reconstruction: *Mgp1*, *Mgp2*, and *Bgp* (Figure 1). To perform this reconstruction, we screened other available cartilaginous fish genomes as well as the genomes of several bony fishes by reciprocal blasts to recover a maximum of Mgp/Bgp sequences in the jawed vertebrate clade. The produced alignment was 129 amino acid long after HmmCleaner (alignment available as **Supplementary Material 3**). The major limitation on the analysis of this phylogeny was the lack of an out-group: no potential Mgp/Bgp sequence could be identified in the available genomic and transcriptomic sequences for cyclostome species (e.g., lamprey or hagfish), and there is currently no identified closely related gene family in jawed vertebrates. Both Mgp and Bgp clades in bony fishes were monophyletic and had a closest monophyletic group made of cartilaginous fish sequences (see **Supplementary Material 5** for the unrooted tree), which made us place the putative root of this tree as resulting in Figure 1, leading to one Mgp and one Bgp clade for jawed vertebrates. This choice implies that one ancestral Bgp and one ancestral Mgp genes were already present in the last common ancestor of extant jawed vertebrates, as previously suggested (Laizé et al., 2005).

In this phylogenetic reconstruction, two cartilaginous fish genes were identified as duplicated copies grouping together as the sister group to a single *Mgp* copy in bony fishes (UF-bootstrap and SH-aLRT support reach acceptable values at this chondrichthyan node, although they are lower than for other deep nodes): the two chondrichthyan copies were named *Mgp1* and *Mgp2* (Figure 1). In the *Bgp* clade, cartilaginous fish sequences were monophyletic and strongly supported by the SH-aLRT statistic and UF-bootstrap, with only one *Bgp* gene in each species, whereas bony fish sequences grouped into two sister clades, suggesting two osteichthyans *Bgp* paralogs well supported by the UF-bootstraps and SH-aLRT (Figure 1). One of these bony fish paralogs is best known as the *osteocalcin/Bgp* gene product in all screened actinopterygians and sarcopterygians (also previously named *OCI*; Cancela et al., 2014). To account for the different nature of the described paralogs, we will further identify this clade as *Bgp1*: although our phylogenetic reconstruction leads to little resolution within this clade, its monophyly is very robust in the tree (SH-aLRT = 99.4; UFboot = 100). The second osteichthyan *Bgp* paralog is herein named *Bgp2*: it includes sequences found only in lissamphibians and sauropsids (including birds). This *Bgp2* gene is predicted but most frequently not annotated in the Ensembl or NCBI databases (for *Chrysemys*, the kiwi bird and the tiger snake) or named *Mgp-like* in *Pogona*, *osteocalcin-like* in *Xenopus laevis* and other lissamphibians, *osteocalcin* in *X. tropicalis* or *osteocalcin 3*

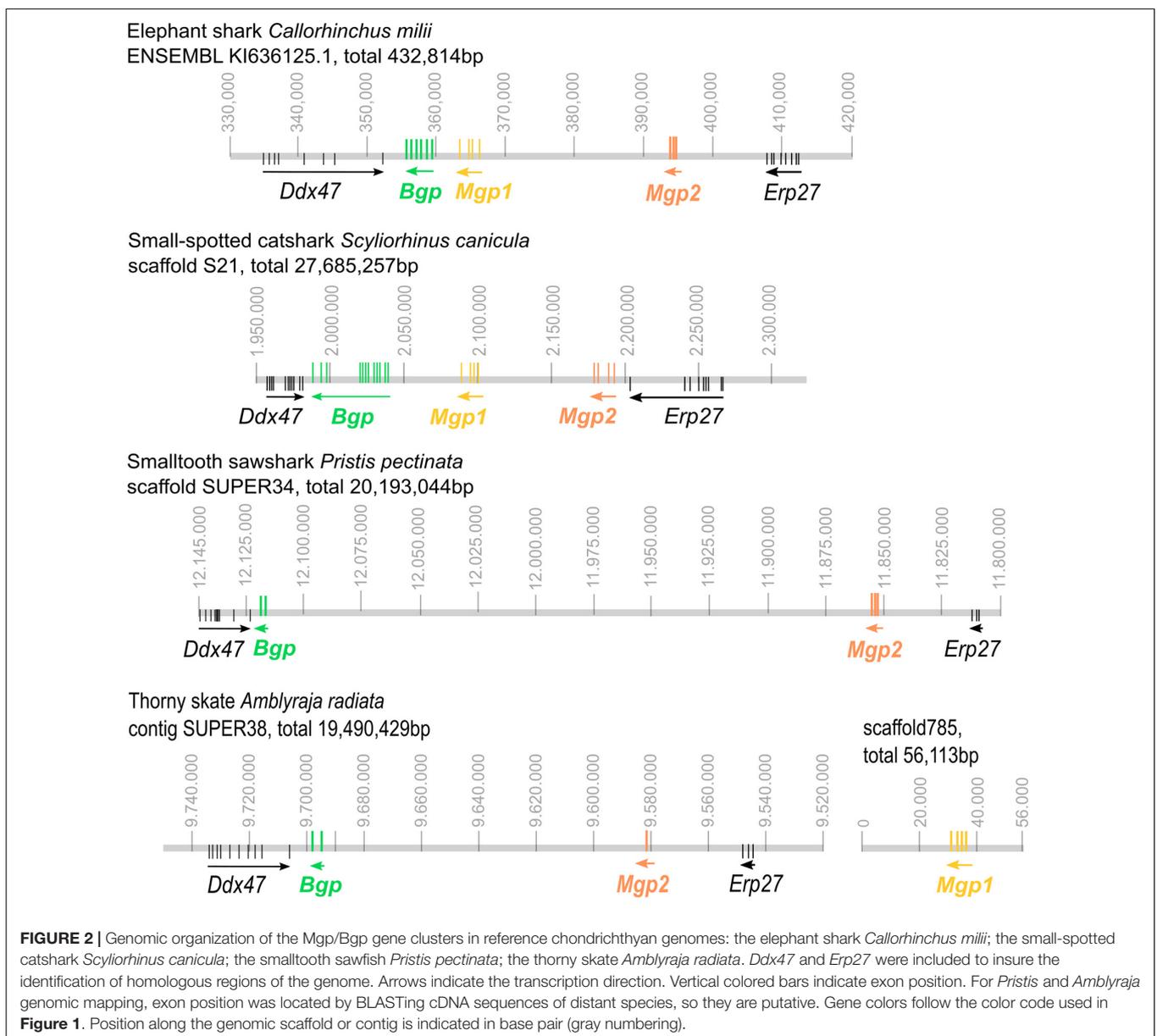


**FIGURE 1 |** Maximum likelihood phylogenetic tree based on Bgp and Mgp amino-acid sequences (107 sequences, 129 positions) with JTT + I + G4 evolution model in IQ-TREE. Node support was evaluated with 1000 ultra-fast bootstrap replicates (shown on all nodes) and SH-aLRT (UFbootstrap/SH-aLRT), shown only on deeper nodes. Colored boxes indicate osteichthyan and chondrichthyan monophyletic clades. See text for gene name nomenclature.

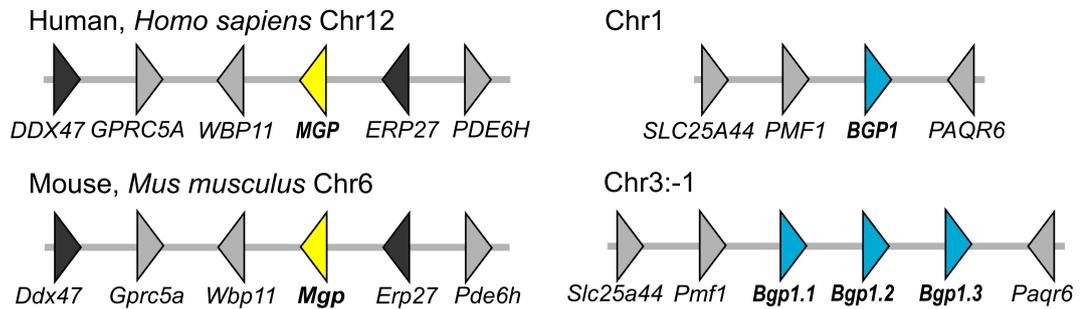
in the chicken [see all references to the extracted sequences in **Supplementary Material 1**; this paralog has also previously been named OC3 (Cancela et al., 2014)]. As a consequence, this topology suggests an event of duplication of an ancestral bony fish *Bgp* gene leading to these *Bgp1* and *Bgp2* paralogs (**Figure 1**). Another event of duplication is deduced from the two sister clades observed within teleost fishes in the *Bgp1* group: this and synteny data (see below) support these paralogs to originate from the teleost-specific whole-genome duplication (Amores et al., 1998), so we followed the accepted gene nomenclature and named them *bgp1a* (usually annotated *bglap* or *osteocalcin* in public databases) and *bgp1b* [previously named OC2 (Cancela et al., 2014), or *bglap-like* in databases, see **Supplementary Material 1**].

## Genomic Organization of the Mgp and Bgp Genes in Jawed Vertebrates

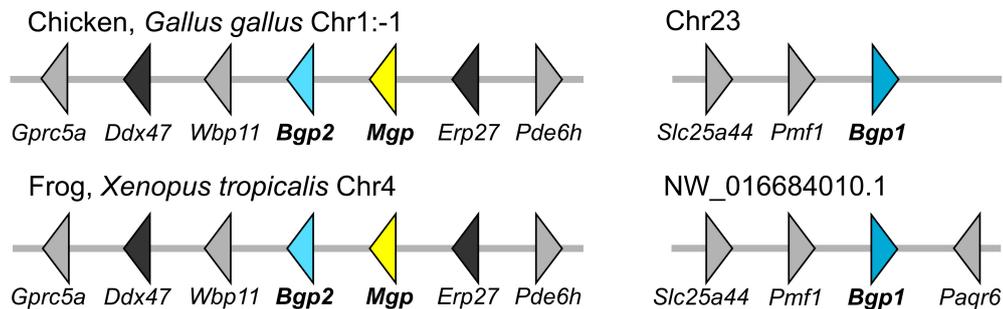
All three coding sequences were predicted in the available elephant shark genome and all assigned to a single genomic contig (**Figure 2**) together with two genes bordering the syntenic regions, *Erp27* and *Ddx47*, as identified in other syntenic regions from bony fishes (see **Figure 3**). The identified cDNA sequences of *Mgp1*, *Mgp2*, and *Bgp* could be assigned to a single scaffold in the small-spotted catshark draft genome in synteny with *Ddx47* and *Erp27* (see **Figure 2**). In two batoid genomes (*Amblyraja* and *Pristis*), *Mgp2* and *Bgp* genes could be assigned to a single contig together with *Erp27* and *Ddx47*. However, the *Mgp1* gene was located on another scaffold in the *Amblyraja* genome, outside of the locus identified by the presence of *Erp27* and



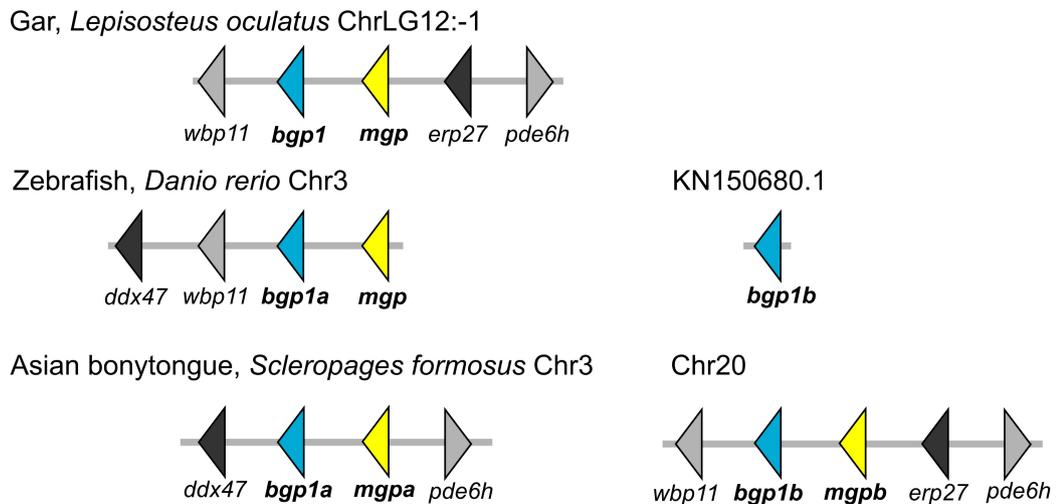
**A Mammals**



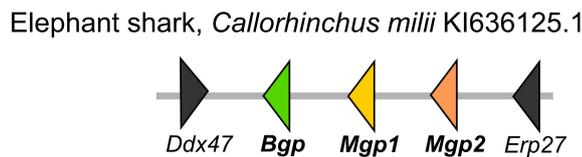
**B Amphibians and sauropsids**



**C Actinopterygians**



**D Chondrichthyans**



**FIGURE 3 |** Genomic organization of the *Mgp/Bgp* gene clusters in reference osteichthyan genomes as annotated in currently available databases. **(A)** Two mammalian genomes with separated *Mgp* and *Bgp* loci; **(B)** Two non-mammalian tetrapod genomes with one *Bgp* locus, and one tandem *Bgp* and *Mgp* genes on homologous loci; **(C)** One non-teleost actinopterygian with only one locus where *Mgp* and *Bgp* genes are tandemly organized, and two teleost genomes with two loci, where *mgp* and *bgp* are tandemly organized or single. **(D)** The elephant shark as a representative of chondrichthyans. Several syntenic genes were selected to support the homology of the compared loci. Distance between genes is not to scale. Gene names and corresponding color refer to our phylogenetic analyses, and the correspondence to gene names in databases is found in the **Supplementary Material 1**.

*Ddx47* (Figure 2), and no *Mgp1* gene could be identified in the *P. pectinata* genome.

The comparison to the genomic data in bony fishes was made in two steps. First, an overview of the genomic locations in tetrapods shows that in the human genome, there are separated loci for the *Mgp* (chromosome 12) and the *Bgp1* (BGLAP on chromosome 1) genes for which we highlighted the position of syntenic genes (Figure 3). All occurrences of the tetrapod *Bgp2* gene (as identified in our phylogenetic reconstruction) are in the *Mgp* locus in lissamphibians and sauropsids (Figure 3, and verified by BLAST on Ensembl available genomes of *P. sinensis*, *M. unicolor*, and *P. vitticeps*, not shown).

In a second step, we searched the homologous loci in actinopterygians (ray-finned fishes) outside of teleost fishes: the syntenic markers linked to *Bgp1* in tetrapods would not co-localize with any known sequence of either *Bgp* or *Mgp* in actinopterygians (Figure 3). The actinopterygian *Bgp1* gene was located in the *Mgp* locus, as defined by the presence of syntenic markers such as *Erp27* and *Ddx47* (Figure 3, and verified by BLAST on the available genome of *E. calabaricus*). This *Bgp1* copy is identified as *Bglap* or *Osteocalcin-like* (however, not annotated in the spotted gar) in the available databases (but see Supplementary Material 1 for predicted gene IDs). Within teleost fishes, the zebrafish *D. rerio* is usually used as a reference species, however, the contig where *bgp1b* is located is very short and does not give syntenic gene markers, while *bgp1a* and *mgp* are located close to each other on chromosome 3 (Figure 3). In Figure 3, we illustrate the genomic loci in the Asian bonytongue *S. formosus*, where each of the two teleost-specific copies of *bgp1* are found adjacent to one *mgp* gene that we named *mgpa* and *mgpb*, each of these genomic loci with either a sequence coding for *erp27* or *ddx47*, but both regions including a paralog of the *pde6h* gene. In all other teleost genomes that we have screened (see the sequences chosen for the phylogenetic reconstruction, and Supplementary Material 1), the *mgp* sequence was found in synteny with the *bgp1a* sequence, together with *ddx47/wbp11*, while the *bgp1b* sequence was found with *erp27* but without another copy of *mgp* (not shown).

## Protein Domains

The prediction of functional protein domains by InterproScan and SMART led to the recognition of a signal peptide for all sequences, but of a general Gla domain only in Bgp and Mgp1 proteins, excluding the Mgp2 sequences of the small-spotted catshark or elephant shark. The FIMO algorithm also identified a furin cleavage site in the Mgp2 sequence (see Figure 4). This was unexpected as it is typical for Bgp proteins but not of Mgp (Laizé et al., 2005). To further describe the presence, absence, and conservation of functional protein domains, we aligned characterized protein sequences of either Mgp or Bgp proteins (from human, mouse or chicken, and zebrafish) to those of the small-spotted catshark and elephant shark (Supplementary Material 6 and 7) and identified the expected location of specific functional domains of Mgp and Bgp proteins as previously described (Laizé et al., 2005).

The central motif for the Gla domain ExxxExC could be identified in the Mgp2, as well as in Mgp1 and Bgp sequences

(Figure 4). However, the C-terminal part of the Gla domain was poorly aligned in the Mgp2 sequences, suggesting a divergent Gla domain in the Mgp2 paralog. In addition, no phosphorylation site could be identified in the Mgp2 sequences.

Mgp1 sequences (both from the small-spotted catshark and elephant shark) displayed well-conserved signal peptide, phosphorylation site, and general Gla domain (Figure 4). The expected ANxF site upstream to the Gla domain and supposed to participate in the docking site for the gamma-carboxylase (Viegas et al., 2013) was conserved in the elephant shark but modified to AHSF in the small spotted catshark questioning the functionality of this site (Figure 4).

In Bgp protein sequences, a signal peptide was also well conserved, followed by a furin cleavage site in the elephant shark that was not predicted in the small spotted catshark sequence because of a modification to KKSKR (Figure 4). A well-conserved Gla domain including the highly conserved Gla motif ExxxExC was present in the elephant shark and small-spotted catshark Bgp sequences.

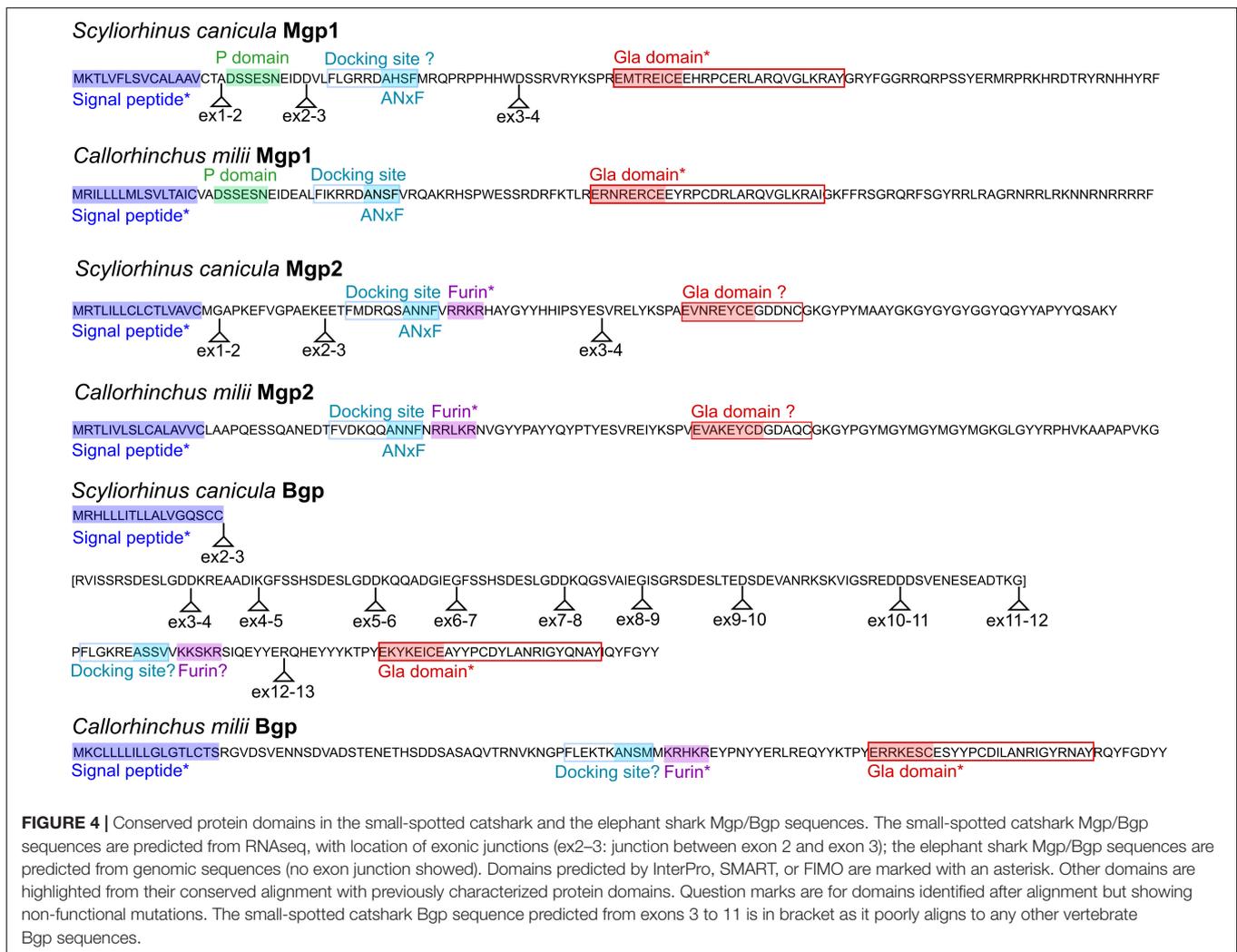
By aligning each cDNA sequence with the genomic locus, we could map the exonic junctions on the full length protein sequences of the small-spotted catshark: *Mgp1* and *Mgp2* display conserved intron/exon structure [four exons, ATG and peptide signal coding sequence in the first exon, docking site coding sequence in the third exon, and Gla domain in the fourth exon: see Figure 4 and compare to bony fishes (Laizé et al., 2005; Viegas et al., 2013)]. On the other hand, the small-spotted catshark Bgp sequence displayed a divergent exon-intron structure [as compared to bony fishes (Laizé et al., 2005; Viegas et al., 2013)]: 13 exons and a series of imperfect repeat sequences between exons 3 and 12 (exons 3, 5, 7 code for very similar protein sequences), revealing important divergence of the gene structure. Because our cDNA sequence is reconstructed from RNAseq data, we cannot exclude the existence of splicing variants that would not include these extra-exons. Also, the elephant shark sequence does not include these repeated exons so they may be specific for the lesser spotted catshark (so a product of recent evolution).

## Gene Expression Patterns in the Embryonic Small-Spotted Catshark *Scyliorhinus canicula*

All three identified *Mgp/Bgp* sequences generated distinct expression patterns in the small-spotted catshark embryos by *in situ* hybridization. The selected stages of development were chosen in order to cover one time point before and another after the initiation of mineralization in the developing vertebrae (Enault et al., 2016) and during tooth development.

### Mgp1 Expression

In the 6 cm long embryo, the expression of *Mgp1* was detected in the developing vertebrae: in the cartilaginous core of neural arches, in a cartilaginous ring surrounding the notochord and also in notochordal cells (Figures 5A,B). At this stage, these zones of expression are not mineralized (Figure 5C; see Enault et al., 2016), but neural arches and the cartilage surrounding the notochord will show strong mineralization in



**FIGURE 4 |** Conserved protein domains in the small-spotted catshark and the elephant shark Mgp/Bgp sequences. The small-spotted catshark Mgp/Bgp sequences are predicted from RNAseq, with location of exonic junctions (ex2–3: junction between exon 2 and exon 3); the elephant shark Mgp/Bgp sequences are predicted from genomic sequences (no exon junction showed). Domains predicted by InterPro, SMART, or FIMO are marked with an asterisk. Other domains are highlighted from their conserved alignment with previously characterized protein domains. Question marks are for domains identified after alignment but showing non-functional mutations. The small-spotted catshark Bgp sequence predicted from exons 3 to 11 is in bracket as it poorly aligns to any other vertebrate Bgp sequences.

embryos measuring 7.7 cm (Figure 5E; see Enault et al., 2016). On the 7.7 cm long embryo, the expression of *Mgp1* was no longer detected in neural arches, appeared faint in the cartilage surrounding the notochord, but was still strong in the notochord which is not a site of mineralization (Figure 5D). In the Meckel’s cartilage, the expression of *Mgp1* was not detected in developing teeth but was detected in a sub-perichondral population of chondrocytes (Figure 5F) at a time when no mineralization has started in the lower jaw cartilage (Figure 5G), but in a zone prefiguring the site of tesseral mineralization (Enault et al., 2015). Further expression in chondrocytes was detected in the pectoral girdle cartilages in a sub-perichondral layer of chondrocytes located in a contact zone between two cartilages (Figure 5H, filled arrowhead). Finally, expression of *Mgp1* was observed in gills, both in the endothelium of the vascular system and in undifferentiated mesenchyme surrounding vascularization (Figure 5H).

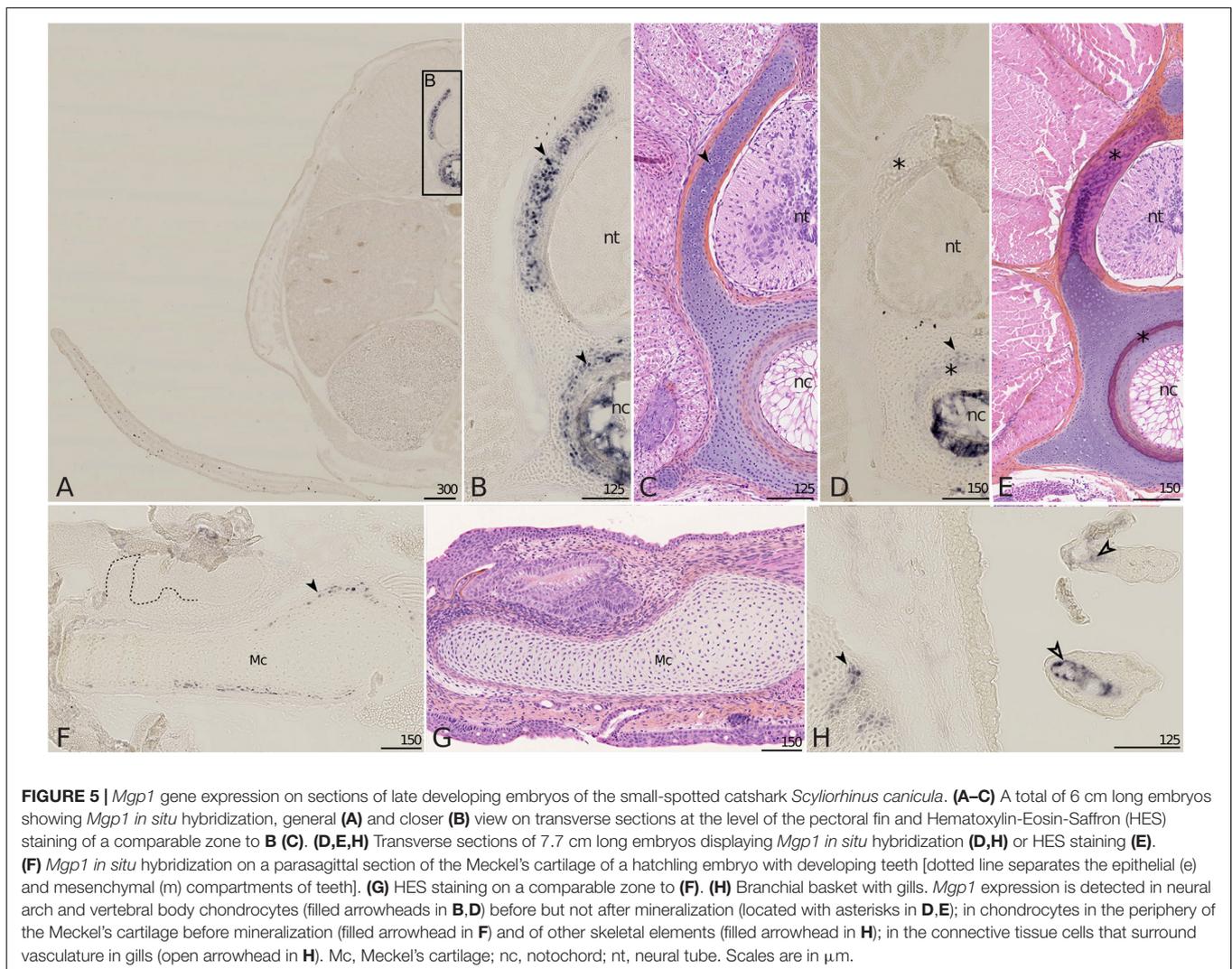
**Mgp2 Expression**

The expression of the *Mgp2* gene in the small-spotted catshark was restricted and could be observed with very strong signal in

the developing fins, in 6 (Figures 6A,B) and 7.7 cm (not shown) long embryos, in the mesenchymal tissue surrounding and most probably synthesizing ceratotrichiae, the semi-rigid fibers that make up the fin support in cartilaginous fishes. Weaker signal was detected in developing unmineralized tooth bud of the lower jaw (Figure 6C).

**Bgp Expression**

Bone Gla protein showed a widespread low-level expression in many connective tissues in the 7.7 cm long embryo (Figures 7A,B) but could not be detected in any chondrocyte population, neither in early (data not shown) or late stage vertebrae (Figure 7B) nor in Meckel’s cartilage (Figure 7E). Stronger detection of *Bgp* expression was observed in the cells of the nerve root (Figure 7B), the mesenchymal cells of scale buds at a placode stage (Figure 7C, filled arrowhead), mesenchymal cells in connective tissues surrounding muscles of the branchial apparatus with strong expression in the zone of attachment between muscle fibers and cartilaginous units (Figure 7D, black arrow), few mesenchymal cells of mineralized teeth (Figure 7E). Some weaker signal could be detected in the epithelium and



**FIGURE 5 |** *Mgp1* gene expression on sections of late developing embryos of the small-spotted catshark *Scylliorhinus canicula*. **(A–C)** A total of 6 cm long embryos showing *Mgp1* *in situ* hybridization, general **(A)** and closer **(B)** view on transverse sections at the level of the pectoral fin and Hematoxylin-Eosin-Saffron (HES) staining of a comparable zone to **B (C)**. **(D,E,H)** Transverse sections of 7.7 cm long embryos displaying *Mgp1* *in situ* hybridization **(D,H)** or HES staining **(E)**. **(F)** *Mgp1* *in situ* hybridization on a parasagittal section of the Meckel's cartilage of a hatching embryo with developing teeth [dotted line separates the epithelial (e) and mesenchymal (m) compartments of teeth]. **(G)** HES staining on a comparable zone to **(F)**. **(H)** Branchial basket with gills. *Mgp1* expression is detected in neural arch and vertebral body chondrocytes (filled arrowheads in **B,D**) before but not after mineralization (located with asterisks in **D,E**); in chondrocytes in the periphery of the Meckel's cartilage before mineralization (filled arrowhead in **F**) and of other skeletal elements (filled arrowhead in **H**); in the connective tissue cells that surround vasculature in gills (open arrowhead in **H**). Mc, Meckel's cartilage; nc, notochord; nt, neural tube. Scales are in  $\mu\text{m}$ .

mesenchyme of non-mineralized tooth buds (**Figure 7E**). *Bgp* expression could also be detected in gill tissues, restricted to the connective mesenchyme that surrounds the vascular system (open arrow), but its expression could not be observed in the vascular endothelium as seen with *Mgp1* (**Figures 7D,F** and compare with **Figure 5H**). Finally, *Bgp* expression was detected in cells of the pectoral fin tip, in the mesenchymal tissue surrounding ceratotrichia in 6 cm long (not shown) and 7.7 cm long embryos (**Figure 7C**, open arrowhead).

### Embryonic Patterns of Expression

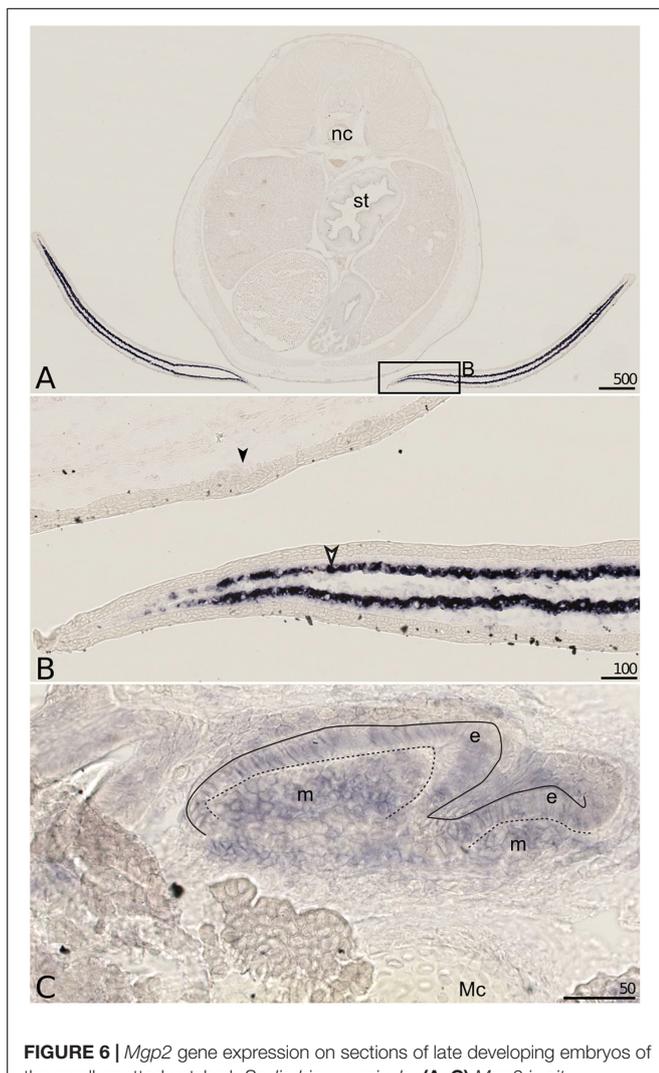
Total RNA extracts obtained from whole embryos of the small-spotted catshark from stage 18 (end of neurulation) to stage 32 (late organogenesis) (Ballard et al., 1993) allowed the evaluation of relative expression levels for *Bgp*, *Mgp1*, and *Mgp2* over the course of organogenesis in the small-spotted catshark (**Figure 8**). *Bgp* expression generally tended to be higher than the expression of the Mgp genes during the stages 18–32, to the exception of *Mgp2* expression at stage 32 (**Figure 8**). The results of the one-way ANOVA testing for gene expression

variation over developmental stages were non-significant for the genes *Mgp1* and *Bgp*. However, the one-way ANOVA for the *Mgp2* gene indicated a difference between group means at the  $P < 0.1$  threshold, probably due to the higher expression level observed at the stage 32. Stage 32 may be the stage of initiation of ceratotrichiae development (there is no sign of ceratotrichiae in pectoral or pelvic fins in stage 30 embryos in Tanaka et al., 2002) explaining the initiation of stronger expression at stage 32.

## DISCUSSION

### An Evolutionary Scenario for Mgp/Bgp Gene Duplicates

Syntenic and phylogenetic data gathered in this study allow drawing an evolutionary scenario for the genomic organization and diversification of the Mgp/Bgp gene family, under a most-parsimonious model of evolution (**Figure 9** and **Supplementary Material 4**). In the bony fishes, our data allow testing previously



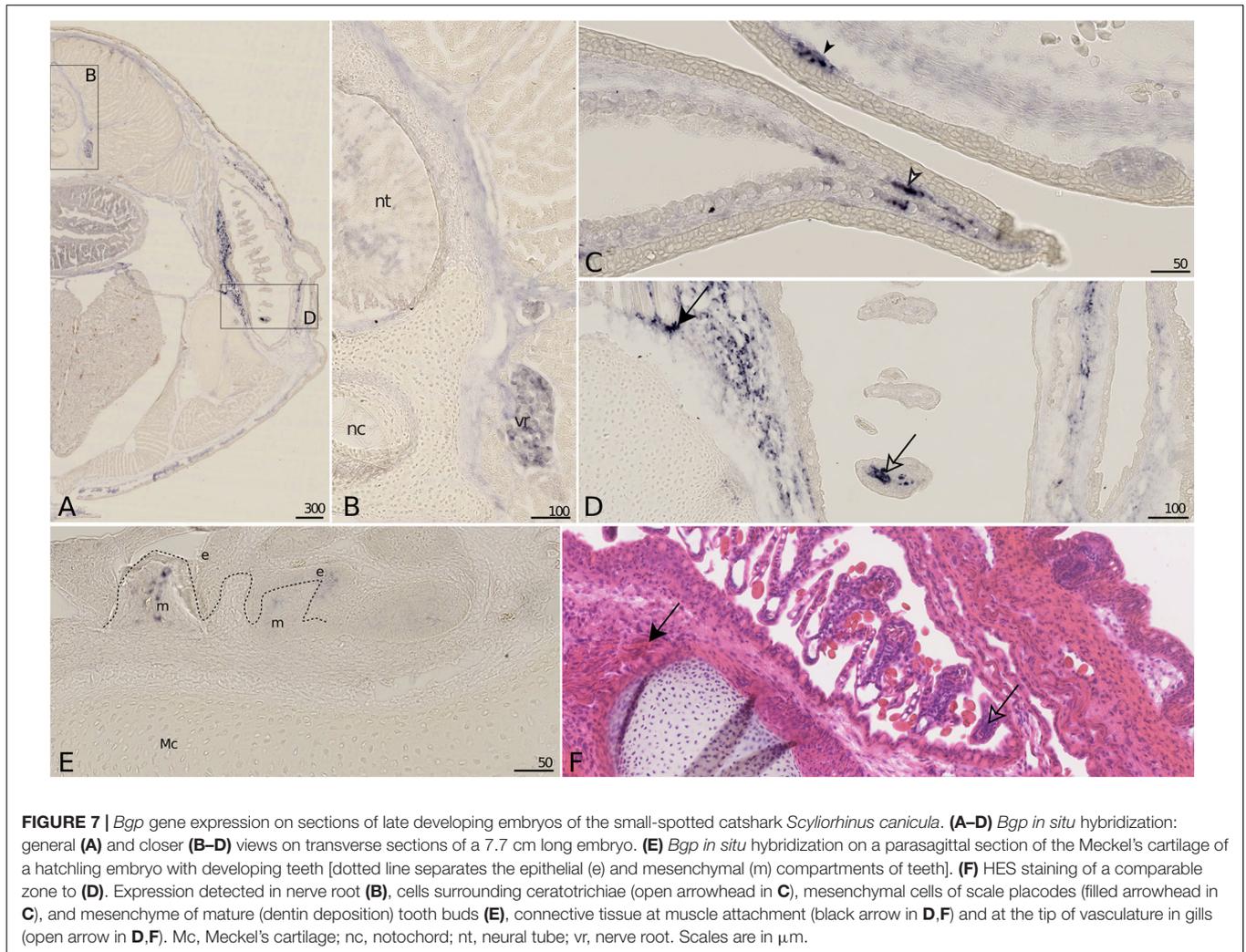
**FIGURE 6** | *Mgp2* gene expression on sections of late developing embryos of the small-spotted catshark *Scyliorhinus canicula*. **(A–C)** *Mgp2* *in situ* hybridization: general **(A)** and closer **(B)** views on transverse sections of a 6 cm long embryo. **(C)** *Mgp2* *in situ* hybridization on a parasagittal section of the Meckel's cartilage (Mc) of a hatchling embryo with developing teeth (e, epithelial compartment; m, mesenchymal compartment of tooth buds; dotted line separates these two compartments). Expression detected in cells surrounding ceratotrichiae (open arrowhead in **B**), epithelial and mesenchymal cells of developing (unmineralized) tooth buds (in **C**). Mc, Meckel's cartilage; nc, notochord; st, stomach. Scales are in  $\mu\text{m}$ .

proposed hypotheses. The phylogenetic relationships between *Bgp1* and *Bgp2* (Figure 1) suggest that these two copies emerged from a gene duplication in the last common ancestor of bony fishes which is congruent with previous identification and phylogenetic reconstruction including *Bgp2* [previously named OC3 (Cancela et al., 2014)] where data from chondrichthyans were missing. This node (and others) still displays poor robustness when tested with SH-aLRT: these low values may be dependent on the little number of positions in our alignment (129 aa), a tendency which amplifies with higher number of protein sequences in the alignment and which cannot be easily corrected for, due to the small length of the studied proteins.

In addition, we show that a translocation of *Bgp1* most probably occurred in the sarcopterygian or tetrapod stem lineages, while *Bgp2* was lost convergently in actinopterygians and mammals (Figure 9). Unfortunately, no sequence homologous to *Bgp* could be identified in the available genomic databases of the coelacanth (NCBI or Ensembl, by TBLASTN search of the gar *Bgp1* sequence), which could have helped in determining more precisely the timing of the *Bgp1* translocation. Finally, our phylogenetic reconstruction and teleost genome data-mining allowed the annotation of the previously named OC2 gene (Cancela et al., 2014) as one of the two *bgp1* paralogs (Figures 1, 9) generated by the teleost-specific whole-genome duplication (Amores et al., 1998). We also identified two *mpg* co-orthologs in the Asian bonytongue genome, in tandem organization with each of the *bgp1a* and *bgp1b* copies, with two *pde6h* gene copies and located in synteny with either *erp27* or *ddx47* (Figure 3), supporting that the duplicated genes could have originated from the teleost-specific genome duplication. However, these two *mpg* copies were found only in one species within the genomes available in Ensembl: only one *mpg*, in synteny with *bgp1a*, is found in all other examined teleost genomes, which would imply a secondary loss of this *mpg* gene duplicate in all examined taxa. As a consequence, further analysis of the genomic data in teleost fishes is still needed to support this scenario.

In the chondrichthyan lineage, we uncovered a specific tandem duplication in the *Mgp* locus leading to the *Mgp1* and *Mgp2* genes. Within chondrichthyans, an additional event of translocation may have occurred for the *Mgp1* copy in the batoid lineage (observed in *A. radiata*, Figure 2). However, because this is a single observation and because the *Amblyraja Mgp1* copy was identified in a short scaffold, we still cannot rule out the possibility of an assembly artifact. Additional genomic data from batoids are necessary to test the robustness of this observation.

Our results demonstrate that the location of the actinopterygian *Bgp1* and of the chondrichthyan *Bgp* is in the *Ddx47/Erp27* locus, which suggests an ancestral location of *Bgp* in this locus, later followed by tandem duplication that generated *Bgp1* and *Bgp2* in bony fishes. The ancestral *Mgp* and *Bgp* genes were, in this scenario, tandem duplicates in the last common ancestor of jawed vertebrates (Figure 9). No other closely related genes to *Mgp/Bgp* family have been reported for jawed vertebrates. In addition, no similar sequence was found in the genomic data of the lamprey (although the *Ddx47/Wbp11* locus can be identified on chromosome 3 of the kPetMar1 assembly), nor in *Amphioxus* nucleotide dataset (NCBI). Taken together, these three last arguments let us hypothesize that the evolution of *Mgp/Bgp* family cannot be explained by two rounds of whole genome duplications (Ohno, 1970), that occurred before the divergence of jawed vertebrates and resulted in expansion of many gene families from one to four genes located in different paralogs (Dehal and Boore, 2005). The inability to detect closely related gene families might be explained by several scenarios: (i) complete loss of other paralogs, ancestrally to jawed vertebrates [most frequently observed situation (Blomme et al., 2006)], (ii) rapid and extensive evolution of the coding sequences making sequence similarity searches inefficient, and (iii) *de novo*



**FIGURE 7** | *Bgp* gene expression on sections of late developing embryos of the small-spotted catshark *Scyliorhinus canicula*. **(A–D)** *Bgp in situ* hybridization: general **(A)** and closer **(B–D)** views on transverse sections of a 7.7 cm long embryo. **(E)** *Bgp in situ* hybridization on a parasagittal section of the Meckel's cartilage of a hatchling embryo with developing teeth [dotted line separates the epithelial (e) and mesenchymal (m) compartments of teeth]. **(F)** HES staining of a comparable zone to **(D)**. Expression detected in nerve root **(B)**, cells surrounding ceratotrichiae (open arrowhead in **C**), mesenchymal cells of scale placodes (filled arrowhead in **C**), and mesenchyme of mature (dentin deposition) tooth buds **(E)**, connective tissue at muscle attachment (black arrow in **D,F**) and at the tip of vasculature in gills (open arrow in **D,F**). Mc, Meckel's cartilage; nc, notochord; nt, neural tube; vr, nerve root. Scales are in  $\mu\text{m}$ .

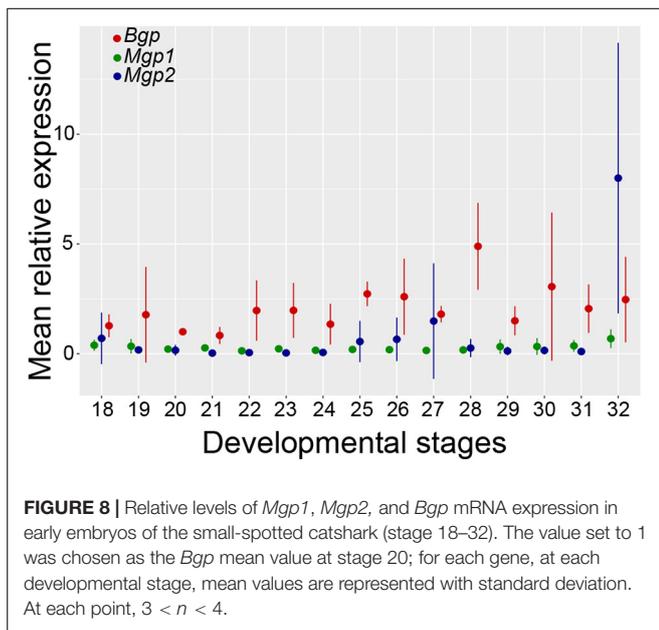
evolution of an ancestral *Bgp/Mgp* gene after the two rounds of genome duplication (Van Oss and Carvunis, 2019).

With the evolutionary scenario presented here, the orthology relationships between jawed vertebrate genes of the *Mgp/Bgp* family are more complex than usually considered, as the *Bgp* gene found in cartilaginous fishes is not a one-to-one ortholog to the *Bgp* copy (*Bgp1*) found in actinopterygian (non-teleost) fishes or in mammals. In addition, the *Bgp1* copy found in sarcopterygian genomes has gone through a translocation event that may have modified the transcriptional regulation, and therefore the function, of its orthologous copy in sarcopterygian fishes.

## Diversity of Expression Patterns in Cartilaginous Fishes and Functional Implications

Previous hypotheses accounting for the evolution of *Mgp/Bgp* sequences relied on partial sequence data and proposed that only a single *Mgp* gene was present in cartilaginous fishes (Cancela et al., 2014). The survey of transcriptomic and genomic data here reveals the presence of three genes, two

*Mgp* genes and one *Bgp*. Few significant aspects of *Mgp/Bgp* gene evolution in chondrichthyans can be derived from the conservation of functional protein domains. *Mgp1*, as partly previously described in shark (Price et al., 1994; Rice et al., 1994; Ortiz-Delgado et al., 2006), displays well-conserved signal peptide, phosphorylation sites, carboxylase docking site, and a full Gla domain. This Gla domain is known to be able to bind calcium when secreted in the extracellular matrix and then acts as an inhibitor of mineralization under the condition that *Mgp* protein is phosphorylated (Schurgers et al., 2007). On the other hand, our data suggest the divergence of the Gla domain in the *Mgp2* protein, although the core Gla motif was observed in our alignments, together with a loss of the phosphorylation domain that follows the signal peptide in other *Mgp* proteins (Figure 4). From these observations, we could expect *Mgp1* to display a conserved *Mgp* function as known in bony fishes, while *Mgp2* may have undergone partial or complete change of function. These observations suggest that after the *Mgp1/Mgp2* duplication event in cartilaginous fishes, the *Mgp2* copy underwent neofunctionalization, while *Mgp1* kept the ancestral function (Ohno, 1970). The chondrichthyan



*Bgp* preprotein as described from transcriptomic data of the small-spotted catshark shows conserved signal peptide, followed by a long stretch of non-conserved amino-acids that partly originate from repeated sequences through addition of new exons (Figure 4). No functional protein domain was predicted in this zone of the protein. It was followed by a putative docking site, and then a better conserved C-terminal sequence including a well conserved Gla domain (Figure 4). In the elephant shark, a furin site is conserved in the N-terminal side of the Gla domain. Provided the cleavage site is indeed functional, the mature *Bgp* protein would then be very similar to *Bgp* in bony vertebrates: cleavage facilitates carboxylation and as a consequence the affinity of *Bgp* for hydroxyapatite (Al Rifai et al., 2017).

We also questioned the function of chondrichthyans *Mgp1*, *Mgp2*, and *Bgp* through the survey of their expression patterns in the small-spotted catshark. Some observed expression patterns are shared with bony vertebrates (Table 1), but some appear to be specific to cartilaginous fishes. Of course, *in situ* hybridization data show which cells express which genes, but do not help in determining if the protein is produced, where and how much of it is secreted, nor what is the gamma-carboxylation status of the *Mgp/Bgp* proteins. Further proteomic studies would be necessary to resolve the protein status in terms of post-translational cleavage and carboxylation. As a result, the following discussion only speculates on the functional implications of gene expression patterns in the small-spotted catshark.

A specific site of expression in the small-spotted catshark was found with a very restricted site of expression of *Mgp2* in the developing ceratotrichiae of the pectoral fin, also observed for *Bgp*. These shark ceratotrichiae are massive collagenous fibers that support the distal fin (Kemp, 1977) without being mineralized. Similar collagen-based fibrils named actinotrichia are found in teleost fishes (Durán et al., 2011), and together with lepidotrichiae (bony hemi-segments), they build up the typical

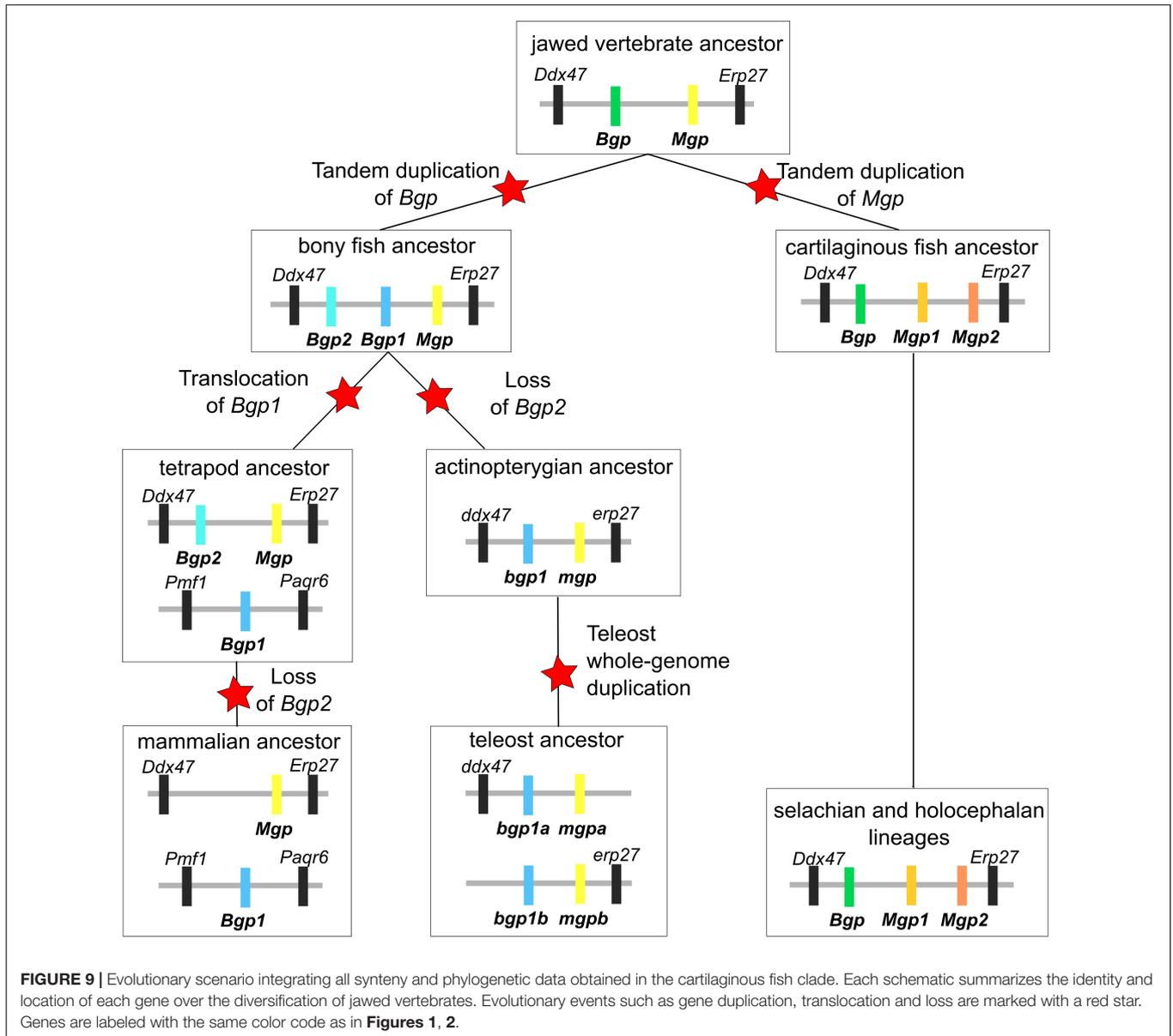
fin rays found in actinopterygians fishes. These collagen-based fibers are supposed to be homologous between cartilaginous and actinopterygian fishes (Zhang et al., 2010). To our knowledge, the expression of *Mgp* or *Bgp1* has never been recorded in fish actinotrichia, although the expression of *Bgp1* was detected in the dermal bone of fin rays in several teleost fishes (Stavri and Zarnescu, 2013; Viegas et al., 2013) and in the cartilaginous supports of fins (Gavaia et al., 2006). These data therefore suggest a cartilaginous fish specific site of expression for *Mgp2* and *Bgp* in developing ceratotrichiae, be it an evolutionary innovation in this lineage, or a consequence of secondary loss of this site of expression in bony fish. The fact that this is a shared zone of expression between *Mgp2* and *Bgp* could support the hypothesis of an ancestral feature (that evolved before the duplication of the gene ancestral to *Bgp* and *Mgp*) or a secondary (chondrichthyan-specific) recruitment of both genes that may share regulatory elements, given their genomic proximity. This strong expression in fin ceratotrichiae, that are not mineralized structures, is not congruent with a mineralization function of *Bgp* proteins in the small-spotted catshark ceratotrichiae.

The remaining range of tissue with expression of the *Bgp* gene in the small-spotted catshark is not congruent either with the hypothesis of a function in the activation of mineralization: it is expressed in several soft tissues such as connective tissues surrounding muscles, nerve root and vasculature of the gills (Figure 7). These sites of expression were previously identified in tetrapods and actinopterygian fishes for both *Mgp* and *Bgp* genes (see Table 1 and discussion below). In most of these soft tissues, the expression of *Mgp* is considered to ensure inhibition of mineralization, but the function of *Bgp* in these tissues is still poorly understood. In the small-spotted catshark, the only site of *Bgp* expression that correlates with tissue mineralization is in the pulp of mineralized teeth, which is similar with other observations in tetrapods and teleost fishes (see Table 1 for references) but may be linked to non-mineralizing cells in the dental pulp, e.g., vascular system or innervation.

Finally, only the expression of *Mgp1* is strongly linked to the dynamic of skeletal mineralization in the small-spotted catshark: it is found expressed in subpopulations of chondrocytes that are specifically pre-mineralization chondrocytes (before areolar mineralization surrounding the notochord; before the initiation of tesseræ mineralization; before globular mineralization in the neural arch (Debiais-Thibaud, 2019); and the expression goes down at the time when mineralization initiates. *Mgp1* is also expressed in the cells of the notochord that never mineralizes. These observations are more congruent with a function of *Mgp1* in the inhibition of mineralization during the maturation of the skeletal tissues in the small-spotted catshark.

## Comparative Analyses and the Evolution of Mgp and Bgp Functions

There is currently no possibility to compare the two *Bgp* copies in tetrapods because expression data have been described only for the *Bgp1* copy in the chicken and the tropical clawed frog, and we did not find any description of *Bgp2* expression (see references in Table 1).



**FIGURE 9 |** Evolutionary scenario integrating all synteny and phylogenetic data obtained in the cartilaginous fish clade. Each schematic summarizes the identity and location of each gene over the diversification of jawed vertebrates. Evolutionary events such as gene duplication, translocation and loss are marked with a red star. Genes are labeled with the same color code as in **Figures 1, 2**.

The early and strong embryonic expression detected for *Bgp* in the small-spotted catshark is reminiscent of other studies showing an early expression of *Bgp1* in the zebrafish (Bensimon-Brito et al., 2012), although others detected neither embryonic nor early larval expression of *Bgp1* in other teleost fish (Pinto et al., 2001). On the other hand, *Mgp* genes are also expressed in early embryos: in the vascular system of the avian embryo (Correia et al., 2016) and developing limbs and lungs of the mouse as early as E10.5 (Luo et al., 1995; Gilbert and Rannels, 2004). All these data suggest a shared and ancestral function of *Mgp* and *Bgp* proteins during embryogenesis, before tissue and cell differentiation. A function in inhibitory interaction with Bmp proteins was shown for the *Mgp* protein in human cells (Zebboudj et al., 2002) as well as with the transforming growth factor- $\beta$  pathway (Oh

et al., 2000) which may explain an early expression during morphogenesis. Another conserved aspect of *Mgp* and *Bgp* genes is their expression in the tissues surrounding certain muscles and the vasculature along the embryonic and adult period. This zone of expression is shared between *Mgp1* and *Bgp* in the small-spotted catshark, similar to previous descriptions in the zebrafish and mammals (Hao et al., 2004; Simes et al., 2004; Viegas et al., 2013). In these sites of expression, it is accepted that *Mgp* and *Bgp* proteins act as mineralization inhibitors, by interacting with the BMP pathway (Yao et al., 2006) or by their properties in their uncarboxylated forms (Schurgers et al., 2005, 2007; Zoch et al., 2016). These two properties might be ancestral characteristics for both *Mgp* and *Bgp* in jawed vertebrates, and of the ancestral gene that gave rise to *Mgp* and *Bgp* by duplication.

**TABLE 1** | Described expression of *Bgp* and *Mgp* genes in selected tissues and selected species of jawed vertebrates compared to data obtained for the small-spotted catshark (this study).

	Small-spotted catshark			Teleost fishes		Xenopus		Mammals		Chicken	
	Mgp1	Mgp2	Bgp	mgp	bgp1a	Mgp	Bgp1	Mgp	Bgp1	Mgp	Bgp1
Embryonic	+	+	+		+ (3)			+ (17)		+ (8)	
Chord	+	–	–		+ (3)					+ (8)	
Chondrocyte	+	–	–	+ (1)	–	+ (10)	– (10)	+ (11)	– (14)	+ (18)	– (12)
Chondrocyte (mineralized matrix)	–	–	–	+ (1)	+ (1)	+ (10)	– (10)	+ (11)	– (14)	+ (18)	+ (12)
Osteoblasts	na	na	na	– (1)	+ (1, 5)	– (10)	+ (10)	+ (6)	+ (14)	– (18)	+ (12)
Early tooth/scale bud	–	+	+		+ (15) regeneration				– (13)	na	na
Late (mineralized) tooth/scale	–	–	+	+ (4)	+ (4)				+ (13)	na	na
Muscle and its connective tissue	–	–	+	+ heart (2, 5)	+ heart (5)	– (7)				+ (19)	
Vasculature and its connective tissue	+	–	+ gills	+ heart (2, 5)	+ arteries(5)			+ (11, 16)	+ (16)	+ (8, 18)	
Nerve root	–	–	+						+ (9)		
Ceratotrachia	–	+	+	–	– (1)	na	na	na	na	na	na

Sources: (1) (Gavaia et al., 2006); (2) (Simes et al., 2003); (3) (Bensimon-Brito et al., 2012); (4) (Ortiz-Delgado et al., 2005); (5) (Viegas et al., 2013); (6) (Coen et al., 2009); (7) (Cancela et al., 2001); (8) (Correia et al., 2016); (9) (Ichikawa and Sugimoto, 2002); (10) (Espinoza et al., 2010); (11) (Luo et al., 1997); (12) (Neugebauer et al., 1995); (13) (Bleicher et al., 1999); (14) (Sommer et al., 1996); (15) (Iimura et al., 2012); (16) (Hao et al., 2004); (17) (Gilbert and Rannels, 2004); (18) (Dan et al., 2012); (19) (Wiedemann et al., 1998). na, not applicable (the anatomical structure does not exist in the specified taxon).

Finally, the expression of the small-spotted catshark *Bgp* in the nerve root is also a characteristic previously described in the mouse (Ichikawa and Sugimoto, 2002) and therefore suggests an ancestral role of the *Bgp* copy in the nervous system of jawed vertebrates. The function of *Bgp* in the nervous system has not been fully uncovered but it has been proposed to be an active neuropeptide in sensory ganglia (Patterson-Buckendahl et al., 2012).

We previously concluded on the putative function of *Mgp1* in the inhibition of mineralization during the maturation of the skeletal tissues in the small-spotted catshark. This observation is shared with all described gene expression patterns in skeletal tissues in other jawed vertebrates. As a consequence, it supports the hypothesis of an ancestral involvement of the *Mgp/Bgp* gene family in the regulation of skeletal mineralization, although limited to the negative regulation of calcium deposition in the cartilage by members of the *Mgp* clade.

## CONCLUDING REMARKS

The description of the *Mgp/Bgp* complement in cartilaginous fishes reveals complex dynamic evolution of this gene family during jawed vertebrate evolution. Although previously reported expression of *Mgp* and *Bgp1* in tetrapods was found involved in the regulation of mineralization in skeletal tissues, only *Mgp1* displays association with skeletal tissue differentiation in the small-spotted catshark embryo, and its expression pattern is congruent with an ability to inhibit mineralization in the step preceding precipitation of calcium in the cartilaginous matrix. The ability to activate mineralization in skeletal tissues may finally be a specificity

of the *Bgp1* bony fish copy: either because it evolved after the divergence with cartilaginous fishes or because cartilaginous fishes have secondarily lost bone-associated genetic toolkits as they lost bone tissues (Donoghue et al., 2006; Brazeau et al., 2020).

## DATA AVAILABILITY STATEMENT

The original contributions presented in this study are included in the article and **Supplementary Material**, further inquiries can be directed to the corresponding authors.

## ETHICS STATEMENT

Handling of small-spotted catshark embryos followed all institutional, national, and international guidelines (European Communities Council Directive of 22 September 2010 [2010/63/UE]): no further approval by an ethics committee was necessary as the biological material is embryonic and no live experimental procedures were carried out.

## AUTHOR CONTRIBUTIONS

MD-T and TH designed the study, analyzed the data, and drafted the manuscript. NL performed the synteny and phylogenetic analyses. CM-M generated qPCR expression data. SV performed the *in situ* expression experiments. TH performed the genome data mining. All authors contributed to the article and approved the submitted version.

## FUNDING

MD-T, CM-M, and NL were supported by the Centre Méditerranéen de l'Environnement et de la Biodiversité (ANR-10-LABX-0004 Skel'Estro project) and TH was supported by the Swedish Research Council (Starting grant 621-2012-4673).

## ACKNOWLEDGMENTS

We thank the national infrastructure France-BioImaging and RIO Imaging platform (Montpellier, France, ANR-10-INBS-04, ANR-10-LABX-0020) for nanozoomer acquisitions. We also thank the MBB and GenSeq platforms for help with the transcriptome analyses and the RHEM at the Institute of Cancer Research of Montpellier (IRCM—U1194) for histology data acquisition. We acknowledge Frédéric Delsuc and Céline Scornavacca for our fruitful discussions on phylogenetic analyses.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.620659/full#supplementary-material>

## REFERENCES

- Al Rifai, O., Chow, J., Lacombe, J., Julien, C., Faubert, D., Susan-Resiga, D., et al. (2017). Proprotein convertase furin regulates osteocalcin and bone endocrine function. *J. Clin. Invest.* 127, 4104–4117. doi: 10.1172/JCI93437
- Amores, A., Force, A., Yan, Y. L., Joly, L., Amemiya, C., Fritz, A., et al. (1998). Zebrafish hox clusters and vertebrate genome evolution. *Science* 282, 1711–1714. doi: 10.1126/science.282.5394.1711
- Ballard, W. W., Mellinger, J., and Lechenault, H. (1993). A series of stages for development of *Scyliorhinus canicula* the lesser spotted dogfish (Chondrichthyes: scyliorhinidae). *J. Exp. Zool.* 267, 1–43. doi: 10.1002/jez.1402670309
- Bensimon-Brito, A., Cardeira, J., Cancela, M. L., Huisseune, A., and Witten, P. E. (2012). Distinct patterns of notochord mineralization in zebrafish coincide with the localization of Osteocalcin isoform 1 during early vertebral centra formation. *BMC Dev. Biol.* 12:28. doi: 10.1186/1471-213X-12-28
- Bleicher, F., Couble, M. L., Farges, J. C., Couble, P., and Magloire, H. (1999). Sequential expression of matrix protein genes in developing rat teeth. *Matrix Biol.* 18, 133–143. doi: 10.1016/S0945-053X(99)00007-4
- Blomme, T., Vandepoele, K., De Bodt, S., Simillion, C., Maere, S., and Van de Peer, Y. (2006). The gain and loss of genes during 600 million years of vertebrate evolution. *Genome Biol.* 7:R43. doi: 10.1186/gb-2006-7-5-r43
- Bordoloi, J., Dihingia, A., Kalita, J., and Manna, P. (2018). Implication of a novel vitamin K dependent protein, GRP/Ucma in the pathophysiological conditions associated with vascular and soft tissue calcification, osteoarthritis, inflammation, and carcinoma. *Int. J. Biol. Macromol.* 113, 309–316. doi: 10.1016/j.ijbiomac.2018.02.150
- Brazeau, M., Giles, S., Dearden, R., Jerve, A., Ariunchimeg, Y. A., Sansom, R., et al. (2020). Endochondral bone in an Early Devonian 'placoderm' from Mongolia. *Nat. Ecol. Evol.* 4, 1477–1484. doi: 10.1038/s41559-020-01290-2
- Cancela, M. L., Laizé, V., and Conceição, N. (2014). Matrix Gla protein and osteocalcin: from gene duplication to neofunctionalization. *Arch. Biochem. Biophys.* 561, 56–63. doi: 10.1016/j.abb.2014.07.020
- Cancela, M. L., Ohresser, M. C. P., Reia, J. P., Viegas, C. S. B., Williamson, M. K., and Price, P. A. (2001). Matrix gla protein in *Xenopus laevis*: molecular cloning, tissue distribution, and evolutionary considerations. *J. Bone Miner. Res.* 16, 1611–1621. doi: 10.1359/jbmr.2001.16.9.1611
- Cavaco, S., Williamson, M. K., Rosa, J., Roberto, V., Cordeiro, O., Price, P. A., et al. (2014). Teleost fish osteocalcin 1 and 2 share the ability to bind the calcium mineral phase. *Fish Physiol. Biochem.* 40, 731–738. doi: 10.1007/s10695-013-9880-9
- Coen, G., Ballanti, P., Silvestrini, G., Mantella, D., Manni, M., Di Giulio, S., et al. (2009). Immunohistochemical localization and mRNA expression of matrix Gla protein and fetuin-A in bone biopsies of hemodialysis patients. *Virchows Arch.* 454, 263–271. doi: 10.1007/s00428-008-0724-4
- Comte, N., Morel, B., Hasic, D., Guéguen, L., Boussau, B., Daubin, V., et al. (2020). TreeRecs: an integrated phylogenetic tool, from sequences to reconciliations. *Bioinformatics* 36, 4822–4824. doi: 10.1093/bioinformatics/btaa615
- Correia, E., Conceição, N., Cancela, M. L., and Belo, J. A. (2016). Matrix Gla protein expression pattern in the early avian embryo. *Int. J. Dev. Biol.* 60, 71–76. doi: 10.1387/ijdb.150365jb
- Dan, H., Simsa-Maziel, S., Reich, A., Sela-Donenfeld, D., and Monsonego-Ornan, E. (2012). The role of matrix Gla protein in ossification and recovery of the avian growth plate. *Front. Endocrinol. (Lausanne)* 3:79. doi: 10.3389/fendo.2012.00079
- Debiais-Thibaud, M., Simion, P., Ventéo, S., Muñoz, D., Marcellini, S., Mazan, S., et al. (2019). Skeletal mineralization in association with Type X collagen expression is an ancestral feature for jawed vertebrates. *Mol. Biol. Evol.* 36, 2265–2276. doi: 10.1093/molbev/msz145
- Debiais-Thibaud, M. (2019). "The evolution of endoskeletal mineralisation in chondrichthyan fish," in *Evolution and Development of Fishes*, eds Z. Underwood, C. Richter, and M. Johanson (Cambridge: Cambridge University Press).
- Dehal, P., and Boore, J. L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* 3:e314. doi: 10.1371/journal.pbio.0030314
- D'Errico, J. A., MacNeil, R. L., Takata, T., Berry, J., Strayhorn, C., and Somerman, M. J. (1997). Expression of bone associated markers by tooth root lining cells, in situ and in vitro. *Bone* 20, 117–126. doi: 10.1016/s8756-3282(96)00348-1
- Di Franco, A., Poujol, R., Baurain, D., and Philippe, H. (2019). Evaluating the usefulness of alignment filtering methods to reduce the impact of errors

**Supplementary Material 1** | References of all sequences used in this manuscript and extracted from open sequence databases or from locally assembled transcriptome data.

**Supplementary Material 2** | Protein sequence alignment generated by MAFFT.

**Supplementary Material 3** | Protein sequence alignment generated by Hmmer automatic cleaning and used to produce the phylogenetic tree reconstruction.

**Supplementary Material 4** | Contracted gene and species trees used with TreeRecs and the visual output for the maximum parsimony scenario.

**Supplementary Material 5** | Unrooted tree obtained after the phylogenetic reconstruction. Colors and gene clade annotations are as in **Figure 1**.

**Supplementary Material 6** | Alignment identifying conserved protein regions of Bgp proteins between the human *Homo sapiens*, the chicken *Gallus gallus*, the zebrafish *Danio rerio*, the small spotted catshark *Scyliorhinus canicula*, and the elephant shark *Callorhynchus milii*: signal peptide in the first 22 amino-acids, furin cleavage site in positions 162–166 of the alignment, core Gla domain starting on position 186.

**Supplementary Material 7** | Alignment identifying conserved protein regions of Mgp proteins between the human *Homo sapiens*, the zebrafish *Danio rerio*, and both Mgp1 and Mgp2 sequences found in chondrichthyan. Signal peptide is found in the first 20 amino-acids, conserved phosphorylation domain in the Mgp1 sequences in positions 28–36, a partially conserved ANxF domain located in position 48–51 with a F in position 42 (considered as a carboxylase docking site), core Gla domain starting on position 83 with poor conservation of the Mgp2 sequence after position 90. Note the furin cleavage site in positions 53–56 of the alignment, found in *Callorhynchus milii* Mgp2 sequence only.

- on evolutionary inferences. *BMC Evol. Biol.* 19:21. doi: 10.1186/s12862-019-1350-2
- Diegel, C. R., Hann, S., Ayturk, U. M., Hu, J. C. W., Lim, K., Droscha, C. J., et al. (2020). An osteocalcin-deficient mouse strain without endocrine abnormalities. *PLoS Genet.* 16:e1008361. doi: 10.1371/journal.pgen.1008361
- Donoghue, P. C. J., and Sansom, I. J. (2002). Origin and early evolution of vertebrate skeletonization. *Microsc. Res. Tech.* 59, 352–372. doi: 10.1002/jemt.10217
- Donoghue, P. C. J., Sansom, I. J., and Downs, J. P. (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. doi: 10.1002/jez.b.21090
- Durán, I., Mari-Beffa, M., Santamaría, J. A., Becerra, J., and Santos-Ruiz, L. (2011). Actinotrichia collagens and their role in fin formation. *Dev. Biol.* 354, 160–172. doi: 10.1016/j.ydbio.2011.03.014
- Enault, S., Adnet, S., and Debiais-Thibaud, M. (2016). Skeletogenesis during the late embryonic development of the catshark *Scyliorhinus canicula* (Chondrichthyes; Neoselachii). *MorphoMuseum* 1:e2. doi: 10.18563/m3.1.4.e2
- Enault, S., Muñoz, D. N., Silva, W. T. A. F. A. F., Borday-Birraux, V., Bonade, M., Oulion, S., et al. (2015). Molecular footprinting of skeletal tissues in the catshark *Scyliorhinus canicula* and the clawed frog *Xenopus tropicalis* identifies conserved and derived features of vertebrate calcification. *Front. Genet.* 6:283. doi: 10.3389/fgene.2015.00283
- Espinoza, J., Sanchez, M., Sanchez, A., Hanna, P., Torrejon, M., Buisine, N., et al. (2010). Two families of *Xenopus tropicalis* skeletal genes display well-conserved expression patterns with mammals in spite of their highly divergent regulatory regions. *Evol. Dev.* 12, 541–551. doi: 10.1111/j.1525-142X.2010.00440.x
- Finn, R. D., Attwood, T. K., Babbitt, P. C., Bateman, A., Bork, P., Bridge, A. J., et al. (2017). InterPro in 2017—beyond protein family and domain annotations. *Nucleic Acids Res.* 45, D190–D199. doi: 10.1093/nar/gkx1107
- Fraser, J. D., and Price, P. A. (1988). Lung, heart and kidney express high levels of mRNA for the vitamin K-dependent matrix Gla protein. *J. Biol. Chem.* 263, 11033–11036. doi: 10.1016/s0021-9258(18)37912-2
- Gavaia, P. J., Simes, D. C., Ortiz-Delgado, J. B., Viegas, C. S. B., Pinto, J. P., Kelsh, R. N., et al. (2006). Osteocalcin and matrix Gla protein in zebrafish (*Danio rerio*) and Senegal sole (*Solea senegalensis*): comparative gene and protein expression during larval development through adulthood. *Gene Expr. Patterns* 6, 637–652. doi: 10.1016/j.modgep.2005.11.010
- Gilbert, K. A., and Rannels, S. R. (2004). Matrix GLA protein modulates branching morphogenesis in fetal rat lung. *Am. J. Physiol. Lung Cell. Mol. Physiol.* 286, 1179–1187. doi: 10.1152/ajplung.00188.2003
- Grant, C. E., Bailey, T. L., and Noble, W. S. (2011). FIMO: scanning for occurrences of a given motif. *Bioinformatics* 27, 1017–1018. doi: 10.1093/bioinformatics/btr064
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., and Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol.* 59, 307–321. doi: 10.1093/sysbio/syq010
- Hao, H., Hirota, S., Ishibashi-Ueda, H., Kushiro, T., Kanmatsuse, K., and Yutani, C. (2004). Expression of matrix Gla protein and osteonectin mRNA by human aortic smooth muscle cells. *Cardiovasc. Pathol.* 13, 195–202. doi: 10.1016/j.carpath.2004.03.607
- Hoang, D. T., Chernomor, O., Haeseler, A., Minh, B. Q., and Vinh, L. S. (2017). UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* 35, 518–522. doi: 10.5281/zenodo.854445
- Ichikawa, H., and Sugimoto, T. (2002). The difference of osteocalcin-immunoreactive neurons in the rat dorsal root and trigeminal ganglia: co-expression with nociceptive transducers and central projection. *Brain Res.* 958, 459–462.
- Iimura, K., Tohse, H., Ura, K., and Takagi, Y. (2012). Expression patterns of runx2, sparc, and bgp during scale regeneration in the goldfish *Carassius auratus*. *J. Exp. Zool. B. Mol. Dev. Evol.* 318, 190–198. doi: 10.1002/jez.b.22005
- Ikeda, T., Nomura, S., Yamaguchi, A., Suda, T., and Yoshiki, S. (1992). In situ hybridization of bone matrix proteins in undecalcified adult rat bone sections. *J. Histochem. Cytochem.* 40, 1079–1088. doi: 10.1177/40.8.1619274
- Janvier, P. (1996). *Early Vertebrates*. New York: Oxford University Press Inc.
- Kaipatur, N. R., Murshed, M., and McKee, M. D. (2008). Matrix Gla protein inhibition of tooth mineralization. *J. Dent. Res.* 87, 839–844. doi: 10.1177/154405910808700907
- Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acid Res.* 30, 3059–3066.
- Katoh, K., and Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* 30, 772–780. doi: 10.1093/molbev/mst010
- Kemp, N. E. (1977). Banding pattern and fibrillogenesis of ceratotrichia in shark fins. *J. Morphol.* 154, 187–204. doi: 10.1002/jmor.1051540202
- Laizé, V., Martel, P., Viegas, C. S. B., Price, P. A., and Cancela, M. L. (2005). Evolution of matrix and bone  $\gamma$ -carboxyglutamic acid proteins in vertebrates. *J. Biol. Chem.* 280, 26659–26668. doi: 10.1074/jbc.M500257200
- Letunic, I., Khedkar, S., and Bork, P. (2021). SMART: recent updates, new developments and status in 2020. *Nucleic Acids Res.* 49, D458–D460. doi: 10.1093/nar/gkaa937
- Luo, G., D'Souza, R., Hogue, D., and Karsenty, G. (1995). The matrix Gla protein gene is a marker of the chondrogenesis cell lineage during mouse development. *J. Bone Miner. Res.* 10, 325–334. doi: 10.1002/jbmr.5650100221
- Luo, G., Ducy, P., McKee, M. D., Pinero, G. J., Loyer, E., Behringer, R. R., et al. (1997). Spontaneous calcification of arteries and cartilage in mice lacking matrix GLA protein. *Nature* 386, 78–81. doi: 10.1038/386078a0
- Neugebauer, B. M., Moore, M. A., and Broess, M. (1995). Characterization of structural sequences in the chicken osteocalcin gene: expression of osteocalcin by maturing osteoblasts and by hypertrophic chondrocytes in vitro. *J. Bone Miner. Res.* 10, 157–163.
- Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol. Biol. Evol.* 32, 268–274. doi: 10.1093/molbev/msu300
- Oh, S. P., Seki, T., Goss, K. A., Imamura, T., Yi, Y., Donahoe, P. K., et al. (2000). Activin receptor-like kinase 1 modulates transforming growth factor- $\beta$ 1 signaling in the regulation of angiogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 97, 2626–2631. doi: 10.1073/pnas.97.6.2626
- Ohno, S. (1970). *Evolution by Gene Duplication*. Berlin: Springer-Verlag.
- Omelson, S., Georgiou, J., Henneman, Z. J., Wise, L. M., Sukhu, B., Hunt, T., et al. (2009). Control of vertebrate skeletal mineralization by polyphosphates. *PLoS One* 4:e5634. doi: 10.1371/journal.pone.0005634
- Onimaru, K., Marcon, L., Musy, M., Tanaka, M., and Sharpe, J. (2016). The fin-to-limb transition as the re-organization of a Turing pattern. *Nat. Commun.* 7:11582. doi: 10.1038/ncomms11582
- Ortiz-Delgado, J. B., Simes, D. C., Gavaia, P. J., Sarasquete, C., and Cancela, M. L. (2005). Osteocalcin and matrix GLA protein in developing teleost teeth: identification of sites of mRNA and protein accumulation at single cell resolution. *Histochem. Cell Biol.* 124, 123–130. doi: 10.1007/s00418-005-0015-y
- Ortiz-Delgado, J. B., Simes, D. C., Viegas, C. S. B., Schaff, B. J., Sarasquete, C., and Cancela, M. L. (2006). Cloning of matrix Gla protein in a marine cartilaginous fish, *Prionace glauca*: preferential protein accumulation in skeletal and vascular systems. *Histochem. Cell Biol.* 126, 89–101. doi: 10.1007/s00418-005-0125-6
- O'Shaughnessy, K. L., Dahn, R. D., and Cohn, M. J. (2015). Molecular development of chondrichthyan claspers and the evolution of copulatory organs. *Nat. Commun.* 6:6698. doi: 10.1038/ncomms7698
- Ota, K. G., Fujimoto, S., Oisi, Y., and Kuratani, S. (2013). Late development of hagfish vertebral elements. *J. Exp. Zool. B. Mol. Dev. Evol.* 320, 129–139. doi: 10.1002/jez.b.22489
- Patterson-Buckendahl, P., Sowinska, A., Yee, S., Patel, D., Pagkalinawan, S., Shahid, M., et al. (2012). Decreased sensory responses in osteocalcin null mutant mice imply neuropeptide function. *Cell. Mol. Neurobiol.* 32, 879–889. doi: 10.1007/s10571-012-9810-x
- Pinto, J. P., Ohresser, M. C. P., and Cancela, M. L. (2001). Cloning of the bone Gla protein gene from the teleost fish *Sparus aurata*. Evidence for overall conservation in gene organization and bone-specific expression from fish to man. *Gene* 270, 77–91. doi: 10.1016/s0378-1119(01)00426-7
- Price, P. A., Rice, J. S., and Williamson, M. K. (1994). Conserved phosphorylation of serines in the Ser-X-Glu / Ser (P) sequences of the vitamin K- dependent matrix Gla protein from shark, lamb, rat, cow, and human. *Protein Sci.* 3, 822–830. doi: 10.1002/pro.5560030511

- Rhie, A., McCarthy, S. A., Fedrigo, O., Damas, J., Formenti, G., London, S. E., et al. (2020). Towards complete and error-free genome assemblies of all vertebrate species. *BioRxiv [Preprint]* doi: 10.1101/2020.05.22.110833
- Rice, J. S., Williamson, M. K., and Price, P. A. (1994). Isolation and sequence of the vitamin K-dependent matrix Gla protein from the calcified cartilage of the soupfin shark. *J. Bone Miner. Res.* 9, 567–576. doi: 10.1002/jbmr.5650090417
- Schurgers, L. J., Spronk, H. M. H., Skepper, J. N., Hackeng, T. M., Shanahan, C. M., Vermeer, C., et al. (2007). Post-translational modifications regulate matrix Gla protein function: importance for inhibition of vascular smooth muscle cell calcification. *J. Thromb. Haemost.* 5, 2503–2511.
- Schurgers, L. J., Teunissen, K. J. F., Knapen, M. H. J., Kwajtaal, M., Van Diest, R., Appels, A., et al. (2005). Novel conformation-specific antibodies against matrix  $\gamma$ -carboxyglutamic acid (Gla) protein: undercarboxylated matrix Gla protein as marker for vascular calcification. *Arterioscler. Thromb. Vasc. Biol.* 25, 1629–1633. doi: 10.1161/01.ATV.0000173313.46222.43
- Simes, D. C., Williamson, M. K., Ortiz-Delgado, J. B., Viegas, C. S. B., Price, P. A., and Cancela, M. L. (2003). Purification of matrix Gla protein from a marine teleost fish, *Argyrosomus regius*: calcified cartilage and not bone as the primary site of mgp accumulation in fish. *J. Bone Miner. Res.* 18, 244–259. doi: 10.1359/jbmr.2003.18.2.244
- Simes, D. C., Williamson, M. K., Schaff, B. J., Gavaia, P. J., Ingleton, P. M., Price, P. A., et al. (2004). Characterization of osteocalcin (BGP) and matrix Gla Protein (MGP) fish specific antibodies: validation for immunodetection studies in lower vertebrates. *Calcif. Tissue Int.* 74, 170–180. doi: 10.1007/s00223-003-0079-4
- Sommer, B., Bickel, M., Hofstetter, W., and Wetterwald, A. (1996). Expression of matrix proteins during the development mineralized tissues. *Bone* 19, 371–380. doi: 10.1016/s8756-3282(96)00218-9
- Stavri, S., and Zarnescu, O. (2013). The expression of alkaline phosphatase, osteopontin, osteocalcin, and chondroitin sulfate during pectoral fin regeneration in *Carassius auratus* gibelio: a combined histochemical and immunohistochemical study. *Microsc. Microanal.* 19, 233–242.
- Tan, M., Redmond, A., Dooley, H., Nozu, R., Sato, K., Kuraku, S., et al. (2019). The whale shark genome reveals patterns of vertebrate gene family evolution. *bioRxiv [Preprint]* doi: 10.1101/685743
- Tanaka, M., Münsterberg, A., Anderson, W. G., Prescott, A. R., Hazon, N., and Tickle, C. (2002). Fin development in a cartilaginous fish and the origin of vertebrate limbs. *Nature* 416, 527–531. doi: 10.1038/416527a
- Van Oss, S. B., and Carvunis, A. R. (2019). De novo gene birth. *PLoS Genet.* 15:e1008150. doi: 10.1371/journal.pgen.1008160
- Venkatesh, B., Lee, A. P., Ravi, V., Maurya, A. K., Lian, M. M., Swann, J. B., et al. (2014). Elephant shark genome provides unique insights into gnathostome evolution. *Nature* 505, 174–179. doi: 10.1038/nature12826
- Viegas, C. S. B., Simes, D. C., Williamson, M. K., Cavaco, S., Laizé, V., Price, P. A., et al. (2013). Sturgeon osteocalcin shares structural features with matrix Gla protein: evolutionary relationship and functional implications. *J. Biol. Chem.* 288, 27801–27811. doi: 10.1074/jbc.M113.450213
- Wen, L., Chen, J., Duan, L., and Li, S. (2018). Vitamin K-dependent proteins involved in bone and cardiovascular health (Review). *Mol. Med. Rep.* 18, 3–15. doi: 10.3892/mmr.2018.8940
- Wiedemann, M., Trueb, B., and Belluocci, D. (1998). Molecular cloning of avian matrix Gla protein. *Biochim. Biophys. Acta Gene Struct. Expr.* 1395, 47–49. doi: 10.1016/S0167-4781(97)00155-3
- Yáñez, M., Gil-Longo, J., and Campos-Toimil, M. (2012). Chapter 19: calcium binding proteins. *Adv. Exp. Med. Biol.* 740, 461–482. doi: 10.1007/978-94-007-2888-2
- Yao, T., Ohtani, K., Kuratani, S., and Wada, H. (2011). Development of lamprey mucocartilage and its dorsal-ventral patterning by endothelin signaling, with insight into vertebrate jaw evolution. *J. Exp. Zool. B. Mol. Dev. Evol.* 316, 339–346. doi: 10.1002/jez.b.21406
- Yao, Y., Zebboudj, A. F., Shao, E., Perez, M., and Boström, K. (2006). Regulation of bone morphogenetic protein-4 by matrix GLA protein in vascular endothelial cells involves activin-like kinase receptor 1. *J. Biol. Chem.* 281, 33921–33930. doi: 10.1074/jbc.M604239200
- Zebboudj, A. F., Imura, M., and Boström, K. (2002). Matrix GLA protein, a regulatory protein for bone morphogenetic protein-2. *J. Biol. Chem.* 277, 4388–4394. doi: 10.1074/jbc.M109683200
- Zhang, J., Wagh, P., Guay, D., Sanchez-Pulido, L., Padhi, B. K., Korzh, V., et al. (2010). Loss of fish actinotrichia proteins and the fin-to-limb transition. *Nature* 466, 234–237. doi: 10.1038/nature09137
- Zhang, Y., Zhao, L., Wang, N., Li, J., He, F., Li, X., et al. (2019). Unexpected role of matrix Gla protein in osteoclasts: inhibiting osteoclast differentiation and bone resorption. *Mol. Cell. Biol.* 39, e00012–e00019.
- Zoch, M. L., Clemens, T. L., and Riddle, R. C. (2016). New insights into the biology of osteocalcin. *Bone* 82, 42–49. doi: 10.1016/j.bone.2015.05.046

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Leurs, Martinand-Mari, Ventéo, Haitina and Debiais-Thibaud. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



# Divergent Expression of *SPARC*, *SPARC-L*, and *SCPP* Genes During Jawed Vertebrate Cartilage Mineralization

Adrian Romero<sup>1†</sup>, Nicolas Leurs<sup>2†</sup>, David Muñoz<sup>1</sup>, Mélanie Debiais-Thibaud<sup>2\*</sup> and Sylvain Marcellini<sup>1\*</sup>

<sup>1</sup>Laboratory of Development and Evolution (LADE), University of Concepción, Concepción, Chile, <sup>2</sup>Institut des Sciences de l'Évolution de Montpellier, ISEM, Univ Montpellier, CNRS, IRD, EPHE, Montpellier, France

## OPEN ACCESS

### Edited by:

Claudio Oliveira,  
São Paulo State University, Brazil

### Reviewed by:

Anthony K. Redmond,  
Trinity College Dublin, Ireland  
Andrew Jheon,  
University of California, San Francisco,  
United States

### \*Correspondence:

Mélanie Debiais-Thibaud  
melanie.debiais-thibaud@  
umontpellier.fr  
Sylvain Marcellini  
smarcellini@udec.cl

<sup>†</sup>These authors have contributed  
equally to this work

### Specialty section:

This article was submitted to  
Evolutionary and Population Genetics,  
a section of the journal  
Frontiers in Genetics

**Received:** 02 October 2021

**Accepted:** 10 November 2021

**Published:** 25 November 2021

### Citation:

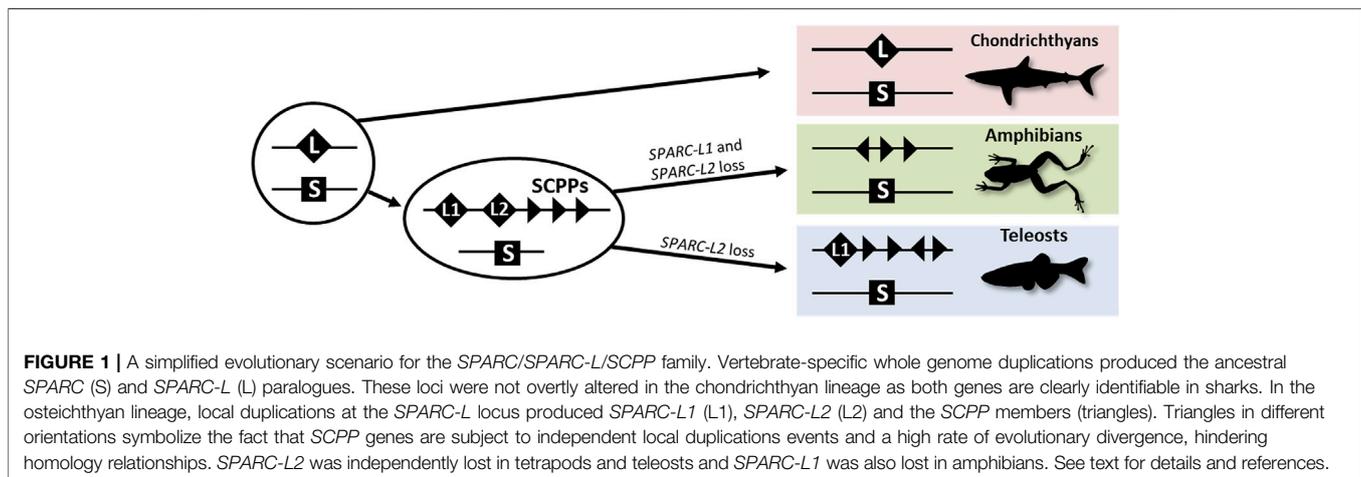
Romero A, Leurs N, Muñoz D,  
Debiais-Thibaud M and Marcellini S  
(2021) Divergent Expression of  
*SPARC*, *SPARC-L*, and *SCPP* Genes  
During Jawed Vertebrate  
Cartilage Mineralization.  
*Front. Genet.* 12:788346.  
doi: 10.3389/fgene.2021.788346

While cartilage is an ancient tissue found both in protostomes and deuterostomes, its mineralization evolved more recently, within the vertebrate lineage. *SPARC*, *SPARC-L*, and the *SCPP* members (Secretory Calcium-binding PhosphoProtein genes which evolved from *SPARC-L*) are major players of dentine and bone mineralization, but their involvement in the emergence of the vertebrate mineralized cartilage remains unclear. We performed *in situ* hybridization on mineralizing cartilaginous skeletal elements of the frog *Xenopus tropicalis* (*Xt*) and the shark *Scyliorhinus canicula* (*Sc*) to examine the expression of *SPARC* (present in both species), *SPARC-L* (present in *Sc* only) and the *SCPP* members (present in *Xt* only). We show that while mineralizing cartilage expresses *SPARC* (but not *SPARC-L*) in *Sc*, it expresses the *SCPP* genes (but not *SPARC*) in *Xt*, and propose two possible evolutionary scenarios to explain these opposite expression patterns. In spite of these genetic divergences, our data draw the attention on an overlooked and evolutionarily conserved peripheral cartilage subdomain expressing *SPARC* or the *SCPP* genes and exhibiting a high propensity to mineralize.

**Keywords:** *SPARC*, *SPARC-L*, *SCPP*, cartilage mineralization, *Xenopus tropicalis*, *Scyliorhinus canicula*, vertebrate evolution

## INTRODUCTION

The evolution of a mineralized skeleton occurred in early vertebrates, in a variety of tissues including superficial dermal scales and teeth, together with internal cartilages, and perichondral bones (Ørving, 1951; Donoghue and Sansom, 2002). In the internal skeleton, several cell types are associated with biomineralization, and the most studied cell model in mammalian organisms is the osteoblast active in the endochondral ossification process (Long and Ornitz, 2013). These osteoblasts are derived from periosteal tissues or from hypertrophic transdifferentiated chondrocytes (Tsang et al., 2015). The process of endochondral ossification, or replacement of cartilage matrix by bone marrow and bone trabeculae, is absent from chondrichthyans and has long been thought to be a derived feature specific to osteichthyans (reviewed in Donoghue and Sansom, 2002; Hirasawa and Kuratani, 2015), although recent paleontological data has challenged this view (Brazeau et al., 2020). Also known to mineralize their matrix are the chondrocytes, not only at the ossification front of endochondral bone growth (in the case of hyaline cartilage), but also in stable forms of mineralized cartilage such as fibrocartilages and other forms of cartilage displaying striking similarities to bony tissues (Beresford, 1981; Dymant



et al., 2015; Paul et al., 2016; Pears et al., 2020; Berio et al., 2021). Even though both perichondral bones and cartilaginous tissues displayed mineralization in the earliest forms of mineralized internal skeletons (Ørvig, 1951; Min and Janvier, 1998; Donoghue et al., 2006; Johanson et al., 2010, 2012; Pears et al., 2020), mineralizing cartilages have been understudied from a genetic and evolutionary perspective in extant vertebrates. A better understanding of the genetic underpinning of the mineralizing chondrocytes is therefore necessary to understand the early steps of the evolution of endoskeletal mineralization in vertebrates.

The evolution of vertebrate endoskeletal mineralization has been discussed in the light of the two rounds of whole-genome duplication (2Rs). These duplications occurred before the diversification of extant jawed vertebrates (Nakatani et al., 2021) and generated gene families with diverging gene functions which may have produced the genetic toolkit required for the cellular ability to mineralize an extracellular matrix (Zhang and Cohn, 2008). The evolution of the *SPARC/SPARC-L/SCPP* gene family has been of great interest in this perspective (Kawasaki and Weiss, 2003; Kawasaki et al., 2005; Kawasaki, 2009; Bertrand et al., 2013; Enault et al., 2018), and is summarized in **Figure 1**. *SPARC-L* and *SPARC* are two paralogues having originated from the 2Rs (Kawasaki and Weiss, 2003; Kawasaki et al., 2005; Kawasaki, 2009; Bertrand et al., 2013; Enault et al., 2018). In bony fishes, independent local duplications at the *SPARC-L* locus generated *SPARC-L1* and *SPARC-L2* and a variable number of tandemly located genes coding for Secretory Calcium-binding PhosphoProteins (SCPPs) that have evolved rapidly since their origin (see **Supplementary Figure S1** and Kawasaki and Weiss, 2003; Kawasaki et al., 2005; Enault et al., 2018). Hence outside of amniotes, homology relationships between *SCPP* duplicates are obscured by independent gene gains and losses together with a high rate of sequence divergence (Kawasaki, 2009). No *SCPP* genes have been identified in cartilaginous fish genomes, making the chondrichthyan *SPARC-L* gene the single orthologue to all *SCPP* genes of bony vertebrates (see

**Figure 1**, **Supplementary Figure S1** and Ryll et al., 2014; Venkatesh et al., 2014; Enault et al., 2018).

The *SPARC* gene (Secreted Protein Acidic and Rich in Cysteine, formerly coined *Osteonectin*) encodes a matricellular protein which is one of the most abundant non-collagenous matrix proteins in mammalian and teleost bone (Schreiweis et al., 2007; Kessels et al., 2014). Secreted by osteoblasts, the *SPARC* protein functions in mineralized tissues by binding both collagen fibrils and calcium, but also by signaling to bone cells (reviewed by Rosset and Bradshaw, 2016). In osteichthyans, the expression of *SPARC* is evolutionary conserved in osteoblasts as well as in odontoblasts (Holland et al., 1987; Li et al., 2009; Espinoza et al., 2010; Enault et al., 2018). In chondrichthyans having secondarily lost the bone tissue (and the osteoblast cell type), *SPARC* is highly expressed in odontoblasts (Enault et al., 2018). The single *SPARC-L* gene in cartilaginous fishes is expressed in enameloid secreting cells in teeth and scales of the catshark *Scyliorhinus canicula* (Enault et al., 2018). In osteichthyans it seems that *SPARC-L1* and *SPARC-L2* are not specifically expressed nor functionally required in the skeleton (McKinnon et al., 2000; Bertrand et al., 2013). In addition, *SPARC-L2* was independently lost in tetrapods and teleosts, and *SPARC-L1* was also lost in amphibians (see **Figure 1** and Kawasaki et al., 2007; Bertrand et al., 2013; Enault et al., 2018), suggesting that these two genes are functionally dispensable. Rather, in osteichthyans, *SCPP* family members are key players of skeletal mineralization. Within amniote *SCPP* genes, *Bone sialoprotein* (*BSP*), *Osteopontin* (*OPN* or *SPP1*) and *Dentin matrix protein 1* (*DMP1*) are strongly expressed by osteoblasts and their protein products are stored in the mineral phase of bone tissue (Ustriyana et al., 2021). Most members of this family are also expressed and functional during tooth development in mammals (either in the production of enamel or/and dentin, reviewed by Nikoloudaki, 2021). In the clawed frog *Xenopus tropicalis* and the zebrafish *Danio rerio* the expression of distinct *SCPP* members has been reported in ameloblasts, odontoblasts, and osteoblasts (Kawasaki et al., 2005; Kawasaki, 2009; Espinoza et al., 2010; Enault et al., 2018). Overall, our knowledge of the evolution of the expression of *SPARC*, *SPARC-L* and the *SCPP* members during

cartilage mineralization remains limited, and, in this study, we examined the expression of these genes during endoskeletal development in *Xenopus tropicalis* and *Scyliorhinus canicula*.

## METHODS

### Specimens, Histological Staining and Cryo-Sections

Lesser spotted catshark (*Scyliorhinus canicula*) embryos were maintained at 17°C at the University of Montpellier, France, until they reached development stage 32 (Ballard et al., 1993; Maxwell et al., 2008). Embryos were taken out of their eggshell, anesthetized and subsequently euthanized by overdose of MS-222 (Sigma) following European animal-care specifications. As substantial growth occurs during stage 32, each individual was measured before fixation and classified into early, intermediate and late stages whose body length measured respectively 5.3, 6.6, and 8.5 cm for histological analyses, and respectively 5.0, 6.3, and 7.9 cm for the Alizarin red S and *in situ* hybridization procedures. Abdominal vertebral portions were fixed 48 h in PFA 4% in PBS 1× at 4°C and were subsequently transferred in ethanol and stored at −20°C until needed.

Adult *Xenopus tropicalis* were maintained following standard protocols established for this species, at the University of Concepcion. Embryos and tadpoles were raised after natural mating and staged according to the Nieuwkoop and Faber developmental table (Nieuwkoop and Faber, 1967). Anesthesia of tadpoles was performed with a MS-222 (Sigma) solution at 2 mg/ml and each specimen was subsequently decapitated in agreement with international bioethical recommendations (Close et al., 1996; Ramlochan Singh et al., 2014).

Dissected organs of both species were embedded in paraffin to generate 7 µm-thick histological sections that were stained with standard protocols [eosin, hematoxylin and safran reaction for catshark (RHEM platform at IRCM, Montpellier); hematoxylin and chromotrope 2R (C3143 Sigma) for frog sections]. The von Kossa protocol was used on paraffin sections of *Xenopus tropicalis* to detect calcium on tissue sections (#10241, Diapath, Italy) following the manufacturer's instruction. Briefly, the von Kossa method is based on the transformation of calcium ions, bound to phosphates, into silver ions brought by a solution of silver nitrate.

Spotted catshark alizarin red S and *in situ* hybridizations were performed on serial, 14 µm thick cryostat sections, cut transversal in the body trunk, at the level of the pectoral fins. Parts of the specimens that were not used for this study were conserved in ethanol at −20°C for further studies on gene expression. Alizarin red S staining was used to detect calcium deposits with a single bath of 0.05% Alizarin Red S (Sigma) in KOH 0.5%, 5 min, before mounting in Mowiol. All slides generated with catshark samples were scanned on a Hamamatsu nanozoomer.

### *In situ* Hybridizations

All probes and *in situ* hybridization procedures used here with *Scyliorhinus canicula* and *Xenopus tropicalis* were previously described (Espinoza et al., 2010; Enault et al., 2018), except for the frog *SCPPA2* gene (GenBank EU642617) for which a 968 bp product was amplified and cloned into pBluescript using the

following primers (5' to 3') Forward- GAG TCA TAC TAC AGG ACC TGC, Reverse-CAT GCA ACT CAG CCA AAG CT.

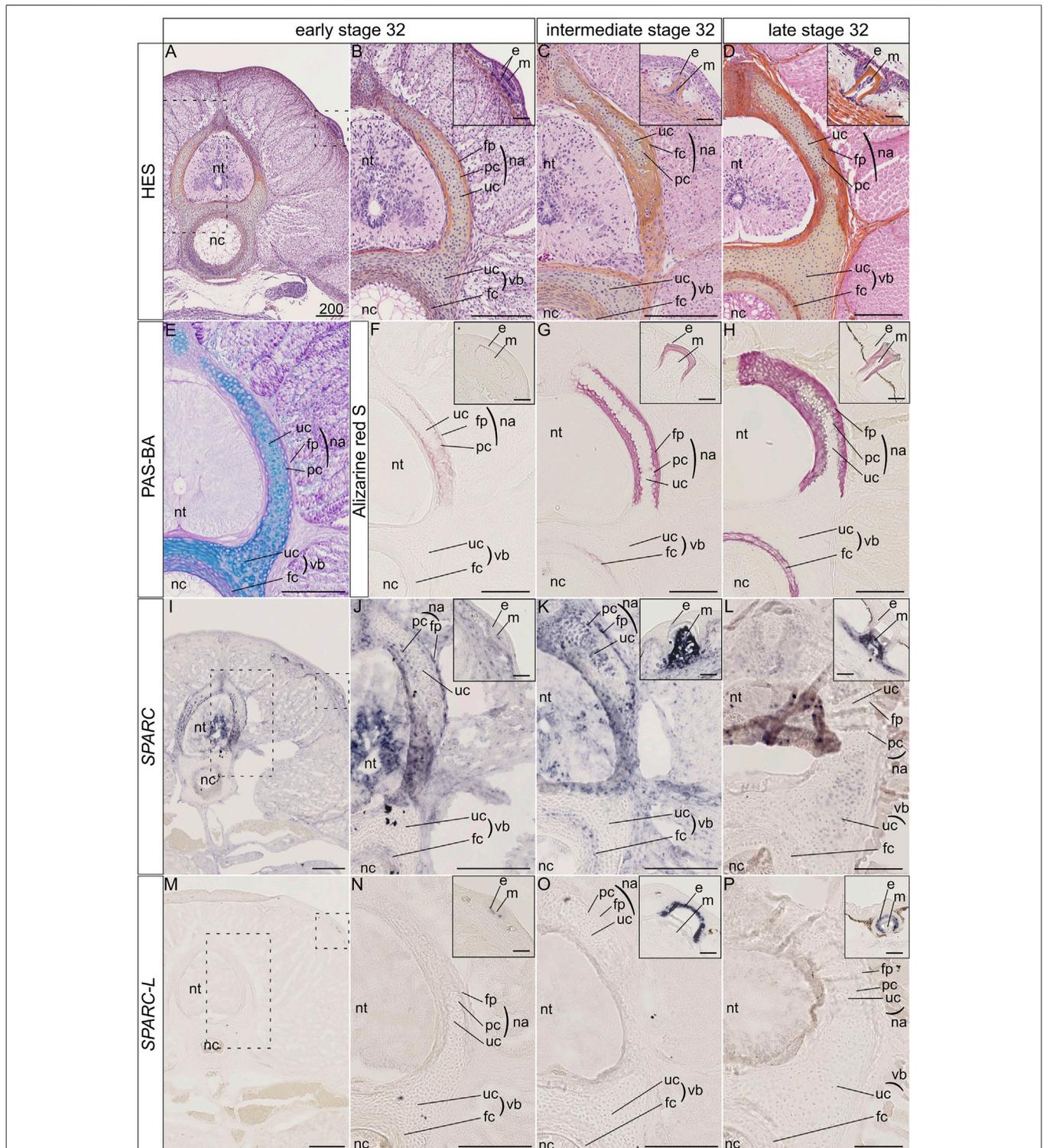
## RESULTS

### *SPARC* and *SPARC-L* Expression in the Development of Vertebrae in the Lesser Spotted Catshark *Scyliorhinus canicula*

The catshark vertebral tissue mineralizes at the level of the neural arches and of the vertebral body (Enault et al., 2015). In the neural arches, mineralization occurs at two juxtaposed sites: the matrix of the most peripheral chondrocytes and the fibrous perichondrial tissue surrounding each neural arch (Figures 2A–H). In the neural arch peripheral cartilage, a faint Alizarin red-positive signal is observable in early stage 32 embryos (Figures 2A,B,E,F), and becomes more intense in intermediate and late stage 32 embryos (Figures 2C,D,G,H). Note that neural arch mineralization never extends to the center of the cartilaginous scaffold (Figures 2A–H and Berio et al., 2021). In addition, the fibrous perichondrial tissue surrounding each neural arch displays a robust mineralization in intermediate and late stage 32 embryos. In the vertebral body, a mineralization ring appears in the cartilage surrounding the notochord of intermediate stage 32 embryos and becomes more mineralized in late stage 32 embryos (Figures 2F–H).

The expression of the *SPARC* gene was detected in the neural tube and several connective tissues such as the dermis and perimysium at all tested developmental stages (Figures 2I–L). We report three major sites of *SPARC* expression in the Sc developing vertebrae: the neural arch chondrocytes, the neural arch fibrous perichondrium, and the vertebral body. In the neural arches of early stage 32 embryos, *SPARC* expression localizes to peripheral chondrocytes (*i.e.*, specifically to the mineralizing cartilage) and to the cells of the fibrous perichondrium (Figures 2I,J). In intermediate stage 32 embryos *SPARC* expression extends to most chondrocytes of the neural arches (Figure 2K). Cells of the mineralized fibrous perichondrial tissue surrounding the neural arches also express *SPARC* in intermediate and late stage 32 embryos (Figures 2J,K). In the vertebral body from early and intermediate stage 32 embryos, *SPARC* is detected as a ring of expression in chondrocytes surrounding the notochord (Figures 2I–K). Our results on late stage 32 embryos show little gene expression of *SPARC* in the vertebral tissues, as only a faint signal was observed in some chondrocytes, (Figure 2L), revealing that the expression of this gene is dynamic and transient in relation to the mineralization processes. We had previously shown that *SPARC* is expressed in developing scales (Enault et al., 2018), and the expected signal in odontoblasts presents on the same section strongly argues against a possible technical problem for the detection method in late stage 32 embryos (Figure 2L, inset).

On the other hand, the expression of *SPARC-L* could not be detected in any endoskeletal tissues, while its expression in the epithelium (*i.e.*, the ameloblast layer) of developing and mineralized scales was observable at all stages (Figures 2M–P), as expected (Enault et al., 2018).



**FIGURE 2 |** Histology, mineralization dynamics and *SPARC* and *SPARC-L* expression pattern in the vertebrae and scales of the small-spotted catshark *Scyliorhinus canicula*. **(A–D)** Hematoxylin-Eosin Safran (HES) histological staining on transverse sections at the level of thoracic vertebrae: A, general view with location of the neural tube (nt) and notochord (nc); B–D, closer views on vertebral tissues as boxed in A, with identification of the fibrous perichondrium (fp), unmineralized cartilage (uc) and peripheral chondrocytes (pc) of the neural arch (na), as well as the unmineralized cartilage (uc) and fibrous cartilage (fc) of the vertebral body (vb); insets in B–D, closer view of the dorsal scales as boxed in A, indicating the location of the epithelial (e) and mesenchymal (m) compartments of scale buds. Stage 32 embryos were staged according to their total length into “early,” “intermediate,” and “late” categories as described in the Material and Method section. **(E)** Periodic Acid Schiff-Alcian (Continued)

**FIGURE 2** | Blue (PAS-BA) histological staining of a section consecutive to A: BA (blue) stains the acid glycosaminoglycans of the hyaline cartilage and PAS (pink) stains the fibrous content of the perichondrium. **(F-H)** Alizarin red S staining locates calcium deposits in mineralizing matrices [of the peripheral chondrocytes (pc), fibrous perichondrium (fp) or fibrous cartilage (fc), and scale enameloid/dentin], in embryos of similar total length as A-D. **(I-L)** *SPARC* gene expression patterns, for sections that are consecutive to those shown in **(F-H)** respectively. **(M-P)** *SPARC-L* gene expression patterns for sections that are consecutive to those shown in **(F-H)** respectively. Scales represent 200  $\mu$ m, except in scale insets where they represent 50  $\mu$ m.

## SPARC and SCPPs Expression in the Development of Limbs and Vertebrae in the Western Clawed Frog *Xenopus tropicalis*

We examined gene expression in NF58 *Xt* limbs because at this stage hypertrophic cartilage is in its most mature state and becomes eliminated and replaced by bone marrow at the diaphysis (Figure 3A). von Kossa staining showed an intense signal in the bone matrix and revealed that the *Xt* hypertrophic cartilage does not mineralize at the diaphysis of NF58 femoral bones (Figure 3B). *SPARC* transcripts were specifically detected in osteocytes and in osteoblasts of the periosteum and endosteum, but not in the cartilage (Figures 3C,D). A similar situation was observed for *BSP* (Figure 3E). *DMP1* was detected in osteocytes as well as in some chondrocytes of the diaphysis (Figure 3F). Transcripts from the *SCPPA2* gene were detected in osteocytes and some osteoblasts, and in many chondrocytes located at the cartilage to bone marrow transition and in the vicinity of the bone matrix of the diaphysis (Figure 3G) and of the epiphysis (Figure 3G').

Stage NF58 vertebrae (Figure 4A) were subjected to von Kossa staining, revealing cartilage mineralization in the dorsal region of the neural arches (Figures 4B,C), as well as in the ventral region located between the neural tube and the notochord (Figures 4B',C'), in agreement with previous observations performed with Alizarin red S (Enault et al., 2015). We found that each of the examined genes displays a distinctive expression pattern. *SPARC* is specifically expressed in osteoblasts of the dorsal neural arches and of the ventral region, but not in chondrocytes (Figures 4D,D',E,E'). *BSP* is expressed in osteoblasts and chondrocytes of both regions, albeit its expression is much stronger in the cartilage of the ventral vertebrae (Figures 4F,F'). *DMP1* is expressed in osteocytes and in chondrocytes located close to the bone matrix of the dorsal neural arch, but is not expressed in the ventral vertebra at this stage (Figures 4G,G'). *SCPPA2* is moderately expressed in some osteocytes and osteoblasts of the dorsal neural arch, and very strongly in chondrocytes of the mineralizing cartilage of both vertebral regions (Figures 4H,H').

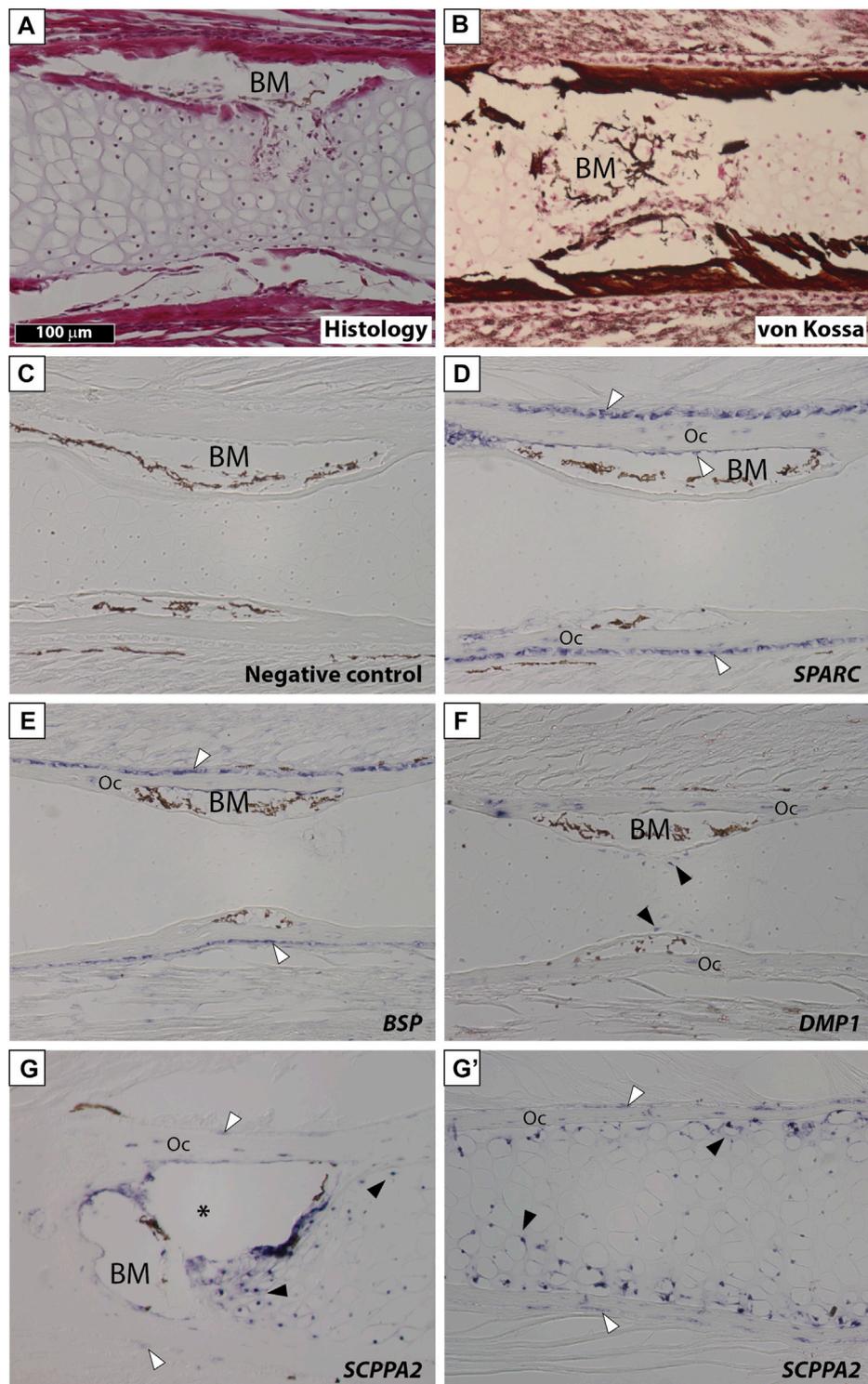
## DISCUSSION

Our findings in *Xt* reveal an evolutionary conservation of the cartilaginous expression of the *SCPP* genes in tetrapods. Indeed, similarly to the situation in *Xt*, *SPARC* is not expressed in mouse chondrocytes (Holland et al., 1987). Rather, *SCPP* genes such as *DMP1* and *BSP* are expressed and required for mouse cartilage development (Chen et al., 1991; Ye et al., 2005; Boulefour et al., 2014; Fujikawa et al., 2015). As indicated by other studies (Yagami et al., 1999; Bandyopadhyay et al., 2008), gene

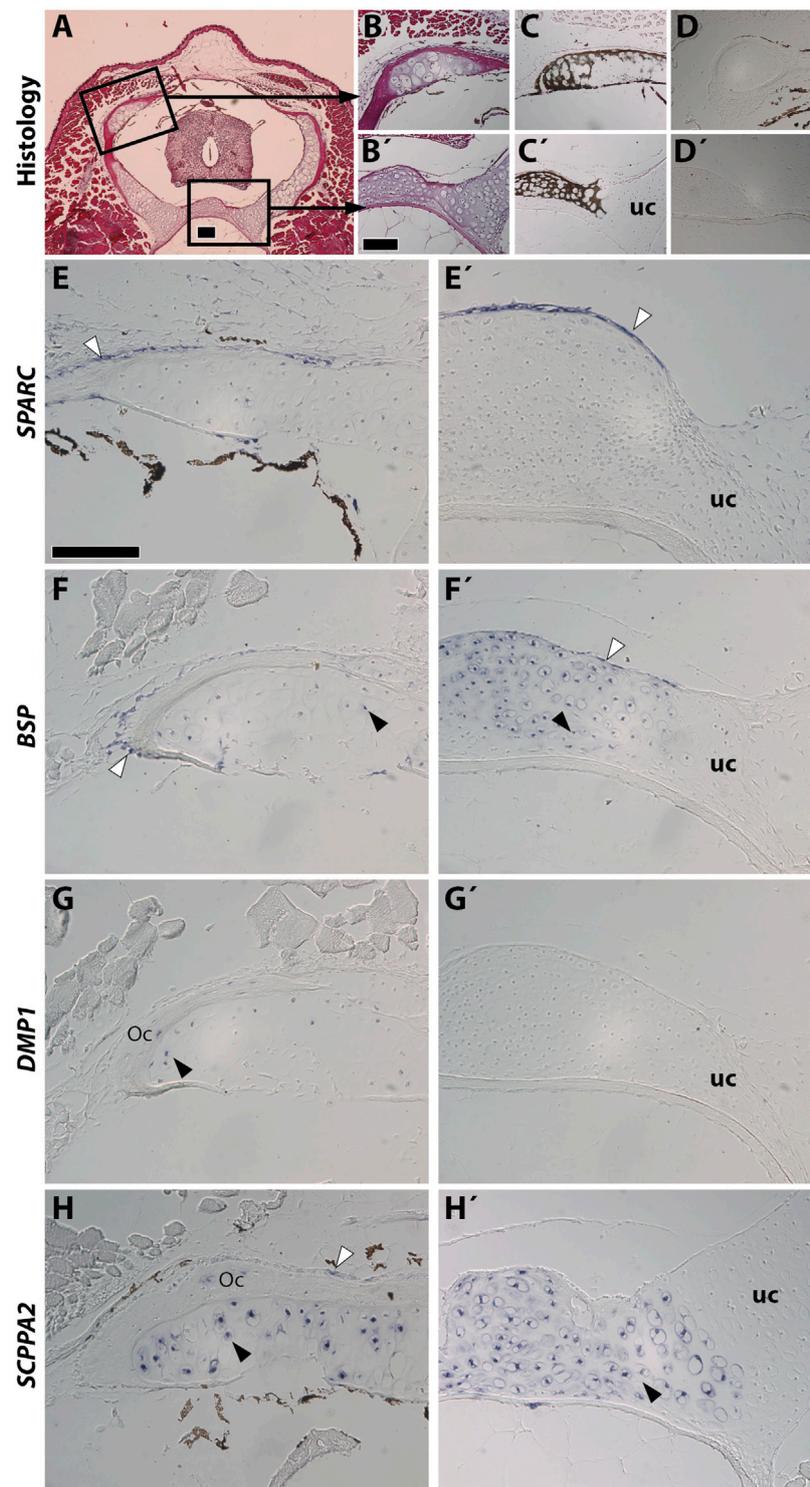
expression in cartilaginous elements can be subdivided in two distinct domains which we will use to discuss our results. First, *SCPP* genes become activated at late stages of hypertrophy, when the cartilage matrix becomes replaced by bone marrow at the mammalian diaphysis (Chen et al., 1991; Fujikawa et al., 2015). Likewise, in *Xt*, *DMP1* is exclusively expressed at the diaphysis (Figure 3F), and *SCPPA2* exhibits a much stronger expression at the diaphysis than the epiphysis region (Figures 3G,G'). A similar situation is observed at the level of the *Xt* vertebrae, where the expression of *SCPP* genes tightly correlates with cartilage maturation and mineralization in the neural arch (for *BSP*, *DMP1*, and *SCPPA2*) as well as in the ventral vertebral region (for *BSP* and *SCPPA2*). Second, the *SCPP* genes harbor a stronger expression in the non-mineralized peripheral cartilage, as observed in mouse for *osteopontin* (Heilig et al., 2016) and *DMP1* (Fujikawa et al., 2015). This situation is similar to the expression of *Xt* *SSCP* genes in chondrocytes located in the vicinity of the bone matrix in long bones and vertebrae (Figures 3, 4). Such a peripheral cartilage domain expresses specific genes, as reported in chick (Bandyopadhyay et al., 2008), and undergoes ectopic mineralization in mutant mouse animals for the *Mgp* (Marulanda et al., 2017) and *Trps1* (Napierala et al., 2008) genes. In summary, *SCPP* genes from frog and mouse are expressed in the mature cartilage of the diaphysis and neural arches, as well as in peripherally located chondrocytes.

Available expression analyses did not report any cartilaginous expression for *SCPP* genes in teleosts (Kawasaki et al., 2005; Kawasaki, 2009; Weigele et al., 2015). Rather, the expression of the *SPARC* gene has been associated to cartilage development in zebrafish, gilthead seabream and the cichlid mouth breeder (Estevao et al., 2005; Redruello et al., 2005; Rotllant et al., 2008; Estevao et al., 2011; Weigele et al., 2015), albeit not in medaka, at least at the examined stages (Renn et al., 2006). Hence the reported cartilaginous expression patterns in teleosts (*SPARC* positive and *SCPP* negative) are opposite to the tetrapod situation (*SPARC* negative and *SCPP* positive), which might be related to drastic difference in the mode of endochondral ossification between these two groups (Cervantes-Diaz et al., 2017). In this respect, our data in the chondrichthyan representative *Sc* is instrumental to understand the evolution of the expression of these genes in the jawed vertebrate endoskeleton.

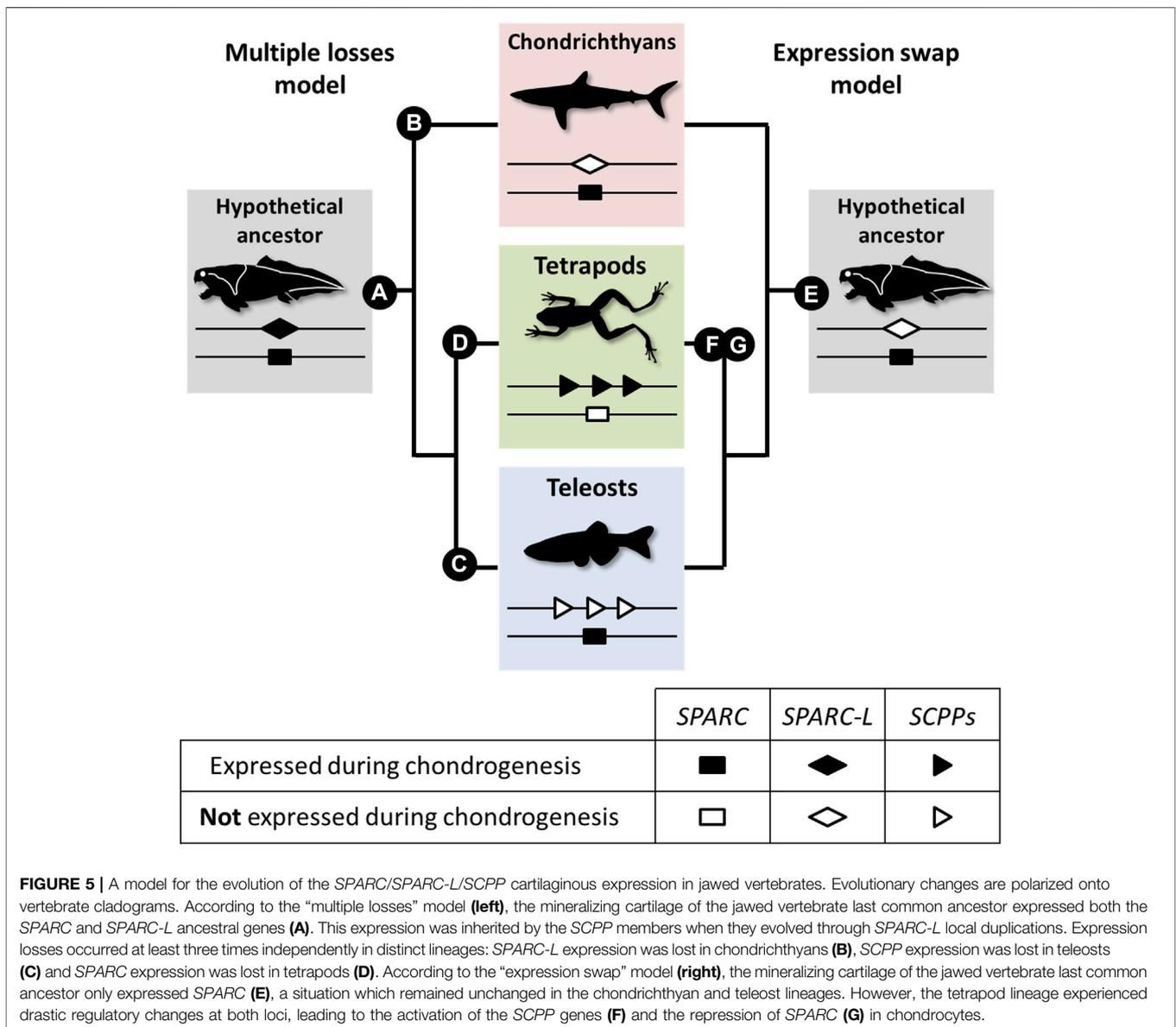
As no *SCPP* genes have been reported in chondrichthyan genomes to date, we focused on the evolutionarily related gene *SPARC-L* (Kawasaki and Weiss, 2003; Bertrand et al., 2013; Venkatesh et al., 2014; Enault et al., 2018). Our finding that *Sc* *SPARC-L* is not expressed in the vertebral cartilage is further confirmed by the robust and expected *Sc* *SPARC-L* expression in the odontodes present on the same sections and serving as convenient internal positive controls (Enault et al., 2018). By



**FIGURE 3** | *SPARC* and *SCPP* gene expression in *Xenopus tropicalis* stage NF58 hindlimbs. Longitudinal sections of *Xenopus tropicalis* stage NF58 femoral bones were subjected to Hematoxylin-Eosin staining (**A**), von Kossa staining (**B**), or *in situ* hybridization using a negative control [*SPARC* sense probe, (**C**)] or an antisense probe for *SPARC* (**D**), *BSP* (**E**), *DMP1* (**F**), and *SCPPA2* (**G**, **G'**). Pictures show the diaphysis in (**A–G**) and the epiphysis in (**G'**). White and black arrowheads show *in situ* hybridization signal in osteoblasts and chondrocytes, respectively. Abbreviations: BM, bone marrow, Oc osteocytes showing *in situ* hybridization signal. Scale bar in (**A**) represents 100  $\mu$ m and applies to all panels. The asterisk shows an artifact due to the cartilage which teared apart and contracted in this region of the section.



**FIGURE 4** | *SPARC* and *SCPP* gene expression in *Xenopus tropicalis* stage NF58 vertebrae. Transversal sections of *Xenopus tropicalis* stage NF58 vertebrae were subjected to Hematoxylin-Eosin staining (**A,B,B'**), von Kossa staining (**C,C'**), or *in situ* hybridization using a negative control [*SPARC* sense probe, (**D,D'**)] or an antisense probe for *SPARC* (**E,E'**), *BSP* (**F,F'**), *DMP1* (**G, G'**), and *SCPPA2* (**H,H'**). Pictures are orientated with dorsal to the top and show the whole vertebrae (**A**), the neural arch (**B–H**) or the vertebral body (**B'–H'**). White and black arrowheads show *in situ* hybridization signal in osteoblasts and chondrocytes, respectively. Abbreviations: uc, unmineralized cartilage, Oc osteocytes showing *in situ* hybridization signal. Scale bars: (**A**), 100  $\mu$ m; (**B**), 100  $\mu$ m in (**B–D'**); (**E**), 50  $\mu$ m in (**E–H'**).



contrast, *SPARC* expression clearly co-localizes with sites of vertebral mineralization. In the neural arches of early stage 32 embryos, *SPARC* is restricted to the mineralizing peripheral cartilage matrix, thereby paralleling the expression of *SCPP* genes in frog (Figures 2, 4) and mouse (Fujikawa et al., 2015; Heilig et al., 2016). Hence, we uncover a novel evolutionarily conserved cartilage domain, as defined by peripherally located chondrocytes expressing *SPARC* in chondrichthyans and the *SCPP* genes in tetrapods. One difference is that this domain mineralizes in chondrichthyans, but not in tetrapods. We propose that dosage variations between pro- and anti-mineralizing proteins might explain evolutionary differences between vertebrate lineages, as might be the case for instance for *MGP* which is a well-conserved cartilage mineralization inhibitor (Yagami et al., 1999; Espinoza et al., 2010; Viegas et al., 2013; Marulanda et al., 2017; Leurs et al., 2021). By examining the

centrum of *Sc* specimens from different developmental stages we show that a ring of *SPARC* expression is already present in Alizarin red-negative early stage 32 embryos, suggesting that the presence of the *SPARC* protein plays a crucial role in the initiation of mineralization. The functional interaction between the mammalian *SPARC* and collagen 1 proteins is important for mineralization (Termine et al., 1981). However, as the shark chondrocytes embedded within a mineralizing cartilage matrix neither expresses *collagen 1a1* nor *collagen 1a2* (Enault et al., 2015), *Sc SPARC* function might be related to other aspects of matrix mineralization, such as its interaction with calcium and hydroxyapatite crystals (Engel et al., 1987; Maurer et al., 1995; Fujisawa et al., 1996).

Our data suggest that chondrichthyans are more similar to teleosts than to tetrapods because the *Sc* mineralizing cartilage is *SPARC* positive and *SPARC-L* negative. Here, we propose two

evolutionary models to explain these divergent expression patterns (see **Figure 5** and **Supplementary Figure S1**). Both models are based on the assumption that *SPARC*, *SPARC-L*, and *SCPP* share some level of functional redundancy, at least during chondrogenesis, as suggested by the fact that both *SPARC* and *SCPP* proteins are extracellular transglutaminase substrates (Aeschlimann et al., 1995; Forsprecher et al., 2011) and bind calcium ions (Engel et al., 1987; Chen et al., 1992; Maurer et al., 1995; Klaning et al., 2014). The “multiple losses” model is reminiscent of the Duplication Degeneration Complementation (DDC) phenomenon (Force et al., 1999) and involves at least three independent changes which abrogated the cartilaginous expression of *SPARC*, of *SPARC-L* or of the *SCPP* genes in distinct lineages (**Figure 5**, left panel). The “expression swap” model involves two changes and implies that, in the tetrapod lineage, cartilaginous expression was gained for the *SCPP* genes and lost for *SPARC* (**Figure 5**, right panel). While the *SPARC/SPARC-L/SCPP* family exhibit a unique evolutionary trajectory (Bertrand et al., 2013; Enault et al., 2018), the “expression swap” model would be similar to the dynamic expression turnover observed between members of the *Keratin*, *Vitellogenin*, and *Globin* vertebrate families (Finn et al., 2009; Vandebergh and Bossuyt, 2012; Opazo et al., 2015). Both scenarios imply regulatory changes that switched off (“multiple losses model”) or on (“expression swap model”) several *SCPP* genes (**Figure 5**). From a regulatory perspective, the idea of coordinated change in the expression of tandemly located *SCPP* genes is consistent with the fact that *BSP* and *DMP1* are included within the same topological association domain in human chromatin (Krietenstein et al., 2020), and that multiple genes can be co-regulated by the same enhancer (reviewed in Zheng and Xie, 2019). To the best of our knowledge, current experimental evidence is not sufficient to discriminate between the two models shown in **Figure 5**. Hence, a detailed analysis of the expression and regulation of *SPARC/SPARC-L/SCPP* genes in chondrocyte from a broader array of chondrichthyan and osteichthyan species will be required to shed light on the genetic mechanisms involved in the evolution of cartilage mineralization that originated deep in the vertebrate lineage (Min and Janvier, 1998; Donoghue et al., 2006; Johanson et al., 2010; 2012).

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusion of this article will be made available by the authors, without undue reservation.

## ETHICS STATEMENT

Handling of small-spotted catshark embryos followed all institutional, national, and international guidelines [European

Communities Council Directive of September 22, 2010 (2010/63/UE)]; no further approval by an ethics committee was necessary as the biological material is embryonic and no live experimental procedures were carried out. The Ethics Committee of the University of Concepcion (Concepcion, Chile) approved all experimental procedures carried out on *Xenopus tropicalis* tadpoles during this study, which were performed following the guidelines outlined in the Biosafety and Bioethics Manual of the National Commission of Scientific and Technological Research (CONICYT, Chilean Government).

## AUTHOR CONTRIBUTIONS

AR and DM performed the histological analyses and *in situ* hybridization procedures in *Xenopus tropicalis*. NL and MD-T performed the histological analyses and *in situ* hybridization procedures in *Scyliorhinus canicula*. SM and MD-T wrote the manuscript and prepared the figures. All authors analyzed and interpreted the data and read and approved the manuscript.

## FUNDING

We acknowledge the MRI platform, member of the national infrastructure France-BioImaging supported by the French National Research Agency (ANR-10-INBS-04, «Investments for the future»), the labex CEMEB (ANR-10-LABX-0004), and NUMEV (ANR-10-LABX-0020). We also wish to acknowledge the “Réseau d’Histologie Expérimentale de Montpellier”—RHEM facility supported by the SIRIC Montpellier Cancer (Grant INCa\_Inserm\_DGOS\_12553), the european regional development foundation and the occitanian region (FEDER-FSE 2014-2020 Languedoc Roussillon) for processing our animal tissues, histology techniques, and expertise. This study was partially funded by the IRP (LIA) MAST (CNRS). Fondecyt 1190926 to SM. AR and DM were ANID fellows 21170393 and 21160069 respectively.

## ACKNOWLEDGMENTS

We thank Stéphanie Ventéo and Camille Martinand-Mari for their help with the biological material and *in situ* hybridization technique.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.788346/full#supplementary-material>

## REFERENCES

- Aeschlimann, D., Kaupp, O., and Paulsson, M. (1995). Transglutaminase-catalyzed Matrix Cross-Linking in Differentiating Cartilage: Identification of Osteonectin as a Major Glutaminyl Substrate. *J. Cel Biol* 129, 881–892. doi:10.1083/jcb.129.3.881
- Ballard, W., Mellinger, J., and Lechenault, H. (1993). A Series of Stages for Development of *Scyliorhinus canicula* the Lesser Spotted Dogfish (Chondrichthyes: Scyliorhinidae). *J. Exp. Zool* 267, 1–43. doi:10.1002/jez.1402670309
- Bandyopadhyay, A., Kubilus, J. K., Crochiere, M. L., Linsenmayer, T. F., and Tabin, C. J. (2008). Identification of Unique Molecular Subdomains in the Perichondrium and Periosteum and Their Role in Regulating Gene Expression in the Underlying Chondrocytes. *Develop. Biol.* 321, 162–174. doi:10.1016/j.ydbio.2008.06.012
- Beresford, W. A. (1981). Chondroid Bone, Secondary Cartilage, and Metaplasia. *The American Journal of Surgical Pathology* 5, 405.
- Berio, F., Broyon, M., Enault, S., Pirot, N., López-Romero, F. A., and Debais-Thibaud, M. (2021). Diversity and Evolution of Mineralized Skeletal Tissues in Chondrichthyan. *Front. Ecol. Evol.* 9, 1–19. doi:10.3389/fevo.2021.660767
- Bertrand, S., Fuentealba, J., Aze, A., Hudson, C., Yasuo, H., Torrejon, M., et al. (2013). A Dynamic History of Gene Duplications and Losses Characterizes the Evolution of the SPARC Family in Eumetazoans. *Proc. R. Soc. B* 280, 20122963. doi:10.1098/rspb.2012.2963
- Bouletfour, W., Boudiffa, M., Wade-Gueye, N. M., Bouët, G., Cardelli, M., Laroche, N., et al. (2014). Skeletal Development of Mice Lacking Bone Sialoprotein (BSP) - Impairment of Long Bone Growth and Progressive Establishment of High Trabecular Bone Mass. *PLoS One* 9, e95144. doi:10.1371/journal.pone.0095144
- Brazeau, M. D., Giles, S., Dearden, R. P., Jerve, A., Ariunchimeg, Y., Zorig, E., et al. (2020). Endochondral Bone in an Early Devonian 'placoderm' from Mongolia. *Nat. Ecol. Evol.* 4, 1477–1484. doi:10.1038/s41559-020-01290-2
- Cervantes-Diaz, F., Contreras, P., and Marcellini, S. (2017). Evolutionary Origin of Endochondral Ossification: the Transdifferentiation Hypothesis. *Dev. Genes Evol.* 227, 121–127. doi:10.1007/s00427-016-0567-y
- Chen, J., Shapiro, H. S., Wrana, J. L., Reimers, S., Heersche, J. N. M., and Sodek, J. (1991). Localization of Bone Sialoprotein (BSP) Expression to Sites of Mineralized Tissue Formation in Fetal Rat Tissues by *In Situ* Hybridization. *Matrix* 11, 133–143. doi:10.1016/s0934-8832(11)80217-9
- Chen, Y., Bal, B. S., and Gorski, J. P. (1992). Calcium and Collagen Binding Properties of Osteopontin, Bone Sialoprotein, and Bone Acidic Glycoprotein-75 from Bone. *J. Biol. Chem.* 267, 24871–24878. doi:10.1016/s0021-9258(18)35844-7
- Close, B., Banister, K., Baumans, V., Bernoth, E.-M., Bromage, N., Bunyan, J., et al. (1996). Recommendations for Euthanasia of Experimental Animals: Part 1. *Lab. Anim.* 30, 293–316. doi:10.1258/002367796780739871
- Donoghue, P. C. J., Sansom, I. J., and Downs, J. P. (2006). Early Evolution of Vertebrate Skeletal Tissues and Cellular Interactions, and the Canalization of Skeletal Development. *J. Exp. Zool.* 306B, 278–294. doi:10.1002/jez.b.21090
- Donoghue, P. C. J., and Sansom, I. J. (2002). Origin and Early Evolution of Vertebrate Skeletonization. *Microsc. Res. Tech.* 59, 352–372. doi:10.1002/jemt.10217
- Dymnt, N. A., Breidenbach, A. P., Schwartz, A. G., Russell, R. P., Aschbacher-Smith, L., Liu, H., et al. (2015). Gdf5 Progenitors Give Rise to Fibrocartilage Cells that Mineralize via Hedgehog Signaling to Form the Zonal Enthesis. *Develop. Biol.* 405, 96–107. doi:10.1016/j.ydbio.2015.06.020
- Enault, S., Muñoz, D. N., Silva, W. T. A. F., Borday-Birraux, V., Bonade, M., Oulion, S., et al. (2015). Molecular Footprinting of Skeletal Tissues in the Catshark *Scyliorhinus canicula* and the Clawed Frog *Xenopus tropicalis* Identifies Conserved and Derived Features of Vertebrate Calcification. *Front. Genet.* 6, 283. doi:10.3389/fgene.2015.00283
- Enault, S., Muñoz, D., Simion, P., Ventéo, S., Sire, J.-Y., Marcellini, S., et al. (2018). Evolution of Dental Tissue Mineralization: an Analysis of the Jawed Vertebrate SPARC and SPARC-L Families. *BMC Evol. Biol.* 18, 127. doi:10.1186/s12862-018-1241-y
- Engel, J., Taylor, W., Paulsson, M., Sage, H., and Hogan, B. (1987). Calcium Binding Domains and Calcium-Induced Conformational Transition of SPARC/BM-40/osteonectin, an Extracellular Glycoprotein Expressed in Mineralized and Nonmineralized Tissues. *Biochemistry* 26, 6958–6965. doi:10.1021/bi00396a015
- Espinoza, J., Sanchez, M., Sanchez, A., Hanna, P., Torrejon, M., Buisine, N., et al. (2010). Two Families of *Xenopus tropicalis* Skeletal Genes Display Well-Conserved Expression Patterns with Mammals in Spite of Their Highly Divergent Regulatory Regions. *Evol. Dev.* 12, 541–551. doi:10.1111/j.1525-142X.2010.00440.x
- Estêvão, M. D., Redruello, B., Canario, A. V. M., and Power, D. M. (2005). Ontogeny of Osteonectin Expression in Embryos and Larvae of Sea Bream (*Sparus auratus*). *Gen. Comp. Endocrinol.* 142, 155–162. doi:10.1016/j.ygcen.2004.11.018
- Estêvão, M. D., Silva, N., Redruello, B., Costa, R., Gregório, S., Canário, A. V. M., et al. (2011). Cellular Morphology and Markers of Cartilage and Bone in the marine Teleost *Sparus auratus*. *Cell Tissue Res* 343, 619–635. doi:10.1007/s00441-010-1109-y
- Finn, R., Kolarevic, J., Kongshaug, H., and Nilsen, F. (2009). Evolution and Differential Expression of a Vertebrate Vitellogenin Gene Cluster. *BMC Evol. Biol.* 9, 2. doi:10.1186/1471-2148-9-2
- Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y.-l., and Postlethwait, J. (1999). Preservation of Duplicate Genes by Complementary, Degenerative Mutations. *Genetics* 151, 1531–1545. doi:10.1093/genetics/151.4.1531
- Forsprecher, J., Wang, Z., Goldberg, H. A., and Kaartinen, M. T. (2011). Transglutaminase-mediated Oligomerization Promotes Osteoblast Adhesive Properties of Osteopontin and Bone Sialoprotein. *Cell Adhes. Migration* 5, 65–72. doi:10.4161/cam.5.1.13369
- Fujikawa, K., Yokohama-Tamaki, T., Morita, T., Baba, O., Qin, C., and Shibata, S. (2015). An *In Situ* Hybridization Study of Perlecan, DMP1, and MEPE in Developing Condylar Cartilage of the Fetal Mouse Mandible and Limb Bud Cartilage. *Eur. J. Histochem.* 59, 2553. doi:10.4081/ejh.2015.2553
- Fujisawa, R., Wada, Y., Nodasaka, Y., and Kuboki, Y. (1996). Acidic Amino Acid-Rich Sequences as Binding Sites of Osteonectin to Hydroxyapatite Crystals. *Biochim. Biophys. Acta (Bba) - Protein Struct. Mol. Enzymol.* 1292, 53–60. doi:10.1016/0167-4838(95)00190-5
- Heilig, J., Paulsson, M., and Zaucke, F. (2016). Insulin-like Growth Factor 1 Receptor (IGF1R) Signaling Regulates Osterix Expression and Cartilage Matrix Mineralization during Endochondral Ossification. *Bone* 83, 48–57. doi:10.1016/j.bone.2015.10.007
- Hirasawa, T., and Kuratani, S. (2015). Evolution of the Vertebrate Skeleton: Morphology, Embryology, and Development. *Zoolog. Lett.* 1. doi:10.1186/s40851-014-0007-7
- Holland, P. W., Harper, S. J., Mcvey, J. H., and Hogan, B. L. (1987). *In Vivo* expression of mRNA for the Ca<sup>++</sup>-Binding Protein SPARC (Osteonectin) Revealed by *In Situ* Hybridization. *J. Cel Biol* 105, 473–482. doi:10.1083/jcb.105.1.473
- Johanson, Z., Kearsley, A., Den Blaauwen, J., Newman, M., and Smith, M. M. (2010). No Bones about it: An Enigmatic Devonian Fossil Reveals a New Skeletal Framework-A Potential Role of Loss of Gene Regulation. *Semin. Cel Develop. Biol.* 21, 414–423. doi:10.1016/j.semcdb.2009.10.011
- Johanson, Z., Kearsley, A., Den Blaauwen, J., Newman, M., and Smith, M. M. (2012). Ontogenetic Development of an Exceptionally Preserved Devonian Cartilaginous Skeleton. *J. Exp. Zool.* 318B, 50–58. doi:10.1002/jez.b.21441
- Kawasaki, K., Buchanan, A. V., and Weiss, K. M. (2007). Gene Duplication and the Evolution of Vertebrate Skeletal Mineralization. *Cells Tissues Organs* 186, 7–24. doi:10.1159/000102678
- Kawasaki, K., Suzuki, T., and Weiss, K. M. (2005). Phenogenetic Drift in Evolution: the Changing Genetic Basis of Vertebrate Teeth. *Proc. Natl. Acad. Sci.* 102, 18063–18068. doi:10.1073/pnas.0509263102
- Kawasaki, K. (2009). The SCPP Gene Repertoire in Bony Vertebrates and Graded Differences in Mineralized Tissues. *Dev. Genes Evol.* 219, 147–157. doi:10.1007/s00427-009-0276-x
- Kawasaki, K., and Weiss, K. M. (2003). Mineralized Tissue and Vertebrate Evolution: the Secretory Calcium-Binding Phosphoprotein Gene Cluster. *Proc. Natl. Acad. Sci.* 100, 4060–4065. doi:10.1073/pnas.0638023100
- Kessels, M. Y., Huitema, L. F. A., Boeren, S., Kranenbarg, S., Schulte-Merker, S., Van Leeuwen, J. L., et al. (2014). Proteomics Analysis of the Zebrafish Skeletal Extracellular Matrix. *PLoS One* 9, e90568. doi:10.1371/journal.pone.0090568
- Kläning, E., Christensen, B., Sørensen, E. S., Vorup-Jensen, T., and Jensen, J. K. (2014). Osteopontin Binds Multiple Calcium Ions with High Affinity and

- Independently of Phosphorylation Status. *Bone* 66, 90–95. doi:10.1016/j.bone.2014.05.020
- Krietenstein, N., Abraham, S., Venev, S. V., Abdennur, N., Gibcus, J., Hsieh, T.-H. S., et al. (2020). Ultrastructural Details of Mammalian Chromosome Architecture. *Mol. Cell* 78, 554–565. e557. doi:10.1016/j.molcel.2020.03.003
- Leurs, N., Martinand-Mari, C., Ventéo, S., Haitina, T., and Debiais-Thibaud, M. (2021). Evolution of Matrix Gla and Bone Gla Protein Genes in Jawed Vertebrates. *Front. Genet.* 12, 620659. doi:10.3389/fgene.2021.620659
- Li, N., Felber, K., Elks, P., Croucher, P., and Roehl, H. H. (2009). Tracking Gene Expression during Zebrafish Osteoblast Differentiation. *Dev. Dyn.* 238, 459–466. doi:10.1002/dvdy.21838
- Long, F., and Ornitz, D. M. (2013). Development of the Endochondral Skeleton. *Cold Spring Harbor Perspect. Biol.* 5, a008334. doi:10.1101/cshperspect.a008334
- Marulanda, J., Eimar, H., Mckee, M. D., Berkvens, M., Nelea, V., Roman, H., et al. (2017). Matrix Gla Protein Deficiency Impairs Nasal Septum Growth, Causing Midface Hypoplasia. *J. Biol. Chem.* 292, 11400–11412. doi:10.1074/jbc.M116.769802
- Maurer, P., Hohenadl, C., Hohenester, E., Göhring, W., Timpl, R., and Engel, J. (1995). The C-Terminal Portion of BM-40 (SPARC/osteonectin) Is an Autonomously Folding and Crystallisable Domain that Binds Calcium and Collagen IV. *J. Mol. Biol.* 253, 347–357. doi:10.1006/jmbi.1995.0557
- Maxwell, E. E., Fröbisch, N. B., and Heppleston, A. C. (2008). Variability and Conservation in Late Chondrichthyan Development: Ontogeny of the winter Skate (*Leucoraja ocellata*). *Anat. Rec.* 291, 1079–1087. doi:10.1002/ar.20719
- Mckinnon, P. J., Mclaughlin, S. K., Kapssetaki, M., and Margolskee, R. F. (2000). Extracellular Matrix-Associated Protein Sc1 Is Not Essential for Mouse Development. *Mol. Cell Biol.* 20, 656–660. doi:10.1128/mcb.20.2.656-660.2000
- Min, Z., and Janvier, P. (1998). The Histological Structure of the Endoskeleton in Galeaspid (Galeaspid, Vertebrata). *J. Vertebr. Paleontol.* 18, 650–654. doi:10.1080/02724634.1998.10011091
- Nakatani, Y., Shingate, P., Ravi, V., Pillai, N. E., Prasad, A., Mclysaght, A., et al. (2021). Reconstruction of Proto-Vertebrate, Proto-Cyclostome and Proto-Gnathostome Genomes Provides New Insights into Early Vertebrate Evolution. *Nat. Commun.* 12, 4489. doi:10.1038/s41467-021-24573-z
- Napierala, D., Sam, K., Morello, R., Zheng, Q., Munivez, E., Shivdasani, R. A., et al. (2008). Uncoupling of Chondrocyte Differentiation and Perichondrial Mineralization Underlies the Skeletal Dysplasia in Tricho-Rhino-Phalangeal Syndrome. *Hum. Mol. Genet.* 17, 2244–2254. doi:10.1093/hmg/ddn125
- Nieuwkoop, P. D., and Faber, J. (1967). *Normal Table of Xenopus laevis (Daudin)*. Netherlands: North Holland.
- Nikoloudaki, G. (2021). Functions of Matricellular Proteins in Dental Tissues and Their Emerging Roles in Orofacial Tissue Development, Maintenance, and Disease. *Ijms* 22, 6626. doi:10.3390/ijms22126626
- Opazo, J. C., Hoffmann, F. G., Natarajan, C., Witt, C. C., Berenbrink, M., and Storz, J. F. (2015). Gene Turnover in the Avian Globin Gene Families and Evolutionary Changes in Hemoglobin Isoform Expression. *Mol. Biol. Evol.* 32, 871–887. doi:10.1093/molbev/msu341
- 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.
- Paul, S., Schindler, S., Giovannone, D., De Millo Terrazzani, A., Mariani, F. V., and Crump, J. G. (2016). Ihha Induces Hybrid Cartilage-Bone Cells during Zebrafish Jawbone Regeneration. *Development* 143, 2066–2076. doi:10.1242/dev.131292
- Pears, J. B., Johanson, Z., Trinajstić, K., Dean, M. N., and Boisvert, C. A. (2020). Mineralization of the Callorhynchus Vertebral Column (Holocephali; Chondrichthyes). *Front. Genet.* 11, 571694. doi:10.3389/fgene.2020.571694
- Ramlochansingh, C., Branoner, F., Chagnaud, B. P., and Straka, H. (2014). Efficacy of Tricaine Methanesulfonate (MS-222) as an Anesthetic Agent for Blocking Sensory-Motor Responses in *Xenopus laevis* Tadpoles. *PLoS One* 9, e101606. doi:10.1371/journal.pone.0101606
- Redruello, B., Estêvão, M. D., Rotllant, J., Guerreiro, P. M., Anjos, L. I., Canário, A. V., et al. (2005). Isolation and Characterization of Piscine Osteonectin and Downregulation of its Expression by PTH-Related Protein. *J. Bone Miner Res.* 20, 682–692. doi:10.1359/JBMR.041201
- Renn, J., Schaedel, M., Volff, J.-N., Goerlich, R., Schartl, M., and Winkler, C. (2006). Dynamic Expression of Sparc Precedes Formation of Skeletal Elements in the Medaka (*Oryzias latipes*). *Gene* 372, 208–218. doi:10.1016/j.gene.2006.01.011
- Rosset, E. M., and Bradshaw, A. D. (2016). SPARC/osteonectin in Mineralized Tissue. *Matrix Biol.* 52–54, 78–87. doi:10.1016/j.matbio.2016.02.001
- Rotllant, J., Liu, D., Yan, Y.-L., Postlethwait, J. H., Westerfield, M., and Du, S.-J. (2008). Sparc (Osteonectin) Functions in Morphogenesis of the Pharyngeal Skeleton and Inner Ear. *Matrix Biol.* 27, 561–572. doi:10.1016/j.matbio.2008.03.001
- Ryll, B., Sanchez, S., Haitina, T., Tafforeau, P., and Ahlberg, P. E. (2014). The Genome of Callorhynchus and the Fossil Record: a New Perspective on SCPP Gene Evolution in Gnathostomes. *Evol. Develop.* 16, 123–124. doi:10.1111/ede.12071
- Schreweis, M. A., Butler, J. P., Kulkarni, N. H., Knierman, M. D., Higgs, R. E., Halladay, D. L., et al. (2007). A Proteomic Analysis of Adult Rat Bone Reveals the Presence of Cartilage/chondrocyte Markers. *J. Cel. Biochem.* 101, 466–476. doi:10.1002/jcb.21196
- Termine, J. D., Kleinman, H. K., Whitson, S. W., Conn, K. M., Mcgarvey, M. L., and Martin, G. R. (1981). Osteonectin, a Bone-specific Protein Linking mineral to Collagen. *Cell* 26, 99–105. doi:10.1016/0092-8674(81)90037-4
- Tsang, K. Y., Chan, D., and Cheah, K. S. E. (2015). Fate of Growth Plate Hypertrophic Chondrocytes: Death or Lineage Extension? *Develop. Growth Differ.* 57, 179–192. doi:10.1111/dgd.12203
- Ustriyana, P., Schulte, F., Gombedza, F., Gil-Bona, A., Paruchuri, S., Bidlack, F. B., et al. (2021). Spatial Survey of Non-collagenous Proteins in Mineralizing and Non-mineralizing Vertebrate Tissues *Ex Vivo*. *Bone Rep.* 14, 100754. doi:10.1016/j.bonr.2021.100754
- Vandebergh, W., and Bossuyt, F. (2012). Radiation and Functional Diversification of Alpha Keratins during Early Vertebrate Evolution. *Mol. Biol. Evol.* 29, 995–1004. doi:10.1093/molbev/msr269
- Venkatesh, B., Lee, A. P., Ravi, V., Maurya, A. K., Lian, M. M., Swann, J. B., et al. (2014). Elephant Shark Genome Provides Unique Insights into Gnathostome Evolution. *Nature* 505, 174–179. doi:10.1038/nature12826
- Viegas, C. S. B., Simes, D. C., Williamson, M. K., Cavaco, S., Laizé, V., Price, P. A., et al. (2013). Sturgeon Osteocalcin Shares Structural Features with Matrix Gla Protein. *J. Biol. Chem.* 288, 27801–27811. doi:10.1074/jbc.M113.450213
- Weigele, J., Franz-Odenaal, T. A., and Hilbig, R. (2015). Expression of SPARC and the Osteopontin-like Protein during Skeletal Development in the Cichlid Fish *Oreochromis mossambicus*. *Dev. Dyn.* 244, 955–972. doi:10.1002/dvdy.24293
- Yagami, K., Suh, J.-Y., Enomoto-Iwamoto, M., Koyama, E., Abrams, W. R., Shapiro, I. M., et al. (1999). Matrix GLA Protein Is a Developmental Regulator of Chondrocyte Mineralization and, when Constitutively Expressed, Blocks Endochondral and Intramembranous Ossification in the Limb. *J. Cel Biol* 147, 1097–1108. doi:10.1083/jcb.147.5.1097
- Ye, L., Mishina, Y., Chen, D., Huang, H., Dallas, S. L., Dallas, M. R., et al. (2005). Dmp1-deficient Mice Display Severe Defects in Cartilage Formation Responsible for a Chondrodysplasia-like Phenotype. *J. Biol. Chem.* 280, 6197–6203. doi:10.1074/jbc.M412911200
- Zhang, G., and Cohn, M. J. (2008). Genome Duplication and the Origin of the Vertebrate Skeleton. *Curr. Opin. Genet. Develop.* 18, 387–393. doi:10.1016/j.gde.2008.07.009
- Zheng, H., and Xie, W. (2019). The Role of 3D Genome Organization in Development and Cell Differentiation. *Nat. Rev. Mol. Cel Biol* 20, 535–550. doi:10.1038/s41580-019-0132-4

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2021 Romero, Leurs, Muñoz, Debiais-Thibaud and Marcellini. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

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

Nicolas Leurs,<sup>1</sup> Camille Martinand-Mari,<sup>1</sup> Sylvain Marcellini,<sup>\*2</sup> and Mélanie Debiais-Thibaud<sup>\*,1</sup>

<sup>1</sup>Institut des Sciences de l'Évolution de Montpellier, ISEM, Univ Montpellier, CNRS, IRD, Montpellier, France

<sup>2</sup>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; Ustrian 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

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

Open Access

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

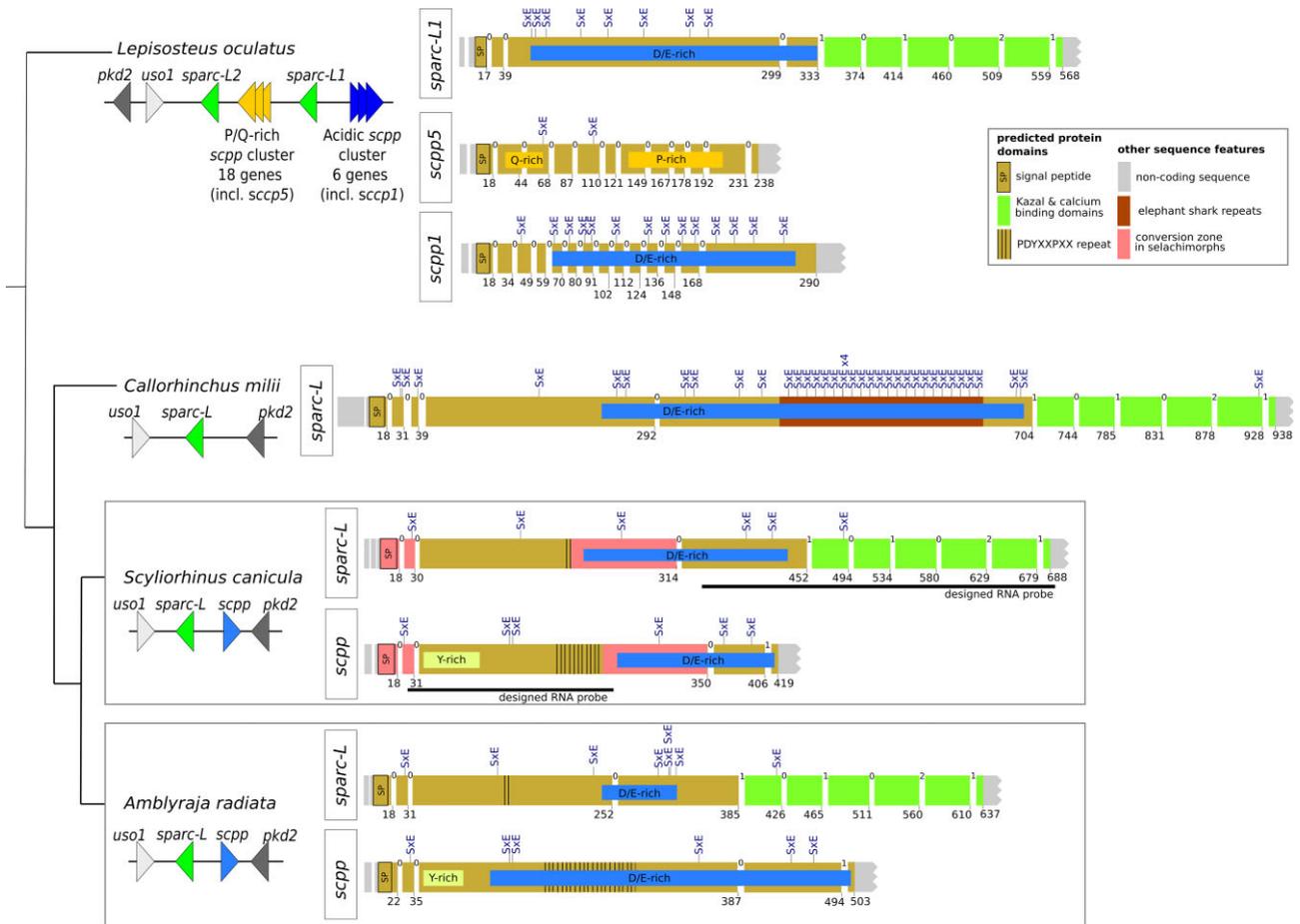
### A Novel Elasmobranch *scpp* Gene

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. 2A, 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. 2C, 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 *Callorhynchus 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.

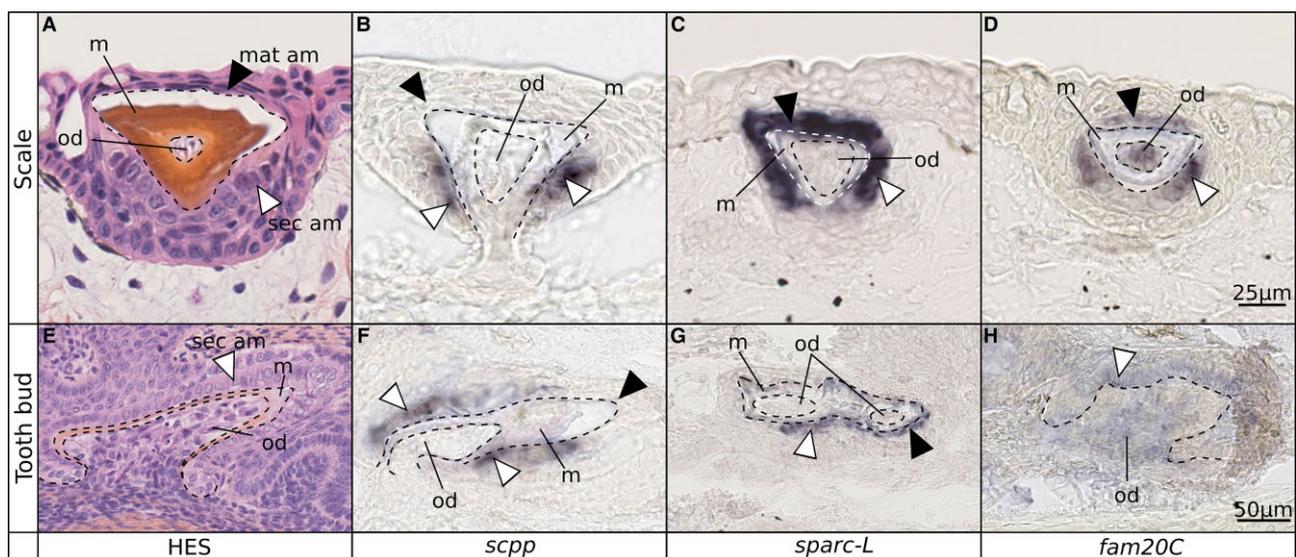
*Scpp* sequences can be readily aligned and compared (fig. 1, supplementary materials S2 and S3, Supplementary Material online). Notably, elasmobranch *Sparc-L* and *Scpp* proteins both contain a variable 8 amino acid-long periodic motif whose consensus is PDYXXPXX and that is absent from the elephant 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 <i>Dpm1</i> , zebrafish <i>spp1</i> )	++	++	+ (only for the elasmobranch <i>sparc-L</i> )	
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 *Dpm1* ([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. \(2013\)](#).

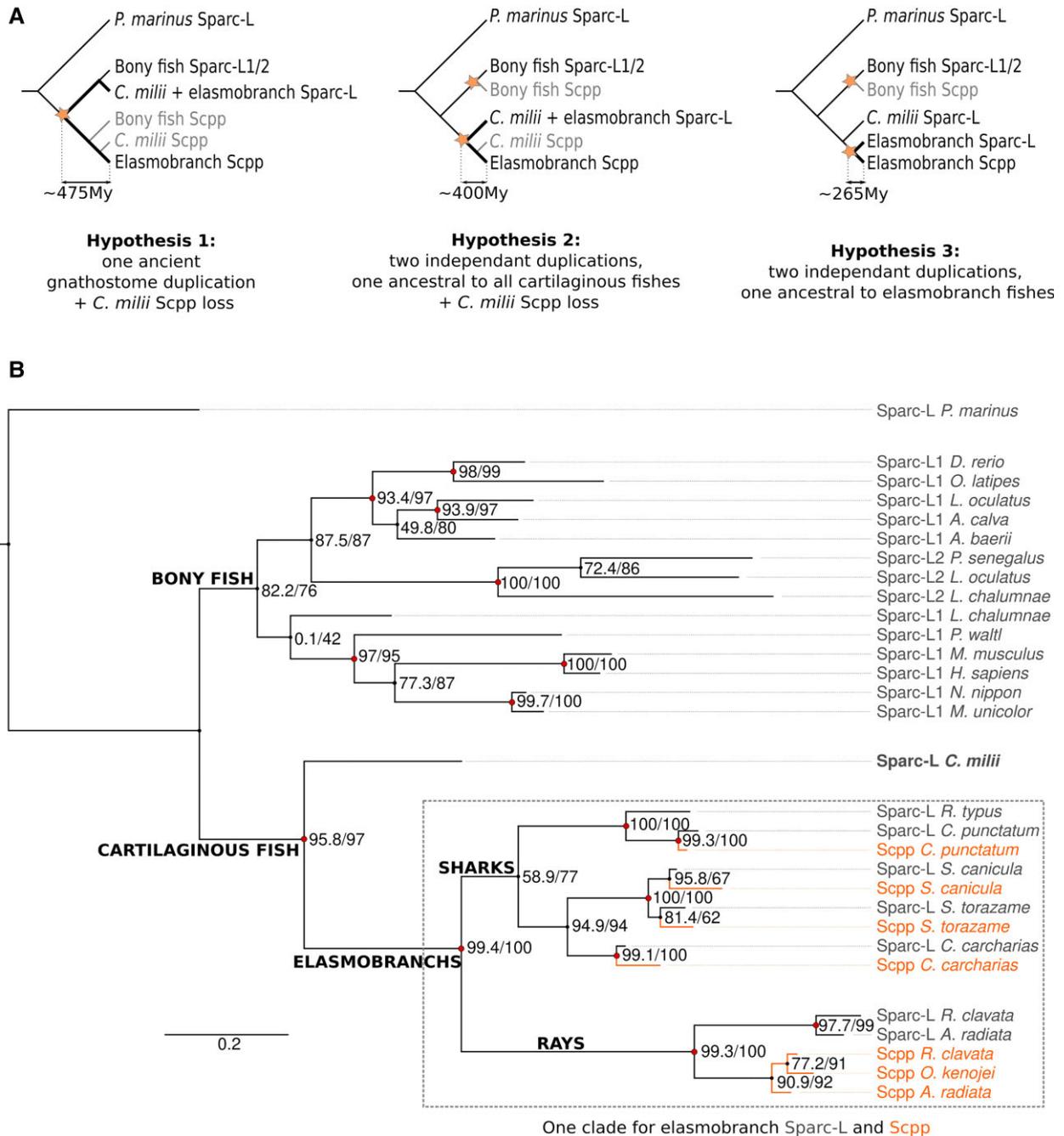


**Fig. 2.** The lesser spotted catshark *scpp* gene is coexpressed with *sparc-L* and *fam20C* in tooth and scale ameloblasts. Sections were performed at the level of developing scales (A–D, 8 cm-long embryos) and teeth (E–H, 9.5 cm-long embryos). Hematoxylin-Eosin-Safran histological staining (A and 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 *fam20C* (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. 3B](#)), a phenomenon reminiscent of concerted evolution due to non-allelic gene conversion between paralogs ([Zhou et al. 2019](#)). From this tree topology, it is impossible to deduce

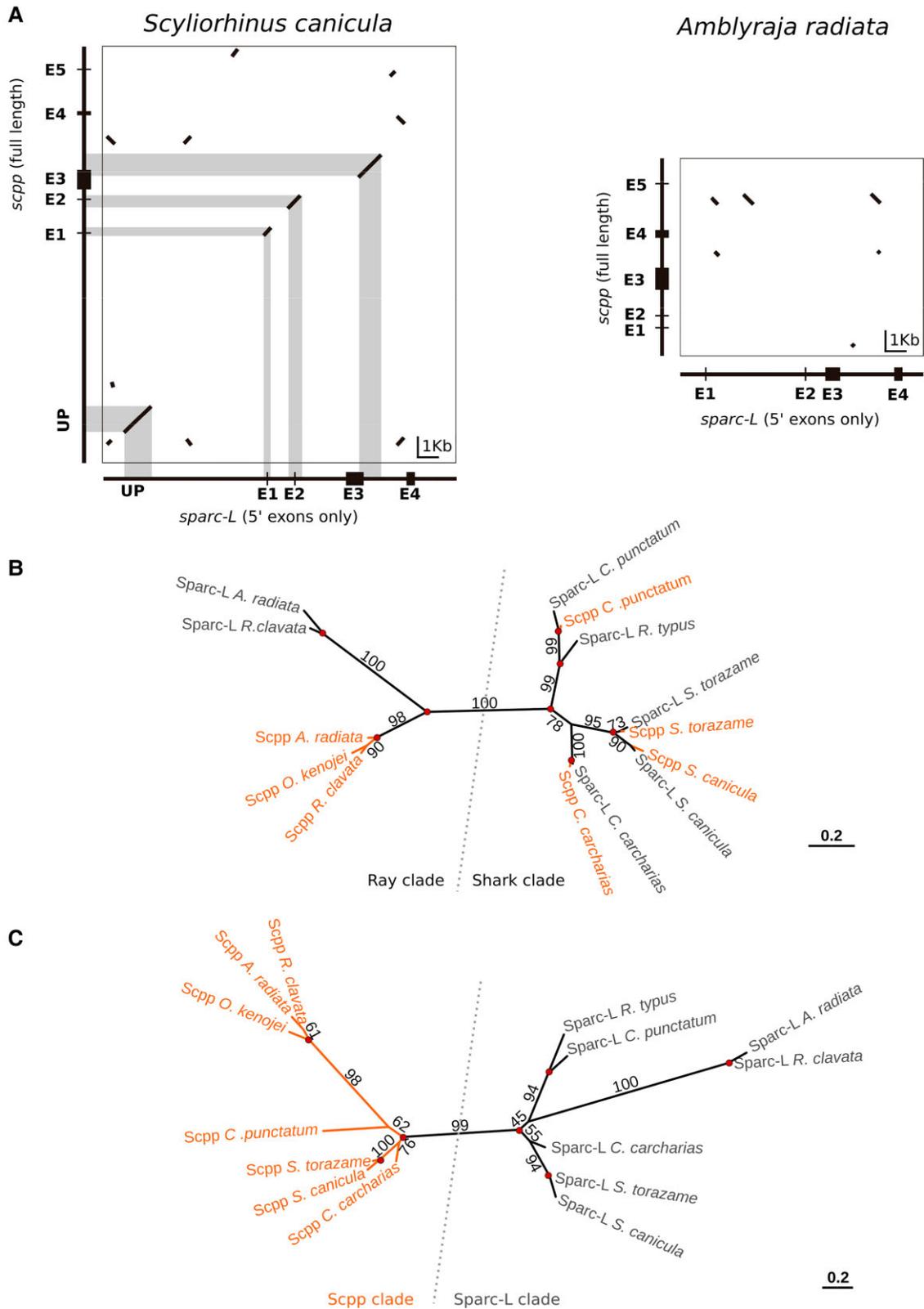
if the *scpp* gene evolved once at the base of the elasmobranchs, or if it results from independent duplications in selachimorphs and batoids. Performing a stringent genomic alignment reveals that gene conversion has occurred in four discrete genomic regions shared between the catshark *sparc-L* and *scpp* genes that encompass the coding exons 1 and 2, and overlaps with the end of the coding exon 3 ([fig. 4A](#), [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. 4B,

[supplementary material S7, Supplementary Material](#) online). By contrast, a phylogenetic analysis excluding the conversion regions the variable PDYXPXX 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).

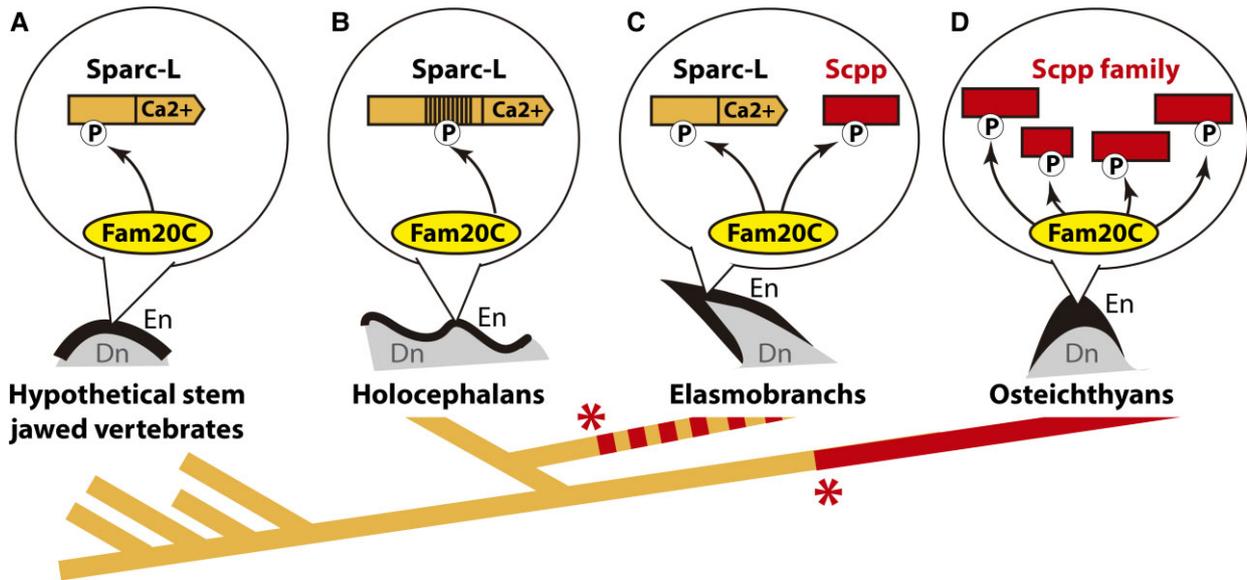
#### An Acidic N-terminal Domain and Fam20°C Phosphorylation Sites are Ancestral Features of Chondrichthyan Sparc-L Proteins

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 (Ca<sup>2+</sup>) 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 *Callorhynchus 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. 5A), and that remained largely unchanged in the holocephalan lineage (fig. 5B). Elasmobranchs and osteichthyans convergently evolved Scpp proteins that respectively function redundantly with Sparc-L (fig. 5C) or eventually took over the role ancestrally played by Sparc-L1/L2 (fig. 5D).

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, Vandeweghe 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, Vandeweghe 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 *Callorhynchus 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, Jermin 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’Hugin 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<sup>C</sup> 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 sialoprotein 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 galeaspid (Galeaspid, 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. *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, Bregenvold I, Kjølge K, Scavenius C, Petersen SV, Enghild JJ, Sørensen ES. 2020. FAM20<sup>C</sup> phosphorylation of the RGDSVVYGLR motif in osteopontin inhibits interaction with the  $\alpha$ v $\beta$ 3 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<sup>C</sup> 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, Blekman 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.

## 2.5 Discussion: Candidates genes and comparative analyses in Evo-Devo

Mineralization-associated candidate genes are the result of an initial comparative analysis between bony and cartilaginous fishes in an effort to explore gene expression in a non-model species, here the small-spotted catshark. This is a powerful method to describe gene repertoires (when the full gene family is considered) and expression patterns in major gnathostome lineages that might explain different skeletal features. As illustrated by the three articles presented in this chapter, there might be substantial variation in the gene repertoire in some families, but also divergence in the expression patterns of homologous genes.

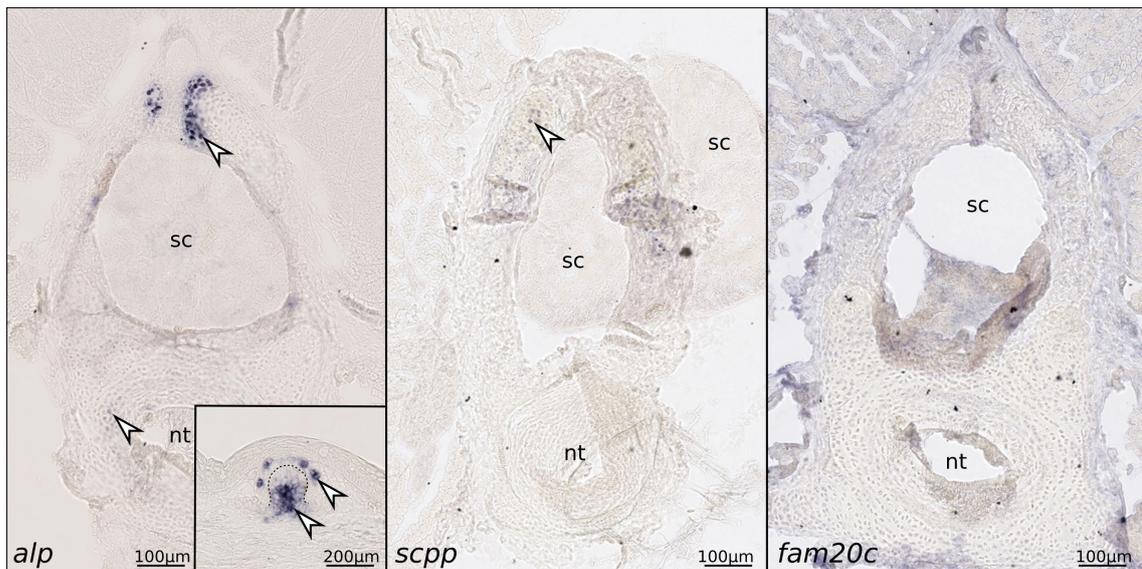
In the case of *mgp* and *bgp* each paralog underwent a specific duplication in different lineages: *mgp* in chondrichthyans and *bgp* in osteichthyans. Functionally, the bony fish *mgp* gene is associated with the inhibition of mineralization while the *bgp1* gene is associated with hydroxyapatite deposition. Furthermore, no descriptions of *bgp2* expression exist in the literature. In chondrichthyans, *mgp1* has a conserved expression pattern compared to bony fish *mgp*, while the *mgp2* paralog presented a completely different expression pattern in non-mineralizing tissues, probably as a consequence of neofunctionalization. Regarding the chondrichthyan *bgp* we did not find any congruent expression with the hypothesis of the activation of mineralization. These results suggest a gnathostome ancestral *mgp* gene whose expression pattern in mineralizing tissues may reflect a role in the inhibition of mineralization, while the ancestral *bgp* gene was either recruited in the bony fish mineralization toolkit or lost in the chondrichthyan toolkit.

We showed that chondrichthyans, teleosteans and tetrapods differ in the expression patterns for *sparc*, *sparc-L* and the *scpp* genes. First, *sparc* is expressed solely in mineralizing tissues of the endoskeleton of chondrichthyans. It has also been shown that it is expressed in bone and odontoblasts of both teleosteans and tetrapods (P. Holland et al., 1987; N. Li et al., 2009). This suggests that an ancestral *sparc* gene was already expressed in these mineralizing tissues before gnathostome diversification. Conversely, *scpps* show a shared expression pattern in pre-secretory ameloblasts in both elasmobranch and osteichthyans. Although both elasmobranch and osteichthyan *scpps* originated from duplications of the *sparc-L* gene, these events appear to have happened independently. The logical conclusion is that these genes are convergent and not inherited from the common ancestor of jawed vertebrates. Additionally, since the *scpp* genes in elasmobranchs and osteichthyans originated from independent duplications of their respective *sparc-L* and still kept similar functions, it suggests that the ancestral *sparc-L* gene already displayed an ancestral role in ameloblasts and therefore in mineralization. Therefore, the genes are convergent, but their expression patterns are ancestral because inherited from *sparc-L*.

These articles showcase various differences of gene repertoires and expression patterns that can be observed when comparing gnathostome lineages. We observed genes that kept the an-

central expression pattern (*mgp1* and *sparc*) and genes that acquired new functions altogether (*bgp1* and *mgp2*). Surprisingly, we even observed genes that originated from independent duplications of the same gene and still display convergent expression patterns (*scpps*). Understanding these dynamics is key to inferring the origins of the mineralization processes in vertebrates (cf Chapter 5).

Along with these articles I produced some additional data that is still preliminary and unpublished, but that will be useful for the following sections. This data concerns *in situ* hybridizations of *alp* in vertebrae and scales, and both *fam20c* and *scpp* in the vertebrae. The *alp* protein has been previously shown to be active at sites of mineralization in old embryos of the swell shark (*Cephaloscyllium ventriosum*) such as the surface of Meckel's cartilage and in the developing tooth (Eames et al., 2007). Additionally, the *in situ* hybridizations of this gene on transversal sections of 6.5cm TL *S. canicula* embryos (Fig 2.1) highlight that the expression of this gene is localized in sites of mineralization such as the centra, the neural arches and both compartments of the scales. In this figure additional ISH assays of *fam20C* and *scpp* extend their expression patterns to endoskeletal structures, since we have only seen them in odontodes in Chapter 1. While *fam20c* is clearly not detected in chondrocytes, *scpp* seems to be associated with mineralizing chondrocytes although this remain to be confirmed.



**Figure 2.1:** Expression patterns of *alp*, *fam20c* and *scpp* in 6.5cm TL *S. canicula* embryos. White arrowheads indicate sites of expression within the vertebrae or the scales (only for *alp*). The dotted line separates the epithelial and mesodermic compartments of the scale. nt: notochord, sc: spinal cord.

# Chapter 3

## The transcriptomic approach that was supposed to work better

### 3.1 Introduction

As mentioned above, when I started my PhD, only a partial and poorly assembled *Scyliorhinus canicula* genome was available, and RNAseq data was scarce. The only data I had was a *de novo* assembled transcriptome based on RNAseq data from a variety of labs around the world (including some unpublished data) resulting from a collaboration with Sylvie Mazan (Observatoire Océanologique Banyuls). This data however was not exploitable in and of itself for finding mineralization associated genes, since the raw data was not necessarily accessible nor useful (e.g. sampling done in unmineralized tissues such as liver, retina, etc...). This reference transcriptome therefore served as a replacement for the non-existing genome. This means that the main way I worked at first was through a candidate gene approach, using this reference to check whether I could find the orthologs of interest.

Candidate gene approaches are useful, although sometimes a studied gene might not be associated with mineralization in chondrichthyans when it is in bony fishes despite all the efforts to carry out the study. Most importantly I might be missing important innovations in chondrichthyans that have no link to what is known in bony fishes. Therefore, I had a strong eagerness to explore transcriptomics in an attempt to have an unbiased assessment of mineralization-associated genes in the small spotted catshark, and to increase the power of our analyses to more genes than what I had done previously. Indeed a transcriptomic approach could allow me to screen many potential mineralization genes, while the candidate gene was more of a one by one step-wise approach.

When I started to work on transcriptomic data generated in the lab (reads produced in early 2020), my main reference was the transcriptome assembled by H el ene Mayeur, in the team of Sylvie Mazan, so all the analyses produced were dependent on this reference (read mapping, see below). However, the small-spotted catshark species was then chosen for genome and RNA

sequencing as part of the commemoration of the 25 years of the Sanger Institute (december 2019). The institute produced a chromosome-level assembly of the genome of my model species as well as an improved reference transcriptome based on fifty RNAseq libraries including major tissues of every organ system (received in may 2020). Therefore, all my transcriptomic data and analyses could benefit from these new references, and I adjusted all the work I did onto this new and improved data. However, very quickly, Sylvie Mazan and her collaborators created yet another reference, this time a gene model (Mayeur et al., 2021), based on the old transcriptome, Sanger Institute’s new reference and NCBI predictions and annotations. Accordingly, in an effort to have the best quality reference, I overhauled my previous analyses to match this latest most complete reference.

In the following, I will be explaining how I used these different data sets (both our locally sequenced RNAseq, as well as Sanger Institute’s RNAseq and the latest gene model from Sylvie Mazan) to try and isolate mineralization-associated genes in *Scyliorhinus canicula*. The idea was to find differentially expressed genes between young and old embryonic vertebrae to isolate the mineralization-associated ones. Although this straight forward approach did not seem to work at first, I used the Sanger Institute RNAseq libraries to define putative skeletal markers and check for differential expression only in this subset. Below, I will present the bio-informatic methods only for the latest version of this analysis, which is still ongoing and unpublished. Finally, a more detailed study of these genes was performed using phylogenies, mRNA *in situ* hybridizations and comparative analyses with functional data on bony fishes.

## 3.2 Materials and methods

Step-wise procedure for all materials and methods are illustrated in Fig 3.1

### 3.2.1 Transcriptomic data generated by the Sanger’s Institute

*S. canicula* tissue libraries from whole embryos at stages 12, 22, 24, 26, 30 and 31 (Ballard et al., 1993) and adult tissues sampled from three individuals (BioSample accessions: 21694013 to 21694062) as following: “Tissue (number of biological replicates)”

Ampullae of Lorenzini (2), Blood (2), Brain+Olfactory sac (3), Chondrocranium (3), Dermal denticles (3), Dental lamina (3), Esophagus (1), Eye (2), Gills (1), Heart (1), Kidney (1), Liver (1), Meckel’s cartilage (3), Muscle (2), Ovary (1), Pancreas (2), Rectal gland (2), Spinal cord (1), Spiral intestine (2), Spleen (1), Stomach (1), Seminal vesicle (1), Testis (2), Uterus (1) and Vertebrae (2).

The paired-end RNAseq reads from adult and embryonic *S. canicula* tissue samples were mapped onto the reference gene model (Mayeur et al., 2021) using BWA MEM (H. Li, 2013). After converting the resulting alignments to BAM files using samtools, mapped reads were filtered (Flag 0x904, MAPQ  $\geq$  3) and sorted in order to extract counts for each library. Count

data was also normalized as Transcripts Per Kilobase Millions (TPMs). When a tissue had several replicates (max 3 replicates), TPM values were averaged, leading to a final number of 31 tissues compared. For each gene, Z-scores were then calculated to evaluate tissue-biased expression as

$$Z = (x - \mu) \cdot \frac{\sqrt{n}}{\sigma}$$

Where for any gene:  $x$  is the observed TPM value in a given tissue,  $\mu$  is the mean of all tissues,  $\sigma$  is the standard deviation of all tissues and  $n$  is the number of tissues.

For all subsequent analyses only genes which are over-represented in skeletal tissues and under-represented in skeletal muscle were used. Under-representation in muscle was used as it is the main contaminant tissue in our samples. This filtering step was performed by selecting genes with Z-scores  $\geq 4$  in all three skeletal tissues and with Z-scores  $\leq 1$  in skeletal muscle. This results in a list of 681 genes, defined as reference skeletal markers.

### 3.2.2 Mineralization-focused RNAseq data

The delineation of our RNA sequencing design was based on current knowledge of catshark endoskeleton mineralization. Three main things were relevant in our decision making: the mineralization of *Scyliorhinus canicula* vertebrae (i) begins in neural arches when embryos reach  $\approx 6.5$ cm total length (TL), (ii) at  $\approx 8$ cm TL all structures meant to mineralize already have and (iii) mineralization occurs first in anterior vertebral centra followed by posterior vertebrae (Enault et al., 2016). Following this, anterior and posterior vertebrae of younger (4.8, 4.9, 5.3cm) and older (8.0, 8.3, 8.5cm) *S. canicula* male embryos were dissected and isolated for RNAseq library preparation. Twelve RNAseq libraries were prepared using TrueSeq mRNA library preparation kit (Illumina, San Diego, California, USA) with 1  $\mu$ g scale total RNA per library. All RNA had an RNA integrity number above 7.6 as determined with a Bioanalyser 2100 (Agilent Technologies, Santa Clara, California, USA). Single end (100 bp) sequencing of RNAseq libraries was performed using Illumina NovaSeq 6000 by MGX. I received and analyzed this raw data. Single end RNAseq reads were mapped and counts extracted exactly as above. Count data was also normalized as Transcripts Per Kilobase Millions (TPMs)

Embryonic vertebrae counts data was kept only for the 681 putative skeletal markers identified above. Additionally, as dissection and isolation of vertebrae was done manually under the stereomicroscope, perfect isolation of embryonic skeletal elements is virtually impossible. To account for differences in cartilage proportion due to size in our samples, a normalization by a known cartilage marker was performed. Counts for each gene, in each sample, were thus divided by (*col2a1* counts/10000) and then rounded up into integers. The resulting data set was subsequently used to perform a differential expression analysis using DESeq2 R package (Love et al., 2014), contrasting young and old embryos. This allows for the identification of up-regulated genes (i.e. stronger values) in older embryos, possibly associated with mineralization.

Differentially Expressed Genes (DEGs) were identified using standard DESeq2 parameters, but using an adaptive shrinkage estimator (Stephens, 2017). Finally, gene names of the data sets before and after screening for skeletal markers were used as input for enrichment analyses using *enrichr* (E. Y. Chen et al., 2013) and the Jensen tissues database (Palasca et al., 2018).

### 3.2.3 mRNA *in situ* Hybridization

*In situ* hybridizations were performed on 14  $\mu\text{m}$  thick cryostat sections of 6,5 cm TL *S. canicula* embryos, cut transversely in the body trunk, at the level of the pectoral fins. All subsequent procedures were previously described (Leurs et al., 2021; see journal article of chapter 1). Slides were scanned on a Hamamatsu nanozoomer. Primers designed to generate the DNA matrix used for RNA probe synthesis are available on demand.

### 3.2.4 Phylogenetic reconstructions

Protein sequences sampled from vertebrate genomes were used to reconstruct phylogenies. To generate each phylogenetic reconstruction, selected sequences were aligned with MAFFT (Kato and Standley, 2013; standard parameters and `-auto` strategy). Because a large proportion of the sequences is predicted from genomes and may include false exons, alignments were then cleaned with HmmCleaner (Di Franco et al., 2019) to remove low similarity segments. Phylogenies were then inferred by Maximum Likelihood via IQ-Tree (version 1.6.1; Nguyen et al., 2014). The best model of amino-acid evolution was selected using ModelFinder (Kalyaanamoorthy et al., 2017) and the Bayesian Information Criterion (Schwarz, 1978). Node support was estimated by performing a thousand ultra-fast bootstrap replicates (UFBoot; Hoang et al., 2017), and single branch tests (SH-aLRT; Guindon et al., 2010). Models and alignment lengths are indicated in figure legends, and full alignments are available on demand.

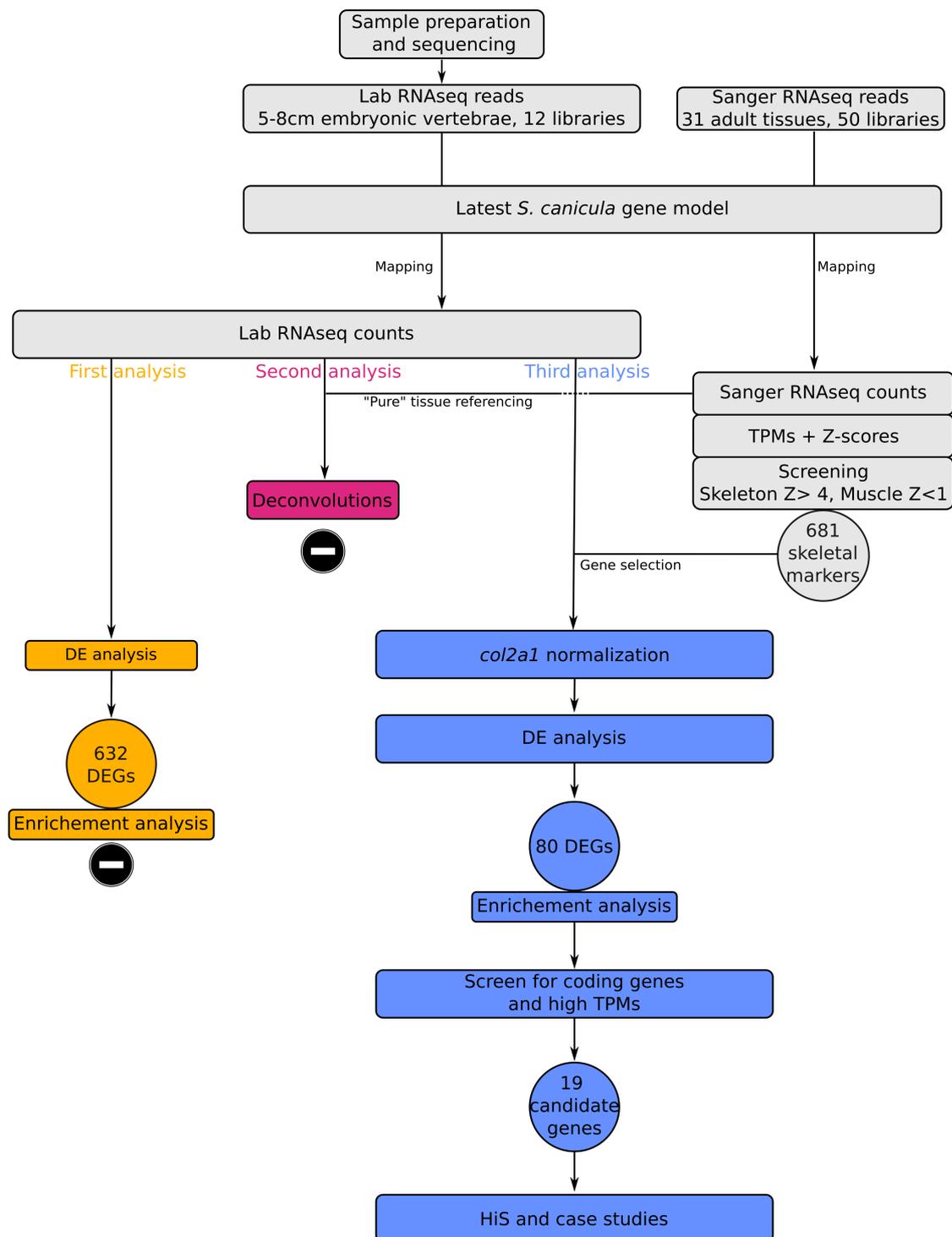
### 3.2.5 Synteny analyses

Synteny was explored in seven reference genomes available on NCBI: *M. musculus* (GCF\_00000-1635.27), *S. canicula* (GCF\_902713615.1), *Lepisosteus oculatus* (GCF\_000242695.1), *D. rerio* (GCF\_000002035.6) and *C. milii* (GCF\_000165045.1), *Carcharodon carcharias* (GCF\_01763951-5.1) and *Amblyraja radiata* (GCF\_010909765.2).

## 3.3 Results and Discussion

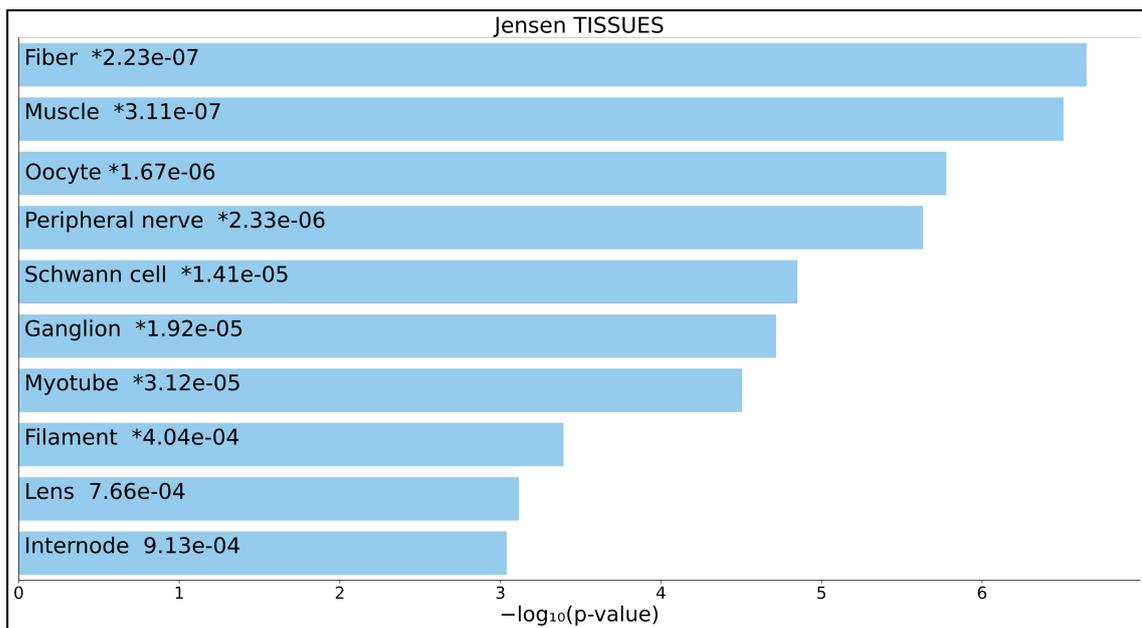
### 3.3.1 Bulk vertebra RNAseq characterization

To discover novel mineralization associated genes in the small-spotted catshark a differential gene expression analysis was performed between RNAseq data of young embryos vertebra



**Figure 3.1:** Graphical representation of the step-wise procedure for RNAseq analysis

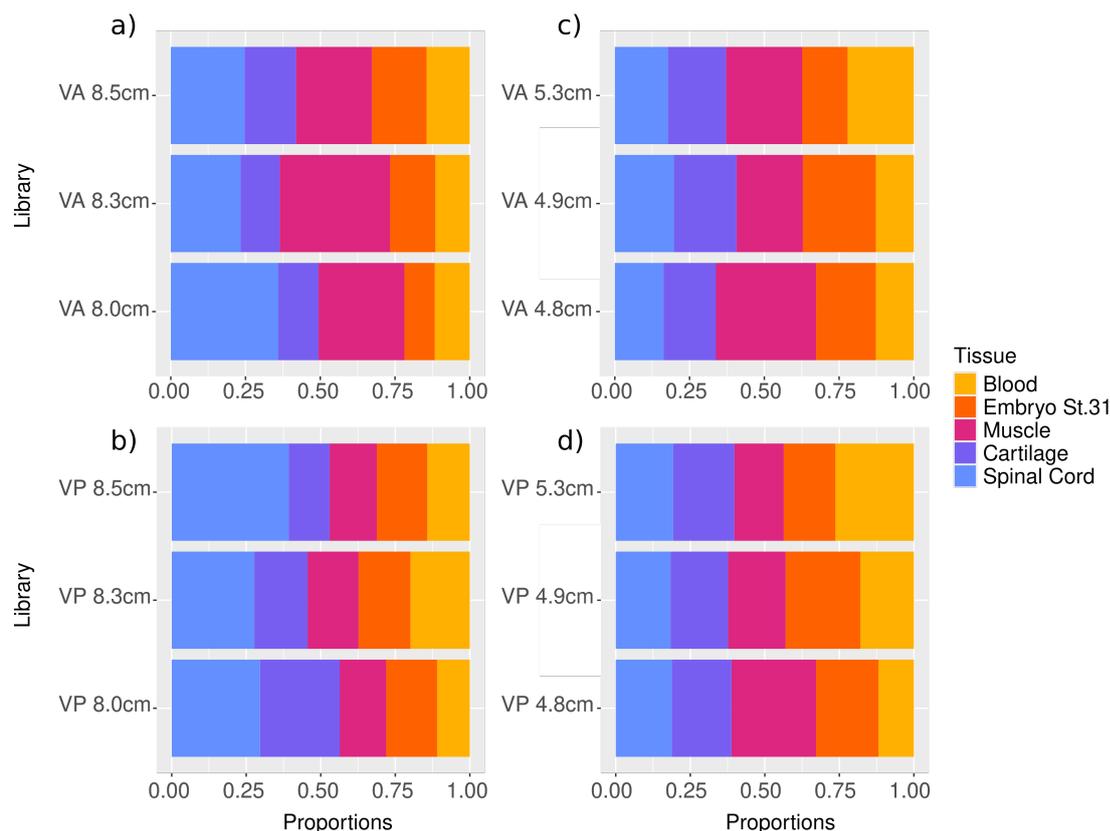
(before the onset of mineralization) and old embryos vertebra (all structures undergoing mineralization). This analysis uncovered 632 differentially expressed genes (DEGs), upregulated in older embryos and presenting a  $\log_2$  fold-change  $\geq 2$  (i.e. 4 fold the expression levels). An enrichment analysis of the annotated DEGs (Fig 3.2) highlights an unexpected result with most GO-terms associated with muscle or nervous tissues. It is noteworthy to mention that many of the genes in *Scyliorhinus canicula* are unannotated (387 out of 632  $\approx 61\%$ ) and that the ontogeny predictions are based on mammal functions, therefore these results should be interpreted with care. Our first examination of the DE analysis therefore seemed blurred by the presence of the muscle and spinal cord tissue that show strong differences in gene expression.



**Figure 3.2:** Bar chart of top enriched terms for the 632 up-regulated genes from young to old embryonic vertebrae ( $\log_2$  fold-change  $\geq 2$ ). Data was extracted from the GO Biological Process 2021 gene set library. The top 10 enriched terms for the input gene set are displayed based on the  $-\log_{10}(\text{p-value})$ , with the actual p-value shown next to each term. The term at the top has the most significant overlap with the input query gene set.

The vertebra used for our RNA sequencing are relatively small ( $\approx 2\text{mm}$ ), which makes complete isolation of the vertebral element from its surrounding muscle tissue virtually impossible by hand. Moreover, the vertebra has neural arches that protect the spinal chord, another tissue that is hard to remove especially since it is not very cohesive and breaks down easily. Similarly, in posterior vertebrae veins and arteries are found in the hemal canals. These tissues are all potentially masking the signal associated with mineralization. However, the results of the enrichment analysis indicate that the main signal being captured, is one of muscle and not skeletal development.

Sequenced samples such as ours are commonly referred to as bulk RNAseq data (i.e. a mix a several tissues in RNAseq) and there are computational methods to try and interpret these data sets, notably one called deconvolution (Avila Cobos et al., 2020). This method in RNAseq consists in estimating the proportions of each tissue in the bulk data using Additional



**Figure 3.3:** Graphical representation of estimated tissue proportions of within embryonic vertebra libraries, using CDSeq (Kang et al., 2019) and Sanger Institute’s RNAseq data of Blood, Embryo Stage 31, Muscle, Cartilage and Spinal Cord. (a) Estimations on old embryos anterior vertebrae, (b) Estimations on old embryos posterior vertebrae, (c) Estimations on young embryos anterior vertebrae, (d) Estimations on young embryos posterior vertebrae. VA: vertebrae, anterior; VP: vertebrae, posterior

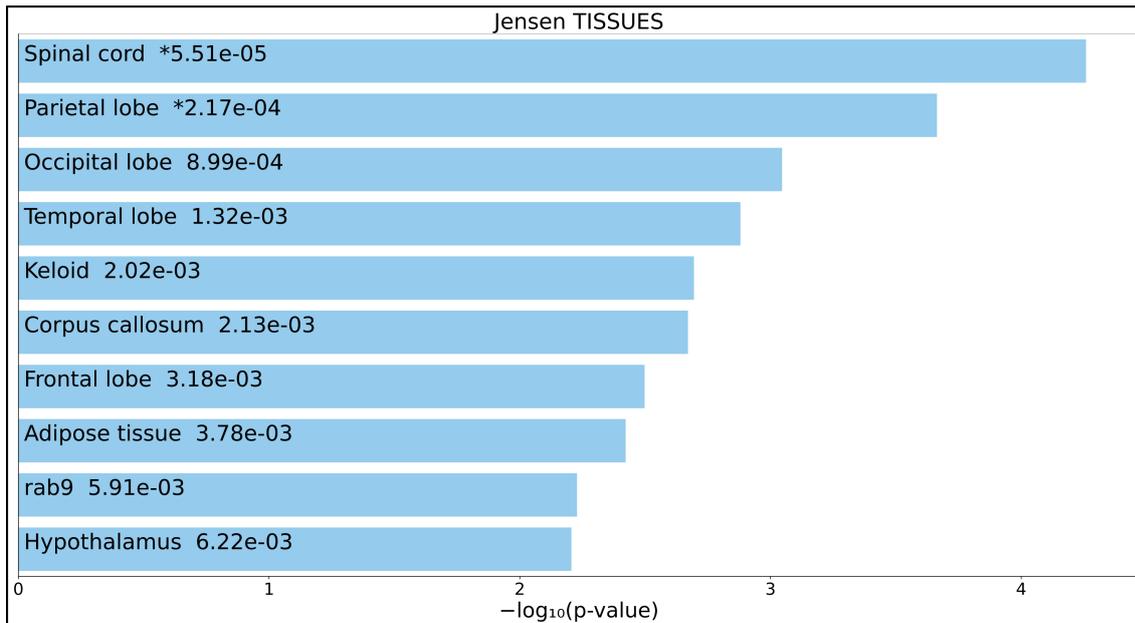
RNAseq data from unique cell lineages (i.e. single-cell RNAseq). There are several ways to deconvolute bulk RNA seq data, the most intuitive being identifying many cell-specific markers (e.g. myocyte markers, neuron markers and chondrocyte markers) such that when the expression profiles are compared to the bulk RNAseq, cell-type proportions can be estimated. However, single cell RNAseq data for the small spotted catshark was not available, and instead only “pure” tissue RNAseq data sets from Sanger Institute were available. We decided, after discussing with Marie Sémon during my PhD committee meeting, to try this method using these fifty Sanger libraries, and estimate tissue proportions instead of cell proportions. However, the method would either estimate unreasonable tissue proportions (e.g.  $\geq 25\%$  of spinal chord in older embryos; Fig 3.3a, b) or estimate close to equal proportions for each tissue, which might be the result of random assignment (e.g. 20% of each tissue Fig 3.3c, d). Additionally, the method would estimate any requested tissue, even when it should not be present in the sample with non-negligible proportions (e.g. 10% of uterus; not shown). The main reason this did not work was that the available data was not optimal for the deconvolution method. Indeed, a “pure” tissue is not realistic since cartilage for instance, is made up of several cell-types. Additionally,

the data was hardly comparable: our RNAseq data was mostly embryonic vertebrae while the Sanger data was mainly adult tissue or whole embryonic stages. Moreover, the Sanger data had a much greater sequencing depth than ours which can highly influence the number of counts and complicate comparisons. Lastly, some of the Sanger data may not have been as pure as previously thought (one chondrocranium library was contaminated with brain tissue, not shown). Therefore, comparing tissues with intrinsically different molecular fingerprints (embryo *vs* adults), and libraries with hugely different sequencing depths was misleading.

### 3.3.2 Sorting RNAseq data to extract skeleton-relevant data

In order to overcome the problem of mixed tissues and of various tissue proportions, we adapted some elements of the deconvolution methods. Using the fifty sanger libraries, count data was normalized as TPMs and Z-scores were calculated. Here, Z-scores are a measure of over or under-representation of expression of a gene in a tissue relative to the mean of expression of that gene in all tissues (so that the sum of all the Z-scores for one gene = 0). Therefore through careful screening of the data through Z-scores, skeleton-specific markers can be isolated. Here only genes with a Z-score  $\geq 4$  in all three skeletal tissues (chondrocranium, Meckel's cartilage and vertebra) all the while having a Z-score  $\leq 1$  in the muscle (the major contaminant in our samples) were kept. This produced a list of 681 genes, considered as reference skeletal markers. Despite the small number of reference genes, the result is not so strongly biased toward "skeleton only" from a Z-score  $\geq 4$  (and see Fig 3.5). Our own RNAseq data was then simplified to only this list of genes and analyzed through differential expression analysis. However, one last element had to be taken into account. Comparing small embryos *vs* big ones is intrinsically flawed since tissue proportions vary through development, specifically here in terms of muscle and cartilage. Therefore a higher number of cartilage cells expressing a given gene in older embryos compared to younger ones will generate higher TPM values that might bias the analysis by masking or promoting a signal because of cell proliferation or variable proportion of a tissue compared to another. This is one of the most prominent issues in transcriptomics and is nicely reviewed in Pantalacci and Sémon, 2015. To solve this, the cross-referenced genes were normalized by a known chondrocyte-specific marker : *col2a1*.

Differential expression analysis of this data set with a log2 fold-change of two, resulted in only eleven DEGs. Considering that our signal might not be as strong as previously thought, genes with a log2 fold-change of one (= two-fold) instead of two, were kept which this time yielded eighty DEGs up regulated in older embryos compared to younger ones. An enrichment analysis of these DEGs (Fig 3.4) shows mostly nervous tissue enrichment but not expected GOterms associated with the skeleton. This, however, might be explained by two main reasons. First, only 50% of these genes are correctly annotated, which means that only forty DEGs are taken into account for the enrichment analysis and may not be representative of the up regulated genes. Second, GOterms are based on functions in mammals, which does not necessarily



**Figure 3.4:** Bar chart of top enriched terms for the 80 up-regulated genes from young to old embryonic vertebrae ( $\text{Log}_2$  fold-change  $\geq 1$ ). Data was extracted from the GO Biological Process 2021 gene set library. The top 10 enriched terms for the input gene set are displayed based on the  $-\log_{10}(\text{p-value})$ , with the actual p-value shown next to each term. The term at the top has the most significant overlap with the input query gene set.

mean that a gene involved in e.g. *corpus callosum* in these models is not involved in cartilage development and mineralization in the catshark. Overall, GO terms are very generic and only allow for a global view of the genes putative functions, when these are correctly annotated.

### 3.3.3 Listing candidate genes

Despite limited enrichment results, eighty DEGs is a manageable number to go through individually, in order to assess their viability as future candidate genes. This assessment was done in several ways. First an assessment of TPM values in our data and their patterns across adult tissues (Sanger institute data) was made to make sure DEGs are highly expressed in embryonic vertebra ( $\text{TPM} \geq 50$ ) and that they show expression patterns with a skew towards skeletal tissues in adults. In addition, BLASTs were used to investigate unannotated genes as well as to check whether annotated ones seem correct. This step can reveal some unannotated genes that, for instance, do not translate into credibly functional proteins making these genes non-coding RNA genes (here 25/80 genes = 31.25%; Table 3.1 and 3.2). Finally, genes with higher TPM values in old embryos will be preferred for the final step: *in situ* mRNA hybridization (ISH), which will localize the expression of these genes, potentially in mineralizing tissues.

ID	Annotation	Log2 fold-change	Mean TPM	Old embryos	TPMs values for ISH	Expression pattern	Blast results	Cloning	ISH
XM_038790375.1	<i>spp2</i>	1,88		205,9719135	Ok	Skeleton skewed	Ok	Ok	Yes
XM_038816264.1	<i>otos</i>	1,28		202,2033045	Ok	Skeleton skewed	Ok	Ok	Yes
XM_038788185.1	<i>spp2</i>	3,31		121,7259071	Ok	Skeleton skewed	Ok	Ok	Yes
XM_038788193.1	<i>spp2</i>	2,62		52,82523345	Ok	Skeleton skewed	Ok	Ok	Yes
XM_038796173.1	<i>ctgf</i>	2,04		52,07531215	Ok	Skeleton skewed	Ok	Ok	Yes
XM_038788196.1	<i>spp2</i>	1,72		51,91032016	Ok	Skeleton skewed	Ok	Ok	Yes
XM_038806248.1	<i>clec3a</i>	1,26		51,90776159	Ok	Skeleton skewed	Ok	Ok	Yes
XM_038792863.1	<i>unannotated</i>	3,32		47,14204106	Ok	Skeleton skewed	<i>scpp</i>	Already done	Yes (Chapter 1)
XM_038786269.1	<i>mia</i>	1,48		42,51696394	Ok	Skeleton skewed	Ok	did not work	-
XM_038781460.1	<i>coch</i>	3,07		17,84650883	Bit low	Skeleton skewed	Ok	Ok	Yes
XM_038795263.1	<i>cytl1</i>	1,1		4,729358271	Too low	Skeleton skewed	Ok	Ok	Yes
XM_038783991.1	<i>unannotated</i>	2,21		51,24566879	Ok	Bimodal	<i>bqp</i>	Already done	Yes (Chapter 1)
XM_038799486.1	<i>snap91</i>	1,36		34,10667272	Ok	Bimodal	Ok	Ok	No
Cluster-18614.1	<i>tgfb2</i>	1,41		17,246013	Bit low	Bimodal	Ok	Ok	Yes
XM_038784999.1	<i>unannotated</i>	3,43		12,78171187	Bit low	Bimodal	uncharacterized protein	Ok	Yes
XM_038782299.1	<i>ccdc9</i>	1,05		3,917049332	Too low	Bimodal	Ok	Ok	No
XM_038778800.1	<i>ptgir</i>	1,56		1,144287275	Too low	Bimodal	Ok	Ok	No
XM_038805081.1	<i>reep5</i>	1,36		558,6641766	Ok	Ubiquitous	Ok	Ok	No
XM_038787500.1	<i>cyl</i>	1,88		124,58665345	Ok	Ubiquitous	Ok	Ok	Yes
XM_038799089.1	<i>borg2</i>	1,23		101,7150357	Ok	Ubiquitous	Ok	Ok	Yes
XM_038796236.1	<i>sympo</i>	1,06		53,72228179	Ok	Ubiquitous	Ok	Ok	Yes
XM_038822098.1	<i>bitbd6</i>	2		52,36651953	Ok	Ubiquitous	Ok	Ok	Yes
XM_038774347.1	<i>mtmr1</i>	1,03		39,5321349	Ok	Ubiquitous	Ok	Ok	Yes
XM_038811488.1	<i>sema3c</i>	1,02		38,55884905	Ok	Ubiquitous	Ok	Ok	No
XM_038798203.1	<i>rab18</i>	1,16		36,48158257	Ok	Ubiquitous	Ok	Ok	No
Cluster-536.0	<i>asb8</i>	1,09		33,64217948	Ok	Ubiquitous	Ok	Ok	No
XM_038815678.1	<i>bcl6</i>	1,21		32,56334407	Ok	Ubiquitous	Ok	Ok	No
XM_038815745.1	<i>rxrba</i>	1,3		32,3078891	Ok	Ubiquitous	Ok	Ok	No

Table 3.1: List of screened DE genes.

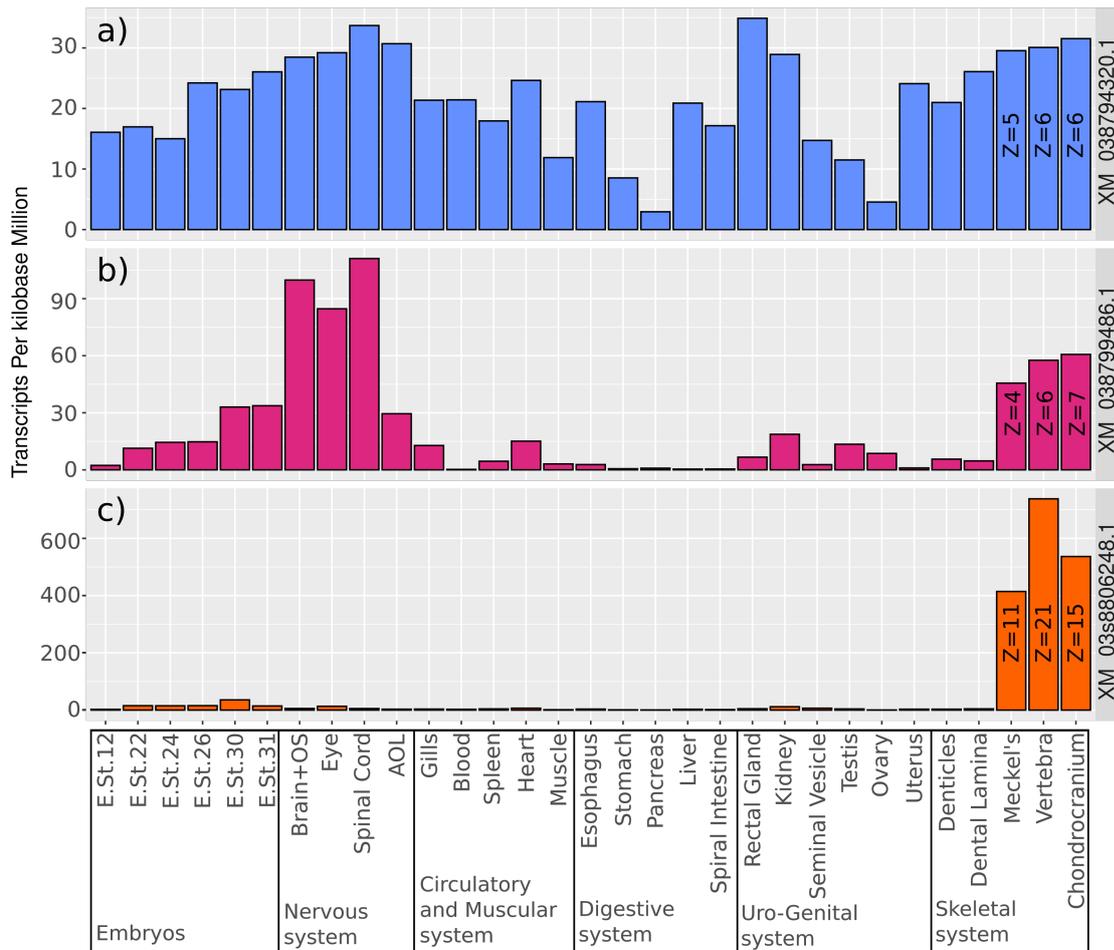
ID	Annotation	Log2 fold-change	Mean TPM	Old embryos	TPMs values for ISH	Expression pattern	Blast results	Cloning	ISH
XM_038799819.1	<i>rlpl1</i>	1,12	28,70145398	Bit low	Ubiquitous	Ok	Ok	No	
XM_038821778.1	<i>c1orf174</i>	1,42	23,39617525	Bit low	Ubiquitous	Ok	Ok	No	
XM_038794320.1	<i>tm2d1</i>	1,02	19,02203063	Bit low	Ubiquitous	Ok	Ok	No	
XM_038809303.1	<i>cebpd</i>	1,4	17,44588624	Bit low	Ubiquitous	Ok	Ok	No	
XM_038813744.1	<i>unannotated</i>	1,06	15,82417287	Bit low	Ubiquitous	uncharacterized protein	Ok	No	
XM_038789767.1	<i>mlkip</i>	1,34	13,54559011	Bit low	Ubiquitous	Ok	Ok	No	
XM_038822144.1	<i>nbl1</i>	1,12	13,53801963	Bit low	Ubiquitous	Ok	Ok	No	
XM_038820050.1	<i>ptbf2</i>	1,06	12,0177971	Bit low	Ubiquitous	Ok	Ok	No	
XM_038812583.1	<i>nisch</i>	1,04	11,81286612	Bit low	Ubiquitous	Ok	Ok	No	
XM_038790999.1	<i>rnf12</i>	1,35	11,0272078	Bit low	Ubiquitous	Ok	Ok	No	
XM_038781248.1	<i>unannotated</i>	1,56	10,70144887	Bit low	Ubiquitous	<i>ppp1</i>	No	No	
XM_038815145.1	<i>aqfl1</i>	1,44	9,749610938	Too low	Ubiquitous	Ok	Ok	No	
XM_038797347.1	<i>nr1d2</i>	1	8,839555105	Too low	Ubiquitous	Ok	Ok	No	
XM_038806253.1	<i>maf</i>	1,11	8,058465221	Too low	Ubiquitous	Ok	Ok	No	
XM_038816355.1	<i>tnk2</i>	1,27	7,596116247	Too low	Ubiquitous	Ok	Ok	No	
XM_038819724.1	<i>cdip1</i>	1,01	7,392861298	Too low	Ubiquitous	Ok	Ok	No	
XM_038822057.1	<i>lgi1</i>	1,03	7,128270934	Too low	Ubiquitous	Ok	Ok	No	
XM_038812531.1	<i>itih3</i>	1,03	6,391719835	Too low	Ubiquitous	Ok	Ok	No	
XM_038819963.1	<i>gaa</i>	1,57	6,28270317	Too low	Ubiquitous	Ok	Ok	No	
Cluster-9325_3	<i>unannotated</i>	1,43	6,149048346	Too low	Ubiquitous	Hypothetical protein (Scyllorhinus torazame)	Ok	No	
XM_038792554.1	<i>hmg20a</i>	1,25	5,725724473	Too low	Ubiquitous	Ok	Ok	No	
XM_038787076.1	<i>rfr7</i>	1	3,904908211	Too low	Ubiquitous	Ok	Ok	No	
XM_038813915.1	<i>clip4</i>	1,29	3,199384887	Too low	Ubiquitous	Ok	Ok	No	
XM_038777944.1	<i>fam20a</i>	1,19	3,142411254	Too low	Ubiquitous	Ok	Ok	No	
XM_038822232.1	<i>cnrm2</i>	1,21	1,530595884	Too low	Ubiquitous	Ok	Ok	No	
XM_038775484.1	<i>hsf1</i>	2	1,503658262	Too low	Ubiquitous	Ok	Ok	No	
XM_038814858.1	<i>edit3</i>	2	34,74816387	Ok	Ubiquitous	Ok	Ok	No	
<b>25 genes</b>	<b>no translation</b>								

Table 3.2: List of screened DE genes.

Assessment of TPM and fold-change values is an important step since several DEGs can present strong fold-changes as a consequence of low TPM values. For instance, a gene having one TPM in young embryos and four in old embryos will show as a DEG having a four-fold increase in expression and come out as a significant result. However, although this result can be seen as an artifact of the analysis, it might also be the result of a small cellular population whose gene expression is diluted in a bigger cellular mass. Finally, an overall view of the gene expression in adult tissues can point towards a specific putative function of the genes. Several cases of gene expression patterns in the Sanger institute's RNAseq data can be identified (Fig 3.5). For example, several genes had quite ubiquitous expression patterns (38/80 genes = 47.5%, Table 3.1, 3.2 and Fig 3.5a) while others had a more bi-modal expression pattern with an expected specificity in skeletal tissues (6/80 genes = 7.5%, Table 3.1, 3.2 and Fig 3.5b). Lastly, some genes had a very skeletal-specific expression patterns with expression detected almost exclusively in skeletal tissues (11/80 genes = 13.75%, Table 3.1, 3.2 and Fig 3.5c). The genes having patterns such as in Fig 3.5c were chosen as the main candidates for this study although some genes having patterns such as in Fig 3.5b and c were also considered (in orange in table 3.1).

Based on these features a total of nineteen genes were studied, presenting all three types of TPM expression patterns (ubiquitous, skeleton skewed and bimodal; Fig 3.5). An encouraging result is to find a previously studied gene: *scpp* which has been presented in Chapter 1 and seems to be expressed in old embryonic vertebrae. This shows that the procedure to isolate some mineralization associated genes did work to some degree. However, another gene presented in Chapter 1: *bgp* shown to be expressed in nervous tissues was also found in this analysis. This indicates a clear contamination of the spinal cord or dorsal ganglia in the vertebral dataset of both our embryonic samples and the adults samples sent to the Sanger Institute. Nonetheless, another interesting result is the presence of four genes, all annotated as *spp2*. This probably means that these are genes of the same family but that do not have a 1:1 orthology relationship to bony fishes hindering correct annotation. These genes were examined through ISH in order to investigate their potential association to mineralized tissues. However, the *spp2* genes turn out to have very similar sequences, therefore primers were defined so that they would amplify all the *spp2* sequences simultaneously. Finally, *mia* could not be successfully cloned for probe synthesis (Table 3.1).

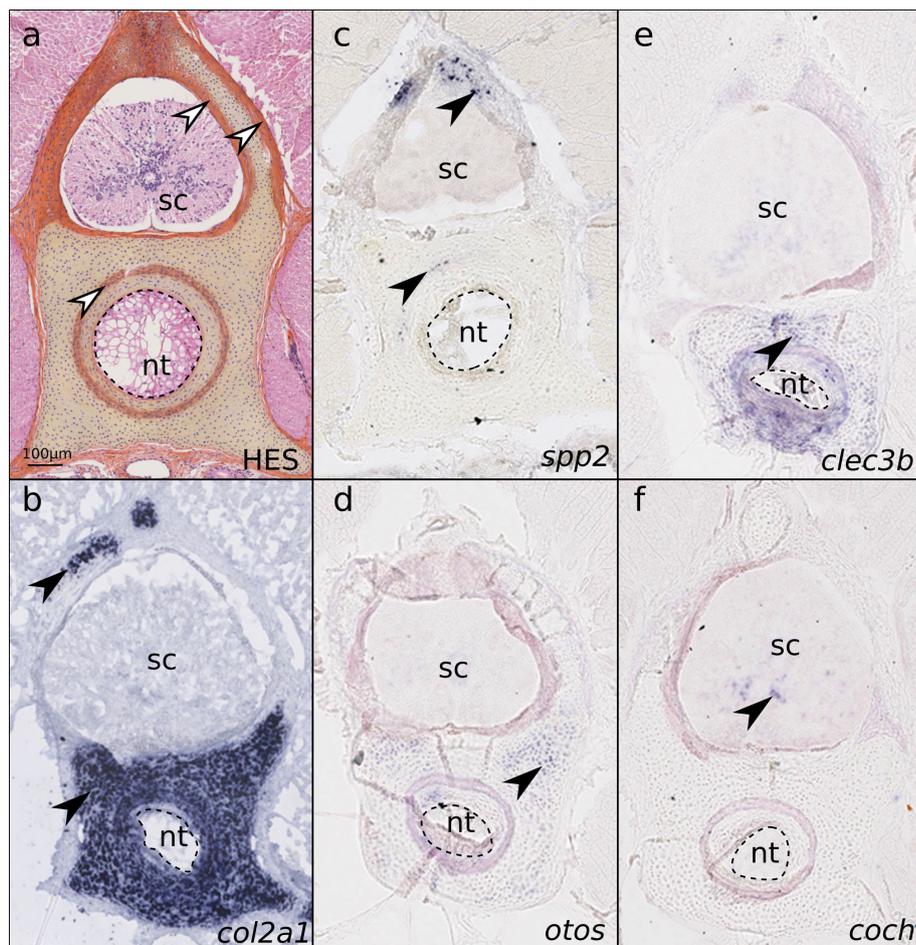
Therefore, excluding *bgp*, *scpp*, *mia* and reducing the four *spp2* genes to only one assay, result in a total of thirteen genes studied through ISH. Fig 3.6 recapitulates the results of ISH assays on 8cm TL catshark transversal sections. A histological section of vertebra (HES) serves as reference, showing the tissues present as well as mineralizing sites (Fig 3.6a, white arrowheads) and a *col2a1* staining in the vertebra highlights cartilaginous tissues (Fig 3.6b). Expected expression in skeletal tissues was only found for three genes out of thirteen: *spp2*, *clec3a* and *otos* (Fig 3.6c, d, e). The *spp2* gene is expressed exclusively in cells that will later mineralize, both in sites of areolar mineralization (in the vertebral body) and in the outer zones



**Figure 3.5:** Expression patterns of three genes representing the main patterns observed in the gene list. E.St.X = Embryo Stage X; OS = Olfactory Sac, AOL = Ampullae of Lorenzini. (a) Ubiquitous, (b) Bimodal, (c) Skeleton skewed. Z-scores are indicated for all three endoskeletal tissues.

of the neural arches (black arrowheads). Both *otos* and *clec3a* however, seem to be less spatially restricted, showing a more general expression pattern in most chondrocytes. For all the other genes no expression was detected in the skeletal tissues but showed instead expression in the spinal chord, exemplified with *coch* (Fig 3.6f). Surprisingly, genes such as *coch* with expression patterns biased toward skeletal tissues in adults and having strong log<sub>2</sub> fold-change values in our RNAseq (here 3.07) are not detected in older embryonic skeletal tissues. The most intuitive way to explain this is to consider *coch* and the other genes as having a function in adult skeletal tissues but not in embryonic ones.

Despite several screenings and a tissue contamination, a signal associated with cartilage was extracted, albeit not as strong as expected. This is explained partly by a spinal cord contamination that was probably quite predominant but also, as mentioned above, by the fact that adult and embryonic tissues are hardly comparable. In fact, our results suggest that adult skeletal markers do not translate perfectly into embryonic skeletal markers.



**Figure 3.6:** Histological staining (a) and ISH assays (b-f) in transversal sections of small-spotted catshark vertebra (8cm TL). (b-e) Cartilage specific genes; (f) *coch* stains specifically the spinal cord, and presents the same pattern as the rest of the genes not shown here. White arrowheads indicate sites of mineralization, black arrowheads indicate sites of expression. nt = notochord; sc = spinal cord

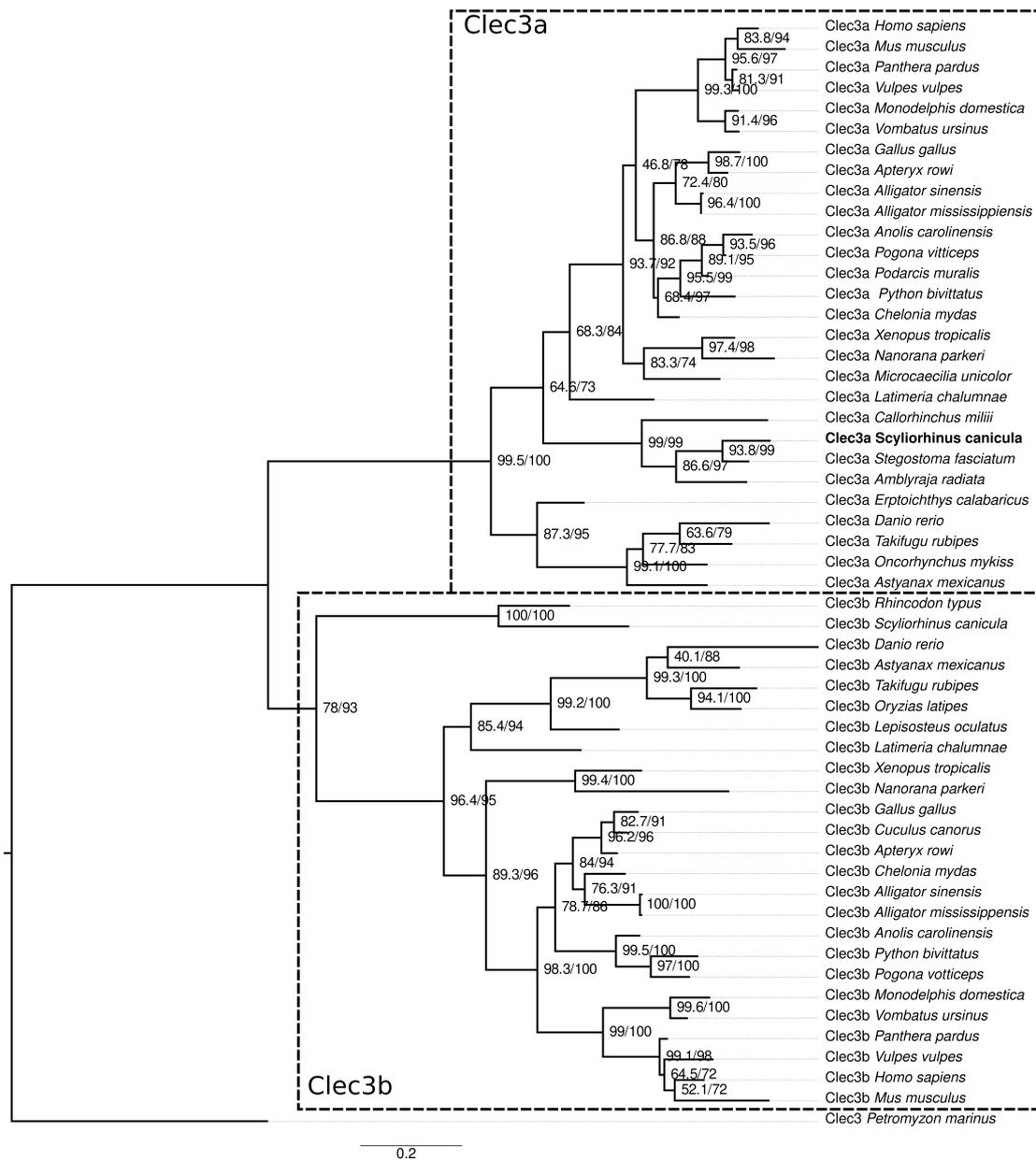
### 3.3.4 Case studies: *clec3a*, *otos*, *spp2*

To further characterize the cartilage associated genes in this study, namely *clec3a*, *otos* and *spp2*, we proceeded to phylogenetic analyses. This allows for correct identification of the genes (i.e. *clec3a* or possibly *clec3b*), particularly important for the several *spp2* genes, and is essential to compare expression patterns between major vertebrate lineages.

#### *clec3a*

Phylogenetic reconstruction of *clec3* protein sequences shows that the *Scyliorhinus canicula* predicted gene is correctly annotated, albeit with an unexpected placement of chondrichthyans within this clade (Fig 3.7, bold leaf). Only one *clec3* sequence was found in the sea lamprey and was used to root the tree. This probably means that *clec3a* and *clec3b* were generated through a whole genome duplication specific to jawed vertebrates. *Clec3a* is part of a family of carbohydrate-binding proteins called lectins. Although the protein is still poorly known, it

is known to be expressed specifically in mice vertebral cartilage and is thought to have a role in the differentiation of chondrocytes in intervertebral discs (X. Chen et al., 2022). Additionally, in mammals, the gene has been shown to have higher expression in osteoarthritic cartilage (Karlsson et al., 2010) and it may function as an oncogene since its suppression seems to inhibit osteosarcoma cell proliferation (Ren et al., 2020). These lines of evidence coupled with the expression pattern in the catshark as well as the up-regulation through development, all point towards an ancestral role of *clec3a* in cartilage development in jawed vertebrates.

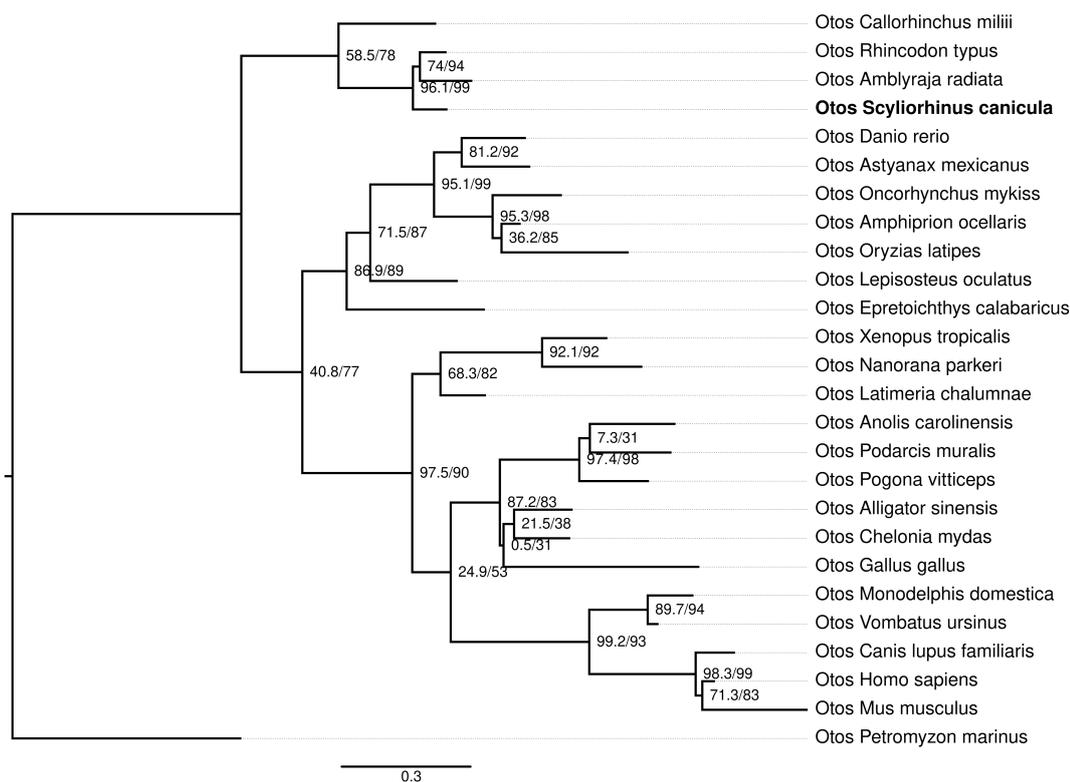


**Figure 3.7:** Phylogenetic reconstruction of vertebrate *clec3* protein sequences. The alignment was made of 54 sequences and was 213 aa long and the cleaning step was used with the `-large` argument. The model of amino acid evolution used was JTT+I+G4. The bold leaf corresponds to the DEG studied here.

**otos**

Phylogenetic reconstruction using otospiralin (*otos*) protein sequences produces an expected tree albeit with low support values in deeper nodes. This is most probably due to the *otos* proteins being quite short and being well conserved, therefore carrying little phylogenetic information. Nonetheless, the phylogeny shows that the *Scyliorhinus canicula* sequence is indeed an *otos* sequence and that the annotation was correct (Fig 3.8, bold leaf).

*Otos* is expressed by the fibrocytes of cochlea of mammals and is essential to normal auditory function (Delprat et al., 2005). However it was also identified as a novel protein potentially involved in zebrafish skeletal development (Kessels et al., 2014) while the paralog of this gene (in teleosts, *otos-like*), has been shown to be expressed in the notochord of zebrafish embryos (Baanannou et al., 2020). Although *otos* is expressed in cochlea, its detection in the skeletal ECM of zebrafish might suggest an additional role in the skeleton of this species. This observation, albeit with no specifics on the tissue this genes might be expressed (bone or cartilage), coupled with the specific expression in the catshark chondrocytes might suggest an ancestral function of *otos* in skeletal development.



**Figure 3.8:** Phylogenetic reconstruction of vertebrate *otos* protein sequences. The alignment was made of 21 sequences and was 98 aa long. The model of amino acid evolution used was JTT+I+G4. The bold leaf corresponds to the DEG studied here.

### *spp2*

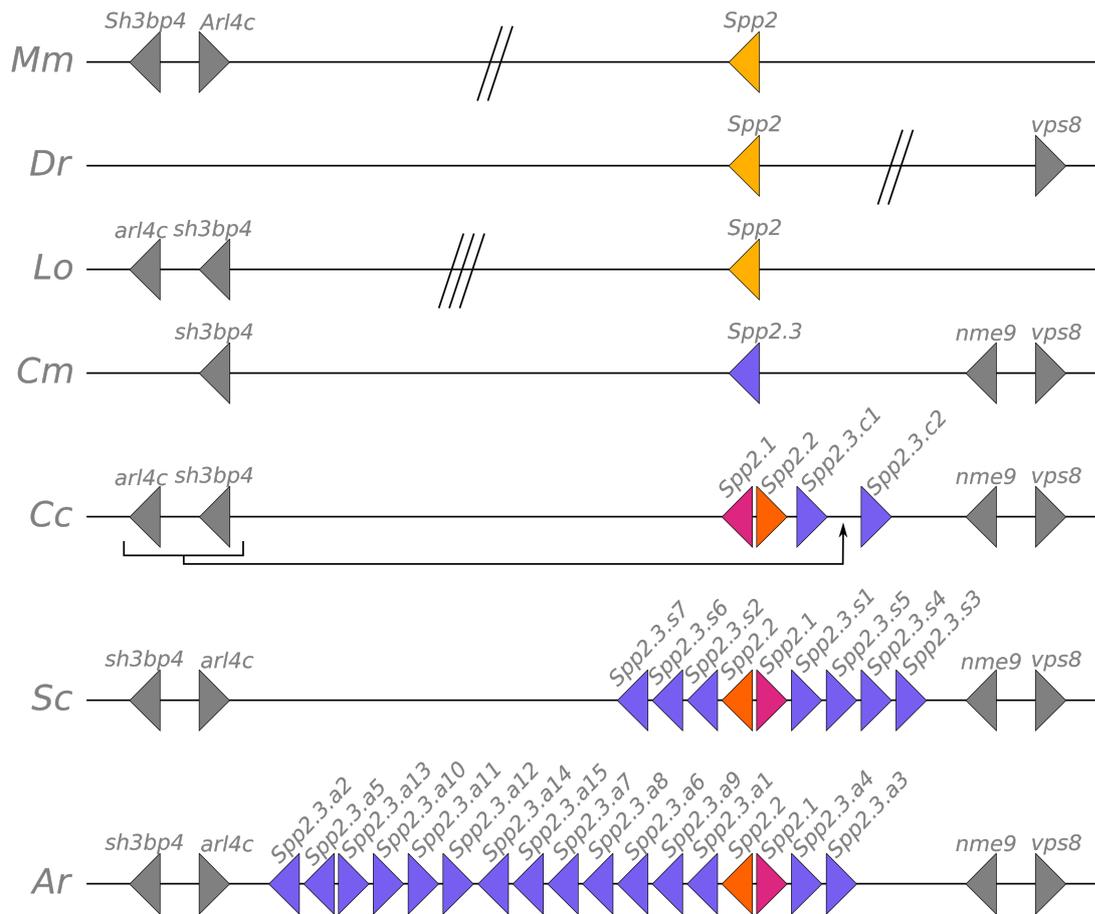
Phylogenetic reconstruction of the evolutionary history of the *spp2* (or *spp24*; Fig 3.9) family shows an unexpected expansion of this gene family in chondrichthyan fishes compared to bony fishes. The phylogeny was rooted using members of the kininogen family, since there is a comparable level of sequence identity between *spp2* genes and the cystatin domains of kininogens (Hu et al., 1995). Inside chondrichthyans there is a variable number of *spp2* genes ranging from one in *Callorhinchus milii* to seventeen in *Amblyraja radiata*. The phylogeny sets apart three main clades in chondrichthyans herein called *spp2.1*, *spp2.2* and *spp2.3*. The branching of the first two suggests that they have been lost in *C. milii*, while *spp2.3* is the gene that has undergone several tandem duplications at different rates in several chondrichthyan species. The short branches in this phylogeny coupled with high sequence similarity indicates quick and successive duplications happening independently in each of the studied species. However, it is interesting to note that these duplications have happened at a higher taxonomical level in Orectolobiformes (i.e. *R. typus* + *C. plagiosum*;  $\approx 100$ My; Hara et al., 2018) based on the topology of this tree. These *spp2* copies have been called following the taxonomical level at which they happened: *Spp2.3.a1-15* for *A. radiata*, *spp2.3.s1-7* for *S. canicula* and *spp2.3.o1-6* in Orectolobiformes. Bold leaves in the tree (Fig 3.9) indicate the four *spp2* sequences which show up as DEGs and are all part of the *spp2.3* clade.

Synteny analysis of these sequences in their genomic context show good conservation of genetic markers (grey colored genes) across jawed vertebrates (Fig 3.10) and further validates their pseudo-orthology relationship (i.e. a 1:9 orthology between bony fishes and *S. canicula*).

The expansion of the *spp2* family in chondrichthyans coupled with the extremely specific expression in mineralizing chondrocytes and in other mineralizing tissues (Fig 3.6c) suggests that these novel genes have been recruited to play a role in the mineralization process at least in the small-spotted catshark. It is important to note that the primers we used for the *in situ* hybridization can amplify the four identified DEGs (*spp2.3.s2* and *spp2.3.s5-7*) but also the other *S. canicula* *spp2* genes in this phylogeny with the exception of *spp2.2*.

In bony fishes *spp2* is thought to modulate the rate and magnitude of bone formation (Murray et al., 2015). It is one of the components, together with Fetuin-A and Mgp, that form calciprotein particles. These particles carry calcium and phosphate ions and can remain in suspension under physiological conditions without precipitating, suggesting that it may be a mechanism to avoid ectopic mineralization, e.g. in vascular tissues (Turner et al., 2021). Additionally, transgenic mice over-expressing *spp2* have been observed to have reduced femoral and vertebral bone mineral density (Sintuu et al., 2008), while in zebrafish a strong increase in abundance of the *spp2* protein was found during development in the skeletal ECM (Kessels et al., 2014). These findings, highlight the role of *spp2* in bone development in bony fishes. This however, contrasts the underwhelming amount of publications on the role of bony fish *spp2* in skeletal development which explains why we might have missed this gene in classical



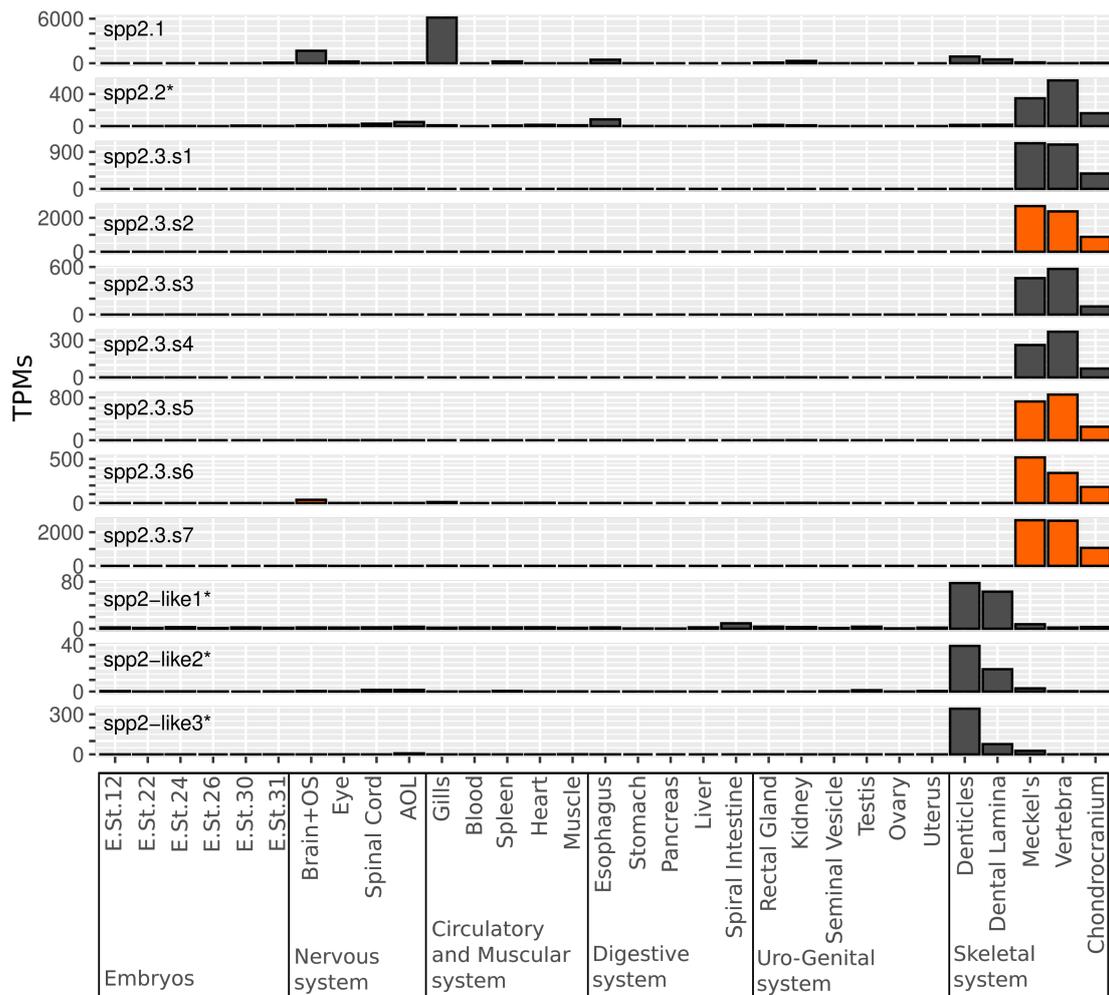


**Figure 3.10:** Synteny of the *spp2* genes in several vertebrate species including: *Mus musculus* (Mm), *Danio rerio* (Dr), *Lepisosteus oculatus* (Lo), *Callorhynchus milii* (Cm), *Carcharodon carcharias* (Cc), *Scyliorhinus canicula* (Sc) and *Amblyraja radiata* (Ar), double and triple bar indicate long genomic stretches. Colors correspond to clades in Fig 3.9

candidate gene approaches.

It is noteworthy to mention three *spp2*-like genes in *S. canicula* and two in *C. carcharias* located in another locus but in the same chromosome, although one of the *S. canicula* copies is actually in another chromosome. RNAseq data for the *S. canicula* copies show divergent expression patterns with high specificity for odontodes (Fig 3.11) and alignments with the other *spp2* genes shows that these sequences do not align very well (not shown), avoiding their use in the phylogenetic reconstruction. They also present highly repeated motifs and are sometimes incorrectly annotated as yeast genes. Although displaying surprising expression patterns in RNAseq data (Fig 3.11), these genes are not differentially expressed in our data, and therefore fall out of the scope of this project. Additionally, there is a possibility that the *spp2.3.s4* gene actually represent two different *spp2* genes in *S. canicula* as a consequence of an incorrect gene prediction (not shown). This however, only impacts the number of *spp2* copies in *S. canicula*.

An ancestral state reconstruction based on the phylogeny indicates that only one *spp2* gene was present in the ancestor of jawed vertebrates. Additionally, a function of this gene in



**Figure 3.11:** TPM distribution among adult tissues for all the identified *spp2* genes, as well as the three *spp2-like* genes in *Scyliorhinus canicula*. Orange colored genes correspond to the identified DEGs in this study. Genes not amplified by the defined probe are marked by an asterisk.

bony fish bone and its expression exclusively in mineralizing cells of the catshark, suggests that this mineralization associated expression was already present in the ancestral jawed vertebrate gene. However, the fact that the expression pattern in the catshark engulfs several very similar sequences and show a very spatially restricted pattern might suggest two different scenarios: (i) that these duplicates are extremely recent and have not had the time to acquire new functions, or become pseudogenized, and are therefore quite redundant or (ii) that these genes operate in a dose dependent manner which would positively select the multiplication of many duplicates sharing high similarity in sequence and function. The latter scenario is more credible when taking into account the fact that the degree of mineralization in each species is correlated with the number of *spp2* copies: from only one copy in *C. milii* a species that displays relatively low levels of mineralization, to seventeen copies in *A. radiata* whose skeleton is extremely mineralized (Berio et al., 2021).

### 3.4 Concluding remarks

In this chapter I explained the several steps, challenges and results of a project that highlights the challenges of using and comparing transcriptomic data in Evo-Devo studies. First, while classical transcriptomic analyses are quite straight forward, here our sampling (dissection) was sub-optimal which significantly increased complications in downstream analyses. In order to overcome the problem of tissue-mix in our sampling, we used several screening methods including, Z-scores, cross referencing, differential expression analyses, manual identification and expression pattern selection to isolate potential genes linked to the mineralization process in the small spotted catshark. Once a set of genes was selected, we further studied these genes through ISH assays as well as phylogenies (and synteny if needed) and comparative analyses with bony fishes, to understand the putative functions of these genes and their evolutionary history.

This project was the first attempt to explore mineralization of chondrichthyans through a transcriptomic approach, in an effort to increase efficiency of mineralization associated gene identification. However, although in theory a transcriptomic approach is more optimal in terms of gene number (especially when trying to be exhaustive) there is one inconvenient to this approach. Since the genes are necessarily selected through differential expression, there are many biological processes that can influence this analysis. For instance, even in pure tissues such as chondrichthyan cartilage, cellular proliferation, differentiation can scramble the desired signal. Additionally, even a “pure” tissue is not made up of a unique cell lineage, therefore containing several molecular fingerprints, e.g. in cartilage we can find fibroblasts and chondrocytes, or in nervous tissue, neurons and glial cells.

We demonstrated that even with sub-optimal sampling, mineralization associated genes such as the chondrichthyan *scpp* can be identified through transcriptomic approaches therefore loosening the dependence from the very derived skeletal biology of bony fishes on which we relied heavily at first. It is worth mentioning that other mineralization associated genes were missed using this approach, such as *collagen X* that does not present Z-scores  $\geq 4$  in endoskeletal tissues (Z-scores = -1.5, -1.3 and 0.3 in chondrocranium, vertebra and meckel’s respectively), or *mgp* that does not display significant variation between young and old embryos.

We identified several genes specifically expressed in the cartilage of the catshark, however only the *spp2* gene family seems to be associated to the mineralization process with the number of copies being positively correlated with the degree of mineralization in chondrichthyan species. These genes might not have been identified through classical candidate gene approaches because of a publication bias: our capacity on defining candidate genes depends strongly on the selected genes that are being studied and published in the literature.



# Chapter 4

## The genes that do not want to fall into ranks

### 4.1 Introduction

Studying candidate mineralization-associated genes in the catshark, has often led to careful examination of the whole gene family to understand their evolutionary history. This, however, can result in a high number of studied genes (when the gene family is big), with many paralogs that end up having no strong link to mineralized tissues. Often, every member of the family is studied in an attempt to be as exhaustive as possible and not miss potentially important orthologs or paralogs for mineralization. In these cases, we have chosen to shift the focus of our study towards a molecular evolution approach by engulfing the gene family as a whole instead of a more classical Evo-Devo approach where only one or two genes are studied at several developmental stages. Therefore, a compromise was made between the number of genes studied and the number of developmental stages included in the studies. A molecular evolution study includes more genes at one stage, while an Evo-Devo approach includes less genes at more stages.

In this chapter I will be presenting some of these molecular evolution approaches that represent the work produced through the supervision of two interns : Nathan Gil (01/21-06/21) and Amirhossein Karimizadeh (04/22-06/22). Nathan worked on the *slrp* gene family and the corresponding article is ready for submission, while Amirhossein did his work on the *mmp* genes which is still a work in progress. Each of these two studies will be presented in this chapter under the form of preliminary manuscripts.

## 4.2 Genomic and transcriptomic characterization of the *slrp* family in *S. canicula*

*Nicolas Leurs, Nathan Gil, Camille Martinand-Mari and Mélanie Debiais-Thibaud*

Institut des Sciences de l'Evolution de Montpellier, ISEM, Univ Montpellier, CNRS, IRD, EPHE, Montpellier, France.

### 4.2.1 Introduction

A diversity of connective tissues has emerged with the evolution of vertebrates (Root et al., 2022; Root et al., 2021). The macromolecular content of their extracellular matrices (ECM) consists firstly of collagen fibers (e.g. Collagen type I in bone and Collagen type II in cartilage; in bony fishes), together with other types of proteins, and lipids. Among non-collagenous proteins, proteoglycans are particularly abundant in the highly hydrated ECMs (proteoglycan content in cartilage: 5-7% w/w; Hardingham, 2006). Proteoglycans constitute a class of proteins linked by post-translational modifications to long un-ramified chains of disaccharide repetitions, called glycosaminoglycans (GAGs). GAGs are negatively charged, highly hydrophilic and participate widely in the function of their proteoglycan (Hardingham, 2006; Karamanos et al., 2018).

Within proteoglycans, the largest known vertebrate family is the Small Leucine-Rich Proteoglycans (SLRPs). In this family, the protein moiety is relatively small (36-42 kDa) and has a distinctive leucine-rich repeat (LRR) (Iozzo and Schaefer, 2015; Nikitovic et al., 2012). SLRPs play critical roles in the structure and assembly of various ECMs and hence in the development, structure and homeostasis of connective tissues (Boskey, 2010; Hardingham, 2006; Nikitovic et al., 2012; Park et al., 2008). For instance, they are known to interact with several types of collagen fibers by regulating their assembly via their protein moiety, while their GAGs control the correct spacing between fibers. SLRPs can also regulate apatite formation and interact with skeletal growth factors such as TGFs and BMPs (Boskey, 2010; Iozzo and Schaefer, 2015; Nikitovic et al., 2012; Schaefer and Iozzo, 2008).

The evolution of early vertebrates was characterized by two rounds of whole genome duplications (2R event; Ohno, 1970). As a result, several preexisting gene families have been expanded and the newly duplicated copies may have either acquired new functions, divided the original function between paralogs or simply become pseudogenic (following a duplication-degeneration-complementation model, Force et al., 1999). In bony vertebrates, twenty-three SLRP paralogs have been identified and most can be found along four different chromosomes (Costa et al., 2018; Iozzo and Schaefer, 2015; Park et al., 2008; Schaefer and Iozzo, 2008). They are classically divided into 5 classes depending on protein sequence similarity and gene chromosomal localization (Schaefer and Iozzo, 2008). Among these five classes, Class I contains asporin (Aspn), biglycan (Bgn), decorin (Dcn), and four "ECM proteins" (Ecm2, Ecm2L, EcmX and EcmXL). Class II includes fibromodulin (Fmod), keratocan (Kera), lumican (Lum),

lumican-like (LumL), osteomodulin (or osteoadherin, Omd) and prolargin (Prelp). Class III consists of epiphycan (Epyc), opticin (Optc) and osteoglycin (or mimecan, Ogn). Class IV encloses chondroadherin (Chad), chondroadherin-like (ChadL), nyctalopin (Nyx) and tsukushin (Tsku). Class V encompasses podocan (Podn) and podocan-like (PodnL). Nephrocan (Npc) does not belong to any of these classes for structural reasons, despite its classification into SLRPs (Costa et al., 2018; Iozzo and Schaefer, 2015; Mochida et al., 2006; Schaefer and Iozzo, 2008). The sixteen SLRPs from classes I to III are considered as “canonical” since they present an extended repeat (called “ear” repeat) in the LRR C-terminal capping motif (LRRCE) of their protein moieties (Park et al., 2008). Despite the expectations of an already diverse SLRP repertoire prior to vertebrate evolution, only few sequences have been identified in the vertebrate sister-group (two Ciona SLRPs with LRRCEs in Park et al., 2008).

Because previous studies were highly biased towards bony vertebrate (osteichthyans) sampling, we took advantage of newly available high-quality genomic data in the cartilaginous fish (chondrichthyan) lineage to assess the evolution of the SLRP gene family in vertebrates. We identified the products of the 2R duplications on clustered and non-clustered SLRPs, which allows us to describe the ancestral state prior to the 2R event. We further identified gnathostome lineage-specific gene losses, and evaluated the divergence of expression patterns following these 2R duplications. In particular, we focused on new data brought by genomic and transcriptomic data in the small-spotted catshark (*Scyliorhinus canicula*), for which tissue-specific expression patterns can be described.

## 4.2.2 Materials and Methods

### Protein sequence sampling

SLRP protein sequences were recovered from public databases (NCBI, Ensembl and the european seabass genome) via BLASTP or TBLASTN on nine sarcopterygians [*Mus musculus*, *Gallus gallus*, *Vombatus ursinus*, *Tinamus guttatus*, *Podarcis muralis*, *Chelonia mydas*, *Xenopus tropicalis*, *Microcaecilia unicolor*, *Latimeria chalumnae*], seven actinopterygians [*Lepisosteus oculatus*, *Danio rerio*, *Dicentrarchus labrax*, *Erpetoichthys calabaricus*, *Sparus aurata*, *Gadus morhua*, *Anguilla anguilla*] and three chondrichthyans [*Callorhynchus milii*, *Rhincodon typus* and *Amblyraja radiata*]. Whenever a *M. musculus*, *G. gallus*, *X. tropicalis* or *M. unicolor* sequence was not identified in the databases, they were respectively replaced by a sequence from *Bos taurus*, *Meleagris gallopavo*, *Nanorana parkeri* or *Rhinatrema bivittatum*. Domestic mouse and zebrafish sequences were further used to screen the small-spotted catshark (*Scyliorhinus canicula*) reference gene model (Mayeur et al., 2021) as well as the locally assembled transcriptome of the thornback ray (*Raja clavata*, Debais-Thibaud, 2019). Reciprocal BLASTs were performed to restrain the recovered *S. canicula* and *R. clavata* sequences to actual SLRPs. Two sequences recovered for *C. milii* (PodnL: XP\_007905284.1 + XP\_007905354.1) and two for *R. clavata* (Ecm2: c112061\_g11\_i2 + c112061\_g10\_i1) were fused together as alignments

showed they were partial sequences of the same gene. Moreover, seven teleostean sequences for EcmXL [*Astyanax mexicanus*, *Chanos chanos*, *Electrophorus electricus*, *Ictalurus punctatus*, *Pangasianodon hypophthalmus*, *Pygocentrus nattereri*, *Sinocyclocheilus rhinoceros*] were recovered via BLAST of the *D. rerio* sequence, to strengthen the resolution in teleosts. Additional *Petromyzon marinus* SLRP sequences from (Ota et al., 2013; Park et al., 2008) were used to screen the *P. marinus* genome in NCBI for additional sequences that may have been missed by the previous studies. Twenty-three SLRP amino-acid sequences were found on the *S. canicula* reference gene model (Mayeur et al., 2021). Alignments of four sequences identified as “Chad-like” with the reference genome (GCF\_902713615.1) showed that some of them were overlapping (XM\_038783897.1 and XM\_038783978.1; XM\_038783861.1 and XM\_038783855.1), which suggests a reconstruction error during the genome assembly: only two sequences seem to be fully independent. Following alignments with *C. milii*, *R. clavata* and *A. radiata*, only two sequences were kept (XM.038783897.1 and XM.038783861.1). As a result, a total of twenty-one protein coding sequences in *S. canicula* were identified for the phylogenetic reconstructions: AspN, Bgn, Chad, two ChadL, two Dcn, Ecm2, Ecm2L, Epyc, Fmod, Kera, Lum, Npc, Nyx, Ogn, Omd, Podn, PodnL, Prelp and Tsku. All sequences used in this study are listed with accession numbers (Suppl Mat on demand).

### Phylogenetic analysis of SLRP phylogeny

The sampled protein sequences from vertebrate genomes (n= 470) were used to reconstruct several phylogenies: one on the whole data set and four others on selected subsets (clades 1-4, as defined below). To generate each phylogenetic reconstruction, selected sequences were aligned with MAFFT (Kato and Standley, 2013; standard parameters and `-auto` strategy). Because a large proportion of the sequences is predicted from genomes and may include false exons, alignments were then cleaned with HmmCleaner (Franco et al., 2019; with `-large` option) to remove low similarity segments. Our final alignments used for subsequent phylogenetic reconstruction included: Full: 1142, clade 1: 548, clade 2: 383, clade 3: 384 and clade 4: 361 amino acids and are available (Suppl Mat on demand). Phylogenies were then inferred by Maximum Likelihood via IQ-Tree (version 1.6.1; Nguyen et al., 2014). The best model of amino-acid evolution was selected using ModelFinder (Kalyaanamoorthy et al., 2017) and the Bayesian Information Criterion (Schwarz, 1978). The two models used were: LG+I+G4 (Full, clades 3 and 4) and JTT+I+G4 (clades 1 and 2). Node support was estimated by performing a thousand ultra-fast bootstrap replicates (UFBoot; Hoang et al., 2017), coupled with the `-bnni` option in case of model violations, and single branch tests (SH-aLRT; Guindon et al., 2010). Phylogeny tree files can be found in (Suppl Mat on demand).

## Synteny analysis of SLRPs

SLRP synteny was explored in five reference genomes available on NCBI: *M. musculus* (GCF\_0001635.27), *S. canicula* (GCF\_902713615.1), *E. calabaricus* (GCF\_900747795.1), *D. rerio* (GCF\_000002035.6) and *C. milii* (GCF\_000165045.1). Data from *D. rerio* and *C. milii* are shown only when they differ from *E. calabaricus* and *S. canicula*, respectively. *E. calabaricus* was chosen over *D. rerio* as a non-teleost actinopterygian devoid of the 3R-generated duplicates.

## SLRP protein domain screening

*S. canicula* SLRPs were screened using LRR finder (Bej et al., 2014), EML (M. Kumar et al., 2019; cell compartment: extracellular) and Sulfinator (Monigatti et al., 2002) that predict, respectively, putative positions for Leucine-Rich Repeats (LRRs), glycosylation, and tyrosine sulfation sites. Additionally, the proteins were screened for two LRRCE motifs: (i) the standard one from Park et al., 2008 typical of canonical SLRPs; and (ii) an unconstrained in the C-terminal region using the ScanProsite tool (de Castro et al., 2006). The model LRRCE motif was (PROSITE syntax):

### Standard LRRCE Motif

[LIV]-X(2)-[LVIYFMA]-X-[LIFM]-X(2)-[NH]-X-[ILVF]-X(2)-[VIMFLY]-X(4)-[FIMLV]-C-X(7,20)-[LYIMV]-X(2)-[ILVTMF]-X-[LVMI]-X(2)-N-X-[IVLMAFT]-**X(8,9)**-[FYMPVAIS]-**X-C**

### LRRCE Motif Unconstrained at C-terminal

[LIV]-X(2)-[LVIYFMA]-X-[LIFM]-X(2)-[NH]-X-[ILVF]-X(2)-[VIMFLY]-X(4)-[FIMLV]-C-X(7,20)-[LYIMV]-X(2)-[ILVTMF]-X-[LVMI]-X(2)-N-X-[IVLMAFT]-**X(8,12)**-C

## Primers for qPCR and *in situ* mRNA hybridization

Forward and reverse primers were defined using Primer3Input version 4.1.0. Primers were constrained to the coding portion of genes and the designed product had to include at least one intron-exon boundary, following the NCBI prediction of these features. The product size ranges were 250-350bp for primer couples designed for RT-qPCR amplification, and 650-800bp for amplifications designed to generate the DNA matrix for RNA probe synthesis (see below *in situ* mRNA hybridization). Only three primers per gene were defined in some instances, one being common to both the RT-qPCR and the *in situ* hybridization (primers are given in Suppl Mat on demand).

## RNAseq mapping on the catshark reference gene model

*S. canicula* tissue libraries from whole embryos at stages 12, 22, 24, 26, 30 and 31 (Ballard et al., 1993) and adult tissues sampled from three individuals were recovered from NCBI (BioSam-

ple accessions: 21694013 to 21694062) and are as following: “Tissue (biological replicates)” Ampullae of Lorenzini (2), Blood (2), Brain+Olfactory sac (3), Chondrocranium (3), Dermal denticles (3), Dental lamina (3), Esophagus (1), Eye (2), Gills (1), Heart (1), Kidney (1), Liver (1), Meckel’s cartilage (3), Muscle (2), Ovary (1), Pancreas (2), Rectal gland (2), Spinal cord (1), Spiral intestine (2), Spleen (1), Stomach (1), Seminal vesicle (1), Testis (2), Uterus (1) and Vertebrae (2). The paired end RNAseq reads from adult and embryonic *S. canicula* tissue samples were mapped onto the reference gene model (Mayeur et al., 2021) using BWA MEM (H. Li, 2013). After sorting and converting the resulting alignments to BAM files using samtools, counts for each library were extracted. Count data was then normalized as Transcripts Per Kilobase Millions (TPMs) and used as such for quantification of expression. When a tissue had several replicates (max 3 replicates), TPM values were averaged, leading to a final number of 31 tissues compared. For each gene, Z-scores were then calculated to evaluate tissue-biased expression as

$$Z = (x - \mu) \cdot \frac{\sqrt{n}}{\sigma} \quad (4.1)$$

Where for any sample:  $x$  is the observed TPM value in a given tissue,  $\mu$  is the mean of all tissues,  $\sigma$  is the standard deviation of all tissues and  $n$  is the number of tissues.

### RT-qPCR

Total RNAs from 22 anterior vertebrae (AV) of *S. canicula* embryos from 5 to 8cm Total Length (TL) were isolated with ReliaPrep RNA tissue Miniprep system according to the supplier’s instructions (Promega) and used for cDNA preparation performed by Superscript II reverse transcription (Invitrogen) with an oligodT primer. Each 1:20 diluted cDNA was run in triplicate on a 384-well plate for each primer pair by using thermal cycling parameters: 95°C for 2 min, 95°C for 10 s, 68°C for 10 s, 72°C for 10 s (45 cycles), and an additional step 72°C for 10 min performed on a LightCycler 480 with the SensiFAST SYBR No-ROX kit (Meridian Bioscience) (qPHD UM2/GenomiX Platform, Montpellier – France). Results were normalized with the expression of three reference genes *eef1a*, *actin* and *gapdh* by geometric mean, and data were further analyzed with the LightCycler 480 software 1.5.1. The reference point used was the highest value of  $\Delta C_p$  for a given gene: all expression values of any given gene are all above or equal to 1-fold. Developmental trajectories were plotted on R (v4.04) using the ggplot2 package.

### *In situ* mRNA hybridization

In situ hybridizations were performed on 14  $\mu$ m thick cryostat sections of 6,5 cm TL *S. canicula* embryos, cut transversely in the body trunk, at the level of the pectoral fins. All subsequent procedures were previously described (Leurs et al., 2021). Slides were scanned on a Hamamatsu nanozoomer. Primers designed to generate the DNA matrix used for RNA probe synthesis are

given in Suppl Mat on demand.

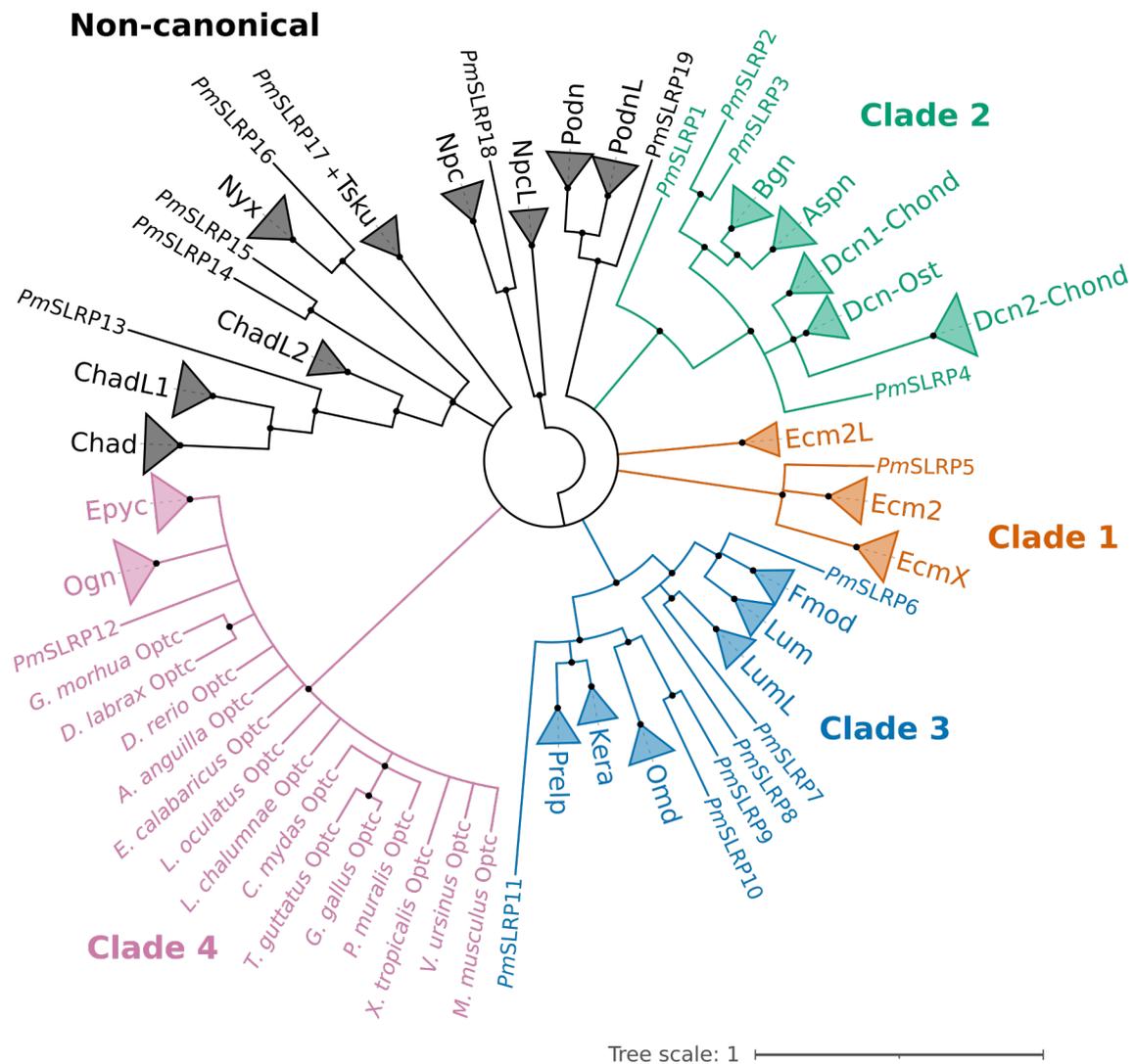
### Embryo collection and ethics statement

Embryos of the small-spotted catshark *S. canicula* originated from a Mediterranean population of adult females housed at Observatoire Océanique de Banyuls, France. Handling of small-spotted catshark embryos followed all institutional, national, and international guidelines [European Communities Council Directive of September 22, 2010 (2010/63/UE)]: no further approval by an ethics committee was necessary as the biological material is embryonic and no live experimental procedures were carried out. Embryos were raised in seawater tanks at 16–18 °C and euthanized by overdose of tricaine (MS222, Sigma) at appropriate stages. Whole embryos were fixed in PFA4% for 48h and then stocked in ethanol 100% before the tissue was sampled for cryostat sectioning.

### 4.2.3 Results

#### A conserved SLRP repertoire in gnathostomes

The complete SLRP phylogeny (herein ‘Full phylogeny’) was rooted using *Npc* and its sister clade as outgroups (Fig 4.1) since these SLRPs were shown to be structurally different to other SLRPs (Mochida et al., 2006). This root branches to a wide multi-furcation where several well-supported monophyletic clades group together (Fig 4.1, Suppl Mat on demand): a *Tsku* clade, a *Nyx* and *Chad*-related clade, a *Podn*-related clade, and four independent clades for the canonical SLRPs. Well-supported gnathostome orthology groups were recovered within these wide clades for all non-canonical SLRPS in class V (*Podn* and *PodnL*) and Class VI (*Nyx*, *Chad* and *Tsku*) and the non-classified member: *Npc* (Mochida et al., 2006). However, we found two *ChadL* gnathostome orthology groups. The previously identified *ChadL* clade is herein identified as “*ChadL1*” and the newly identified clade is named “*ChadL2*”. *ChadL2* sequences were only found in chondrichthyans and in non-amniote sarcopterygians and a single *P. marinus* sequence was associated with these *Chad* clades. *ChadL1* and *chadL2* are tandem genes on orthologous loci in both *C. milii* and *S. canicula* (data not shown) while *chad*, *npc*, *nyx*, *podn*, *podnL* and *tsku* are all located as single SLRPs on different loci, with synteny markers supporting their orthology to other gnathostomes’ loci (not shown). The *npcL* gene (sister group to *Npc*, Fig 4.1) however, was absent in *S. canicula*. The remaining SLRP clades 1-4 consist of canonical SLRPs (Iozzo and Schaefer, 2015; Schaefer and Iozzo, 2008). Support for the monophyly of clades 2, 3 and 4 in our full phylogeny is strong (Fig 4.1). Inclusion of the *Ecm2L* clade into clade 1 (Fig 4.1) will be further discussed with synteny data (see below). Internal relationships within each clade are not well supported in this full phylogeny, especially those in clade 4 (Fig 4.1, but see below).



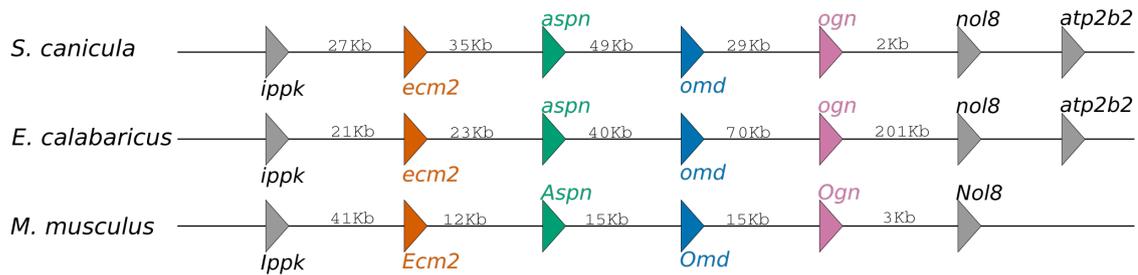
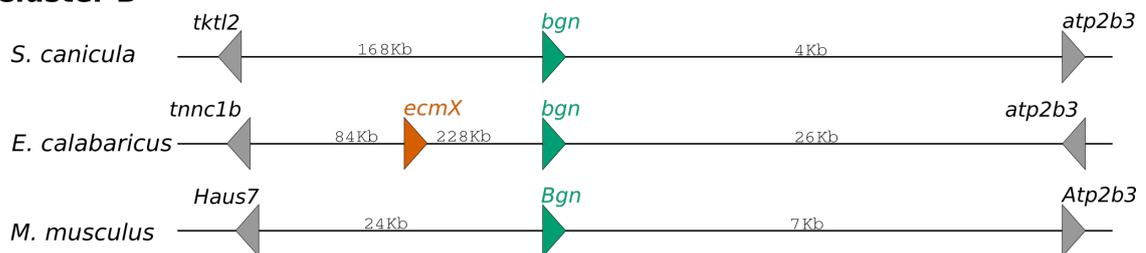
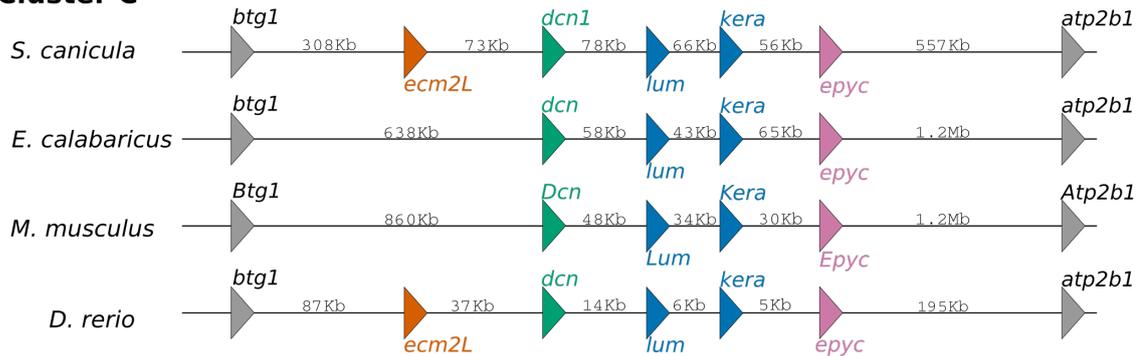
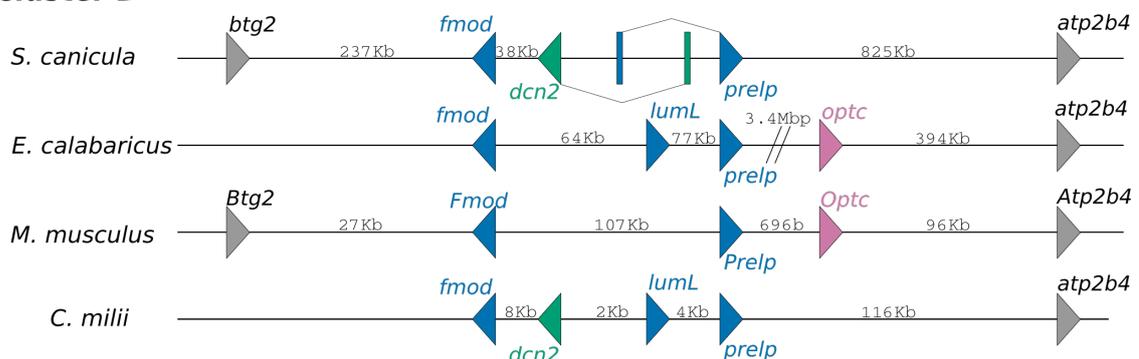
**Figure 4.1:** Phylogenetic relationships of SLRP sequences in vertebrates. Maximum likelihood phylogenetic tree based on amino-acid sequences (470 sequences, 1142 positions) with LG+I+G4 evolution model in IQ-TREE. Node support was evaluated with 1000 ultra-fast bootstrap replicates (and single branch test SH-aLRT, Supp Mat S2). Nodes with values  $\leq 80$  bootstrap were collapsed into multi-furcations. Clades (1-4) of canonical SLRPs are shown in color.

### Genomic organization and the evolution of canonical SLRPs within gnathostomes

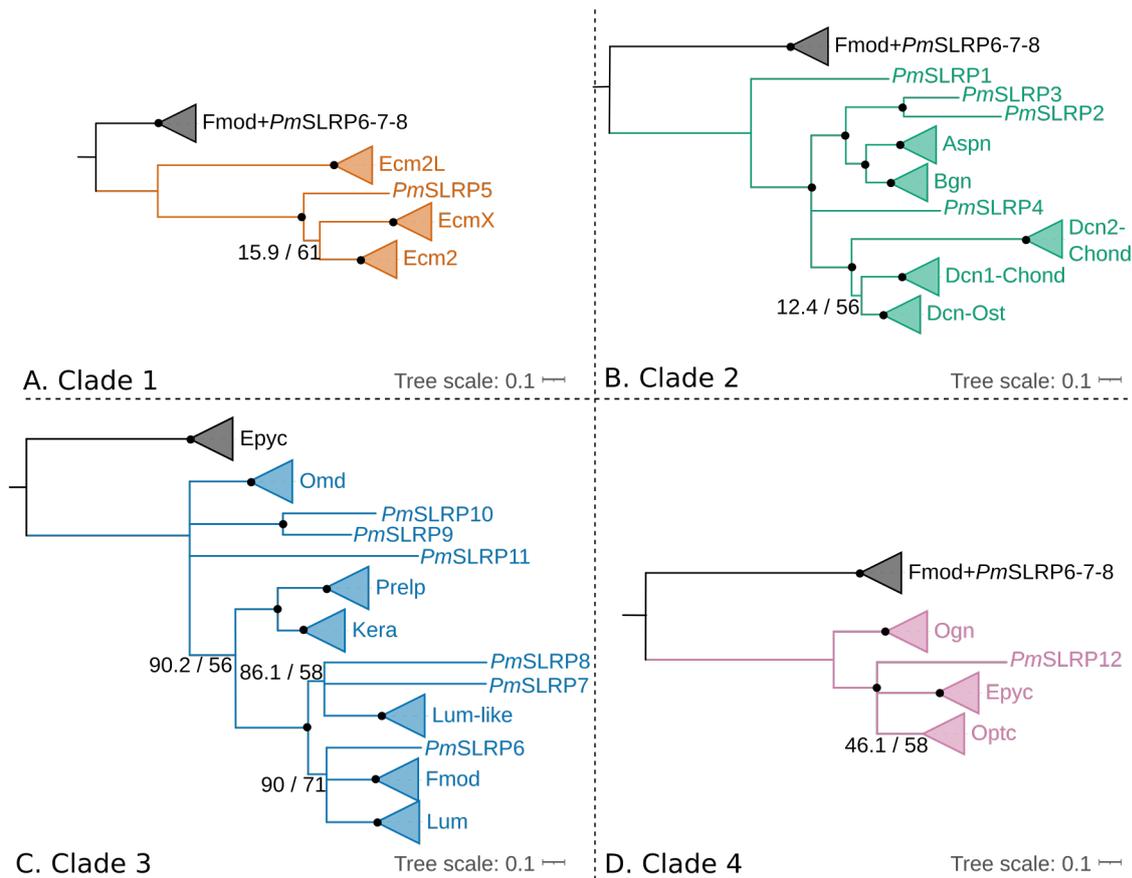
To better describe orthology relationships in canonical SLRPs, we performed a synteny analysis in several vertebrate genomes and generated four focused phylogenies by sub-sampling the sequences of interest. Canonical SLRPs are clustered along four loci that appear conserved between jawed vertebrate genomes (Fig 4.2). These gene clusters will be referred to as clusters A, B, C and D. Genes in cluster A are *ecm2*, *aspn*, *omd* and *ogn*; in cluster B: *ecmX* (found only in *E. calabaricus*) and *bgn*; in cluster C: *ecm2L* (not in *E. calabaricus* nor in *M. musculus*), *dcn1*, *lum*, *kera*, and *epyc*; in cluster D: *fmod*, *dcn2* (only in cartilaginous fishes), *lumL* (found in *E. calabaricus* and *C. milii*), *prelp*, and *optc* (only found in bony fishes). For each cluster, at least one synteny marker is conserved to support orthology relationships in compared genomes (Fig 4.2). Common placement on the first position of *ecm2L* (cluster C), *ecm2* (cluster A) and

*ecmX* (cluster B) supports the inclusion of *Ecm2L* in clade 1, despite low phylogenetic support (Fig 4.1). Clade 4 genes (*ogn*, *epyc*, *optc*) are always located on the last position of the cluster (but absent on cluster B). Clade 2 and 3 genes show less conserved positions (Fig 4.2). Clade 2 genes (*aspn*, *bgn*, and *dcn1*) are well positioned on the second locus in clusters A, B and C but *dcn2* is positioned between *fmod* and *lumL* on cluster D (Fig 4.2). No Clade 3 gene was found on cluster B: one paralog was identified on cluster A (*omd*), two paralogs on cluster C (*lum* and *kera*) and a maximum of three paralogs on cluster D (*fmod*, *lumL* and *prelp*).

Independent phylogenies for each canonical SLRP clade (1 to 4; Fig 4.3) were performed to further characterize internal relationships and assess secondary losses or gains through the evolution of gnathostome SLRPs. Despite its position in the full phylogeny (Fig 4.1), *Ecm2L* was included in the clade 1 phylogeny for its homologous position in cluster C (Fig 4.2C). In clade 1, *Ecm2* is the sister group to *EcmX* albeit low support values (15.9/61; Fig 4.3). *Ecm2L* was lost in most tetrapods and *EcmX* was lost in chondrichthyans and birds (Suppl Mat on demand). A previously identified gene called *EcmXL* (see Suppl Mat on demand; Costa et al., 2018) was identified only in teleost fishes and branches as a sister clade of teleost *EcmX* sequences (Suppl Mat on demand). This topology is typical of the teleost-specific third whole genome duplication (3R event, Jaillon et al., 2004; Kasahara et al., 2007), and in accordance with the current consensus nomenclature (ZFIN Zebrafish Nomenclature Conventions; Ruzicka et al., 2018) teleost *ecmX* and *ecmXL* will be referred to as *ecmXa* and *ecmXb* respectively. No such conservation of duplicates from the teleost whole genome duplication (WGD) was identified in the *Ecm2* or *Ecm2L* clades (Suppl Mat on demand). In clade 2, *Bgn* and *Aspn* are closer to each other than they are to *Dcn*. The *Dcn* clade included one osteichthyan clade and two chondrichthyan clades that we named *Dcn1* and *Dcn2*, suggesting a duplication for which both duplicates were conserved only in chondrichthyans (see Fig 4.1 clade 2, and Fig 4.2D). The monophyly of *Dcn+Dcn1* is weakly supported (12.4/56) but our synteny analysis suggests that the osteichthyan *Dcn* and the chondrichthyan *Dcn1* are found on the orthologous locus on cluster C while *Dcn2* is found on cluster D, thus supporting the orthology relationship between the chondrichthyan *Dcn1* and the osteichthyan ‘*Dcn*’ (Fig 4.2C and D). The 3R duplication in teleosts was only partially conserved in clade 2 with two teleost copies of *Bgn* (*bгна* and *bgnb*). In clade 3, *Lum*, *Fmod* and *LumL* form a monophyletic group, sister to another one formed by *Kera* and *Prelp*. *Omd* comes as an outgroup of these orthology groups. *LumL* was secondarily lost in elasmobranchs but identified in *C. milii*, and lost in tetrapods while *Omd* was partially lost in teleost fishes (Suppl Mat on demand). Teleost 3R duplicates were only partially conserved in the *Fmod* clade. In clade 4, gnathostome *Optc* sequences are recovered monophyletic despite low support values (46.1/58). The *Optc* paralog was lost in chondrichthyans and the teleost-specific duplication is partially conserved (*ogna* and *ognb*) as previously shown by Costa et al. (Costa et al., 2018).

**Cluster A****Cluster B****Cluster C****Cluster D**

**Figure 4.2:** Genomic organization of canonical SLRP gene clusters in reference osteichthyan and chondrichthyan genomes. Letters (A-D) represent the name of each cluster. Distance between genes is indicated (not to scale). Gene names and color code are similar to Figure 1 (for gene accession numbers, see Suppl Mat on demand).



**Figure 4.3:** Phylogenetic relationships in each canonical SLRP clade in vertebrates. Maximum likelihood phylogenetic trees based on SLRP amino-acid sequences in IQ-TREE. clade 1: 86 sequences, 548 positions, JTT+I+G4 model; clade 2: 100 sequences, 383 positions, JTT+I+G4 model; clade 3: 147 sequences, 384 positions, LG+I+G4 model; clade 4: 92 sequences, 361 positions, LG+I+G4 model. Node support was evaluated with 1000 ultra-fast bootstrap replicates (and single branch test SH-aLRT, see Sup Mat 2). Nodes with Ufbootstrap  $\geq 80$  are shown as black dots. All nodes with Ufbootstrap  $\leq 80$  were collapsed except when SH-aLRT was higher than 85 or when nodes are discussed in the text. Color code is the same as in Figure 1.

### Conservation of SLRP protein domains in the catshark

We screened specific SLRP protein domains, such as LRR number and length and putative glycosylation sites, to better understand the degree of conservation or divergence of these proteins relative to bony fishes (see Supp Mat S5). Compared to previously published studies on bony fish SLRP (Matsushima et al., 2021) we found a small degree of variation in the number of LRRs in catshark proteins which for the most part lacked one to two LRRs, compared to mammal sequences. Within catshark canonical SLRPs, we found a standard LRRCE motif in Asp, Ecm2, Dcn1, Dcn2, Lum, Ogn and Epyc. In the remaining sequences only a screening with the C-terminal unconstrained LRRCE yielded hits. The annotated Bgn sequence from the reference genome appeared truncated (only 7 LRRs and no LRRCE motif of any kind) so we used an RNAseq derived sequence (locally assembled transcriptome; Debais-Thibaud, 2019; Suppl Mat on demand) which included 10 LRRs and a standard LRRCE motif. In addition

to the LRR and LRRCE motifs, protein alignment showed good conservation of the cysteine rich motifs on the N terminal capping region of all catshark SLRPs (Low et al., 2021). We found that several *S. canicula* SLRPs sequences showed high similarity in terms of number and position of putative glycosylation sites (e.g. clade 3 SLRPs except Omd) compared to their bony fish orthologs, when data is available.

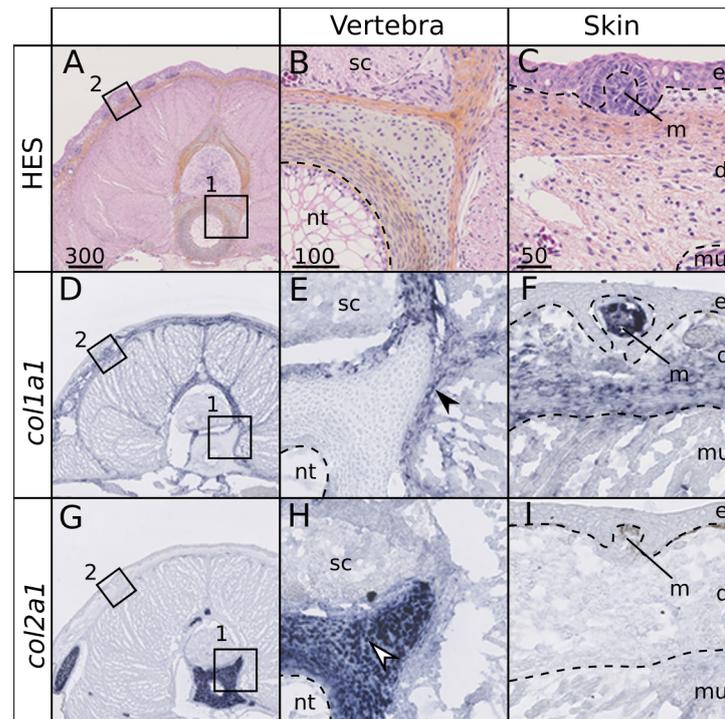
### **Tissue- and embryonic stage-specific expression of SLRP genes in the small-spotted catshark**

Transcriptomic data in a variety of catshark embryonic stages and adult tissues were compared through TPM values (to quantify expression) and Z-score values (a proxy to measure expression bias), both summarized in Table 4.1 (raw Z-score data in Suppl Mat on demand). SLRPs display wide variability in terms of expression levels. Some SLRPs display very low TPM values across all tissues (*dcn2*, *kerA*, *chadL2*, *npc*, *nyx*, *podn* and *podnL*, mean TPM  $\leq 15$ ), while others reach high expression levels in some tissues (e.g. *chad* in vertebrae  $\geq 2700$  TPMs; Table 4.1).

In most sampled tissues, *bgn* displays high levels of expression (TPM  $\geq 50$ ). Expression of most SLRPs is strongly biased (Z-score  $\geq 1$ ) towards endoskeletal tissues (the Meckel's cartilage, vertebrae and chondrocranium samples) and/or exoskeletal tissues (skin denticles and dental lamina). Exceptions to this trend are *dcn2*, *kerA* and *ecm2L* in canonical SLRPs, together with *chadL2* and *npc* in non-canonical SLRPs. *Chad*, *chadL1*, *epyc* and *fmod* are the only SLRPs enriched exclusively in the skeletal system. Several SLRPs from different clusters and different clades have an enriched and strong expression in the ampullae of Lorenzini (AOL), the eye or both (*bgn*, *lum*, *ogn* and *prelp*, Table 4.1), making these sensory organs another major site of SLRP expression. Biased expression is also found for *bgn* and *prelp* in the heart and spleen and for *ogn* and *prelp* in the esophagus (Table 4.1). Compared to the other SLRPs, *tsku* displays a more divergent expression pattern being the only one enriched in both the spiral intestine and the uterus but not in endoskeletal tissues. Finally, *ecm2L* is exclusively enriched in early embryonic stages. Most SLRPs also show strong and/or biased expression in late embryonic stages (stages 30 and 31), when the skeleton is known to engage in cell differentiation. Based on parallel higher enrichment in the skeletal tissues and in late embryonic stages, we further studied SLRPs expression in later *S. canicula* embryonic stages focusing on skeletal development.

	Clade 1					Clade 2					Clade 3					Clade4		Non-canonical SLRPs						
	<i>ecm2</i>	<i>ecm2L</i>	<i>aspn</i>	<i>bgn</i>	<i>dcn1</i>	<i>dcn2</i>	<i>fnod</i>	<i>keru</i>	<i>lum</i>	<i>omd</i>	<i>prelp</i>	<i>epyc</i>	<i>ogn</i>	<i>chad</i>	<i>chadL1</i>	<i>chadL2</i>	<i>npc</i>	<i>ngx</i>	<i>podn</i>	<i>podnL</i>	<i>tsku</i>			
Meckels	63,0	1,0	129,2	388,0	93,0	2,1	144,7	4,5	688,7	33,4	118,5	199,1	152,1	1325,9	234,7	9,5	0,1	2,7	9,5	40,7	21,6			
Vertebrae	64,9	0,9	130,9	711,6	244,5		305,2	4,3	825,9	31,1	222,6	1939,3	246,9	2715,1	152,7	7,4	0,1	2,7	13,6	32,0	15,0			
Chondrocranium	46,7	1,3	54,2	550,2	29,9	1,0	56,1	3,6	259,9	26,3	124,0	46,1	136,9	704,3	90,1	15,2	0,1	2,6	17,4	24,3	18,6			
Denticles	18,7	0,9	30,0	208,0	22,9	0,5	6,4	4,3	93,6	21,6	59,6	1,9	58,7	12,0	0,5	9,9	0,0	2,0	1,1	14,0	33,4			
Dental lamina	32,6	1,3	130,6	170,1	33,0	0,5	7,6	5,2	211,3	32,9	48,5	2,6	39,5	4,0	1,8	6,6	0,0	2,8	1,7	16,2	50,3			
Brain+OS	10,0	1,8	3,4	34,1	2,3	0,7	7,5	4,6	6,9	14,6	13,8	2,3	4,4	3,3	2,7	16,1	0,2	2,2	1,1	12,3	21,5			
Spinal chord	8,9	1,2	3,3	98,3	2,0	2,8	7,5	3,7	18,1	10,7	27,6	2,5	3,0	13,3	4,7	16,5	0,0	2,1	5,3	9,3	8,8			
Eye	24,4	1,3	33,5	284,1	4,1	4,5	7,1	4,8	51,2	17,8	71,6	9,0	55,4	57,7	9,5	13,9	0,0	5,7	9,0	18,0	36,1			
AOL	45,6	1,0	26,5	1048,8	7,4	0,5	6,5	3,7	268,6	22,8	103,8	2,9	70,8	9,8	3,3	11,0	0,1	2,2	8,2	10,1	11,9			
Gills	9,5	1,1	4,6	151,3	2,1	0,5	6,3	4,8	6,1	13,2	25,6	2,3	6,8	1,3	2,0	13,8	0,0	2,4	1,1	11,8	27,0			
Blood	8,7	0,9	6,3	1,2	1,9	0,1	6,6	2,3	0,2	12,9	7,6	2,9	2,8	0,3	0,0	11,2	0,1	1,4	0,3	10,3	13,9			
Spleen	16,4	1,3	7,4	235,2	1,7	0,4	9,2	3,8	0,9	14,2	74,4	2,2	8,0	1,7	0,8	11,7	0,1	1,9	0,6	11,9	7,4			
Heart	13,9	0,5	8,0	322,9	1,8	0,3	4,1	2,2	26,0	7,1	75,2	1,5	25,8	4,0	3,7	47,4	0,0	1,0	7,1	10,0	3,9			
Muscle	11,5	0,3	4,1	119,1	9,9	2,2	2,5	0,9	37,4	6,5	28,0	0,8	41,2	3,7	0,6	4,5	0,1	0,7	5,4	3,3	3,3			
Esophagus	22,3	1,0	11,1	114,6	5,2	0,7	4,1	3,0	65,3	12,8	61,0	3,0	111,1	2,3	0,6	7,9	0,1	1,5	4,0	8,9	13,9			
Stomach	2,8	0,5	1,1	18,7	0,7	0,1	1,5	0,8	2,5	3,8	7,3	0,5	1,7	0,7	0,1	3,1	0,1	0,5	0,2	3,5	31,9			
Pancreas	4,3	0,2	0,6	22,4	0,6	0,1	1,9	0,4	1,3	1,4	5,1	0,3	6,4	0,2	0,3	1,3	0,0	0,2	0,2	1,6	1,5			
Liver	4,1	0,7	1,2	33,1	1,0	0,2	4,4	1,1	0,2	5,3	15,6	1,2	0,9	0,7	0,2	4,7	0,0	0,9	0,3	4,6	82,2			
Spiral intestine	5,9	1,0	1,2	20,8	1,6	0,1	4,3	2,4	0,3	9,5	13,0	1,7	1,8	0,5	0,1	1,7	0,1	1,1	0,4	10,3	108,8			
Rectal gland	13,2	1,3	6,0	186,6	3,7	0,1	5,6	2,7	12,7	11,8	31,4	1,8	26,6	6,0	12,1	23,3	0,1	2,1	0,9	12,1	4,9			
Kidney	14,7	2,1	5,2	69,4	1,9	14,5	9,1	3,1	2,3	12,5	31,3	2,3	5,2	4,6	4,0	12,4	0,0	2,0	2,6	10,5	29,6			
Seminal vesicle	9,7	0,9	4,7	115,8	1,3	0,2	5,3	1,8	9,3	8,6	14,1	0,8	5,9	0,8	10,5	6,1	0,0	1,3	0,5	8,0	4,8			
Ovary	1,6	0,3	0,5	6,3	0,6	0,4	0,8	0,4	0,2	2,3	2,5	0,4	1,7	0,2	0,6	0,8	0,0	0,2	0,2	1,8	11,8			
Uterus	13,0	0,7	5,0	106,8	1,8	0,3	5,6	3,6	3,0	10,9	19,2	2,5	20,9	0,8	0,7	6,7	0,0	1,6	0,8	10,4	66,6			
Testis	13,4	1,1	14,5	58,2	6,2	0,2	7,5	2,9	0,3	12,1	33,1	2,5	5,8	0,9	1,3	4,3	0,1	1,7	3,7	8,3	6,2			
E-St-12	7,5	1,0	1,2	1,0	3,5	0,2	6,0	2,9	1,0	14,8	9,5	1,3	1,3	0,5	1,2	1,8	0,1	1,3	0,3	9,6	22,1			
E-St-22	10,0	141,5	1,5	38,6	2,2	0,5	6,3	5,3	16,1	13,8	17,1	42,6	1,6	0,4	3,9	6,1	0,9	2,1	0,7	11,1	23,0			
E-St-24	13,4	182,2	2,6	119,7	6,2	0,5	12,1	9,0	8,6	17,3	18,9	39,4	2,8	0,9	4,2	9,4	1,1	3,1	1,7	13,9	15,3			
E-St-26	13,2	149,2	2,4	63,1	8,6	0,9	9,9	15,6	24,4	16,2	19,5	49,2	2,7	3,2	6,9	8,7	1,7	2,2	3,4	11,8	26,0			
E-St-30	26,4	4,3	33,0	134,3	50,6	5,9	16,8	48,7	119,1	93,3	51,3	43,5	91,8	37,9	17,4	10,6	0,1	3,3	10,3	12,3	23,0			
E-St-31	19,9	1,8	61,6	121,1	72,7	11,0	8,8	31,1	189,0	133,3	44,7	16,1	129,3	21,3	10,0	11,5	0,1	2,6	8,9	10,3	20,2			

**Table 4.1:** TPM values of canonical SLRPs. Samples with both Z-scores  $\geq 1$  and TPMs  $\geq 50$  are highlighted in orange, which indicate an enriched and strong expression in the sample. Samples with Z-scores  $\geq 1$  but TPMs  $\leq 50$  are highlighted in purple. AOL: ampullae of Lorenzini; E-St-X: embryo stage x (from Ballard et al 1993); OS: olfactory sac.



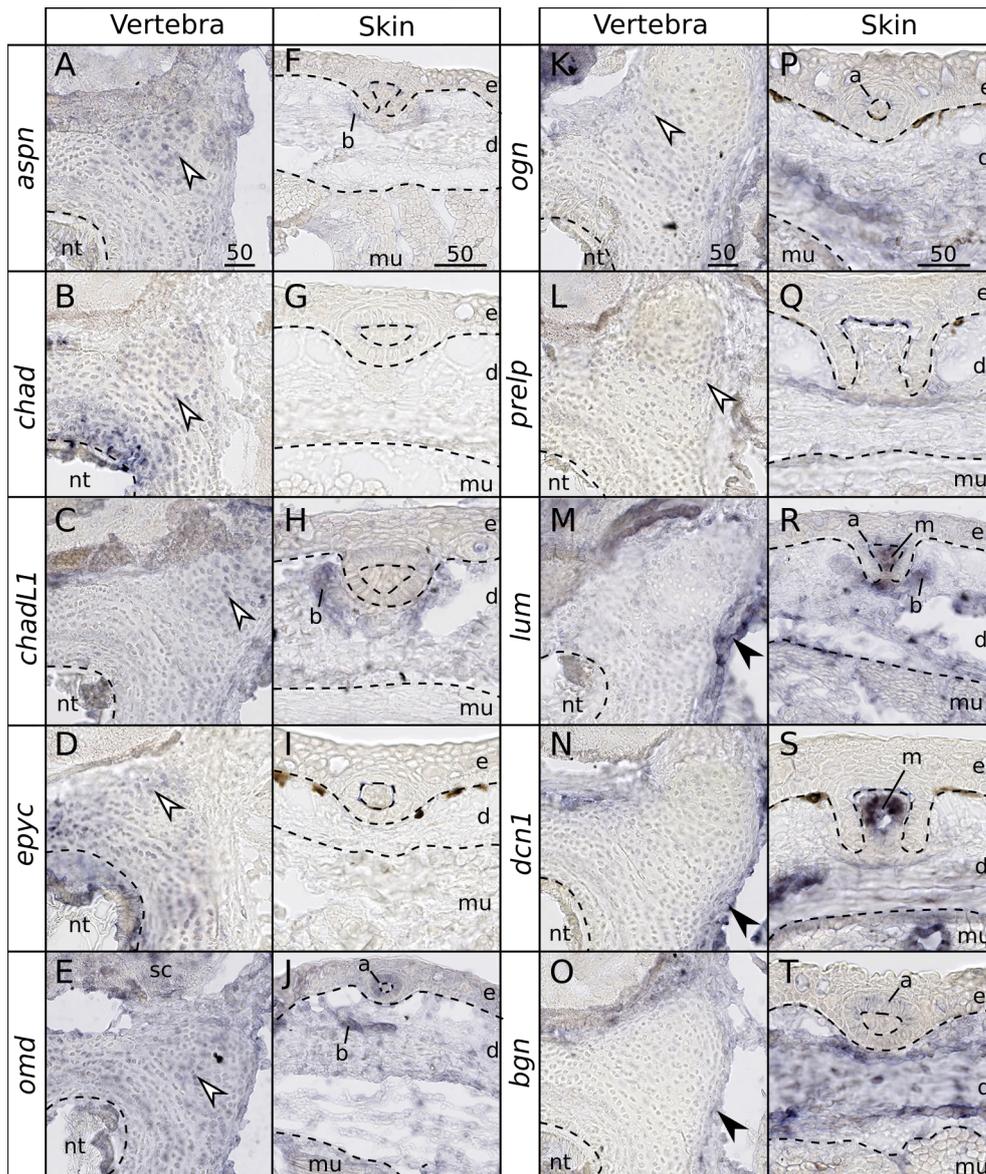
**Figure 4.4:** Histology and gene expression in 6.5 cm *S. canicula* embryo anterior cross sections. A-C: HES staining, D-F: *col1a1* in situ hybridization; G-I: *col2a1* in situ hybridization in vertebra (zoom 1; B, E, H) and skin layers (zoom 2; C, F, I). d: dermis; e: epithelium; m: mesenchyme; mu: muscle; nt: notochord; sc: spinal cord. Black arrowheads indicate expression in the perichondrium; White arrowheads indicate expression in a chondrocyte. Dotted lines mark separations between tissues. Scale bars are in  $\mu\text{m}$ .

### Cell-specific expression of SLRPs in developing skeletal tissues of the catshark

We tested the timing and location dynamics of genes expected to be expressed in endoskeletal tissues (vertebrae) first by relative qPCR measurement (Suppl Mat on demand) and then by *in situ* hybridization on sections of selected developmental stages. In 5- to 8-cm TL catshark embryos, skeletal tissues develop from poorly differentiated cell populations to a variety of differentiated tissues including: non-mineralized and mineralized cartilages, non-mineralized and mineralized perichondrium, mineralized fibrous sheath of the notochord and non-mineralized notochord (Figure 4 and Berio et al., 2021; Enault et al., 2016). Over this embryonic period, skin denticles develop as the dermis and epidermis differentiate; muscle tissue also differentiates (Figure 4). Tested genes were selected based upon TPM values (mean endoskeletal TPM  $\geq 30$ , Table 4.1) or results of qPCR amplification, therefore lowly expressed or not amplified genes were not selected, which was the case of *kera*, *ecm2L*, *dcn2*, *chadL2*, *podn*, *podnL*, *npc*, *nyx* and *tsku*.

No expression could be detected by *in situ* hybridization for *ecm2* and *fmod* (6,5cm TL, not shown). Other SLRPs' expression follow two main patterns comparable to either *col2a1* or *col1a1* expression (Fig 4.4 and Table 4.2) and therefore assigned to chondrocyte or perichondrial cells. The *aspn*, *chad*, *chadL1*, *epyc*, *omd*, *ogn* and *prelp* genes are expressed by chondrocytes

(Fig 4.5, A-E and K-L), while *bgn*, *dcn1* and *lum* display perichondrial cell expression (Fig 4.5, M-O). Additional expression in chordocytes of the notochord was observed for both *chad* and *epyc* and in neural cells of the spinal cord for *omd*. In developing scales, ameloblastic expression is apparent for *bgn*, *lum*, *omd* and *ogn* (Fig 4.5, J, P, R and T), while odontoblastic expression is observed for *dcn1* and *lum* (Fig 4.5, R and S). Additionally, expression was detected in the mesenchyme of scale roots for *aspn*, *chadL1*, *lum* and *omd* (Fig 4.5, F, H, J and R). Expression in the dermal cells is detected for *bgn*, *dcn1*, *lum*, *omd* and *ogn* (Fig 4.5, J, P and R-T). Finally, muscle expression was detected for *dcn1*, *lum*, *omd* and *bgn* (Fig 4.5, J and R-T).



**Figure 4.5:** Expression data in 6.5 cm *S. canicula* embryo anterior cross sections. : mRNA *in situ* hybridizations of several canonical SLRPs in vertebrae (A-E and K-O) and skin layers (F-J and P-T). a: ameloblast; b: scale base; d: dermis; e: epithelium; m: mesenchyme; mu: muscle; nt: notochord; sc: spinal cord. Black arrowheads indicate expression in the perichondrium; White arrowheads indicate expression in a chondrocyte. Dotted lines mark separations between tissues. Scale bars are in  $\mu\text{m}$ .

Table 4.2	Collagen I or II expression	SLRP expression
Nervous system	-	<i>omd</i>
Notochord	-	<i>chad, epyc</i>
Cartilage	<i>col2a1</i>	<i>aspn, chad, chadL1, epyc, ogn, omd, prelp</i>
Perichondrium	<i>col1a1</i>	<i>bgn, dcn1, lum</i>
Ameloblasts	-	<i>bgn, lum, ogn, omd</i>
Odontoblasts	<i>col1a1</i>	<i>dcn1, lum</i>
Scale base	<i>col1a1</i>	<i>aspn, chadL1, lum, omd</i>
Dermis	<i>col1a1</i>	<i>bgn, dcn1, lum, ogn, omd</i>
Muscle	<i>col1a1</i>	<i>bgn, dcn1, lum, omd</i>

**Table 4.2:** Recap of *slrp* expression in *S.canicula* tissues / cells

#### 4.2.4 Discussion

##### Reevaluation of vertebrate SLRP gene repertoire

Chondrichthyans and their phylogenetic placement have proven valuable for inferring general and specific properties regarding the evolution of vertebrate gene families (Debiais-Thibaud, 2019; Leurs et al., 2021; Romero et al., 2021). In this study we recovered previously described orthology groups with high confidence values and discovered several new vertebrate members of the SLRP family, one canonical *Dcn2*, and two non-canonical: *NpcL* and *ChadL2*. We showed that *EcmXL* is not a vertebrate SLRP orthology group, but is instead a teleost specific gene whose emergence is the consequence of a third WGD event, similar to results published by Costa et al. regarding *ogn* (Costa et al., 2018). From their branching to canonical SLRPs, all non-canonical SLRP clades emerged before the vertebrate 2R but our results do not allow for a more accurate understanding of their evolutionary history. A phylogenetic analysis including the five most similar cephalochordate sequences (*Branchiostoma floridae*) did not help resolve deep nodes in our phylogeny nor proposed orthology relationships between the amphioxus and vertebrate sequences, questioning the belonging of these sequences to the SLRP family (Supp Mat S1 and S8). SLRP-related sequences in *Ciona* are scarce (Jones et al., 2007; Park et al., 2008) suggesting that SLRPs have undergone extensive evolution, both by coding-sequence evolution and gene duplication events, in the period just pre-dating the 2Rs in vertebrates.

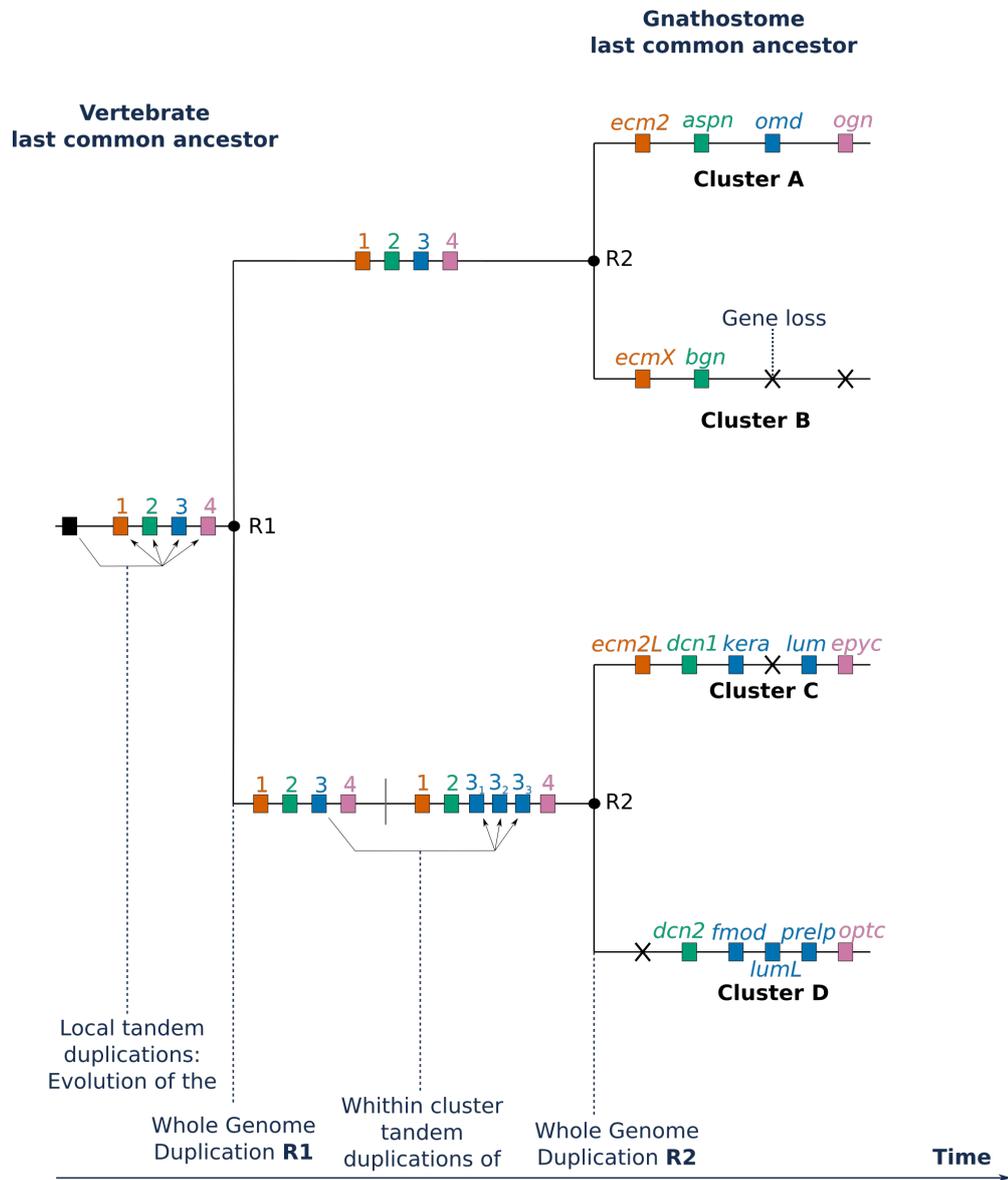
The lamprey *P. marinus* was used as the sister group of jawed vertebrate sequences but its sequences do not show a 1:1 orthology relationship with the rest of the vertebrate sequences, congruent with the hypothesis of parallel second WGD in cyclostomes and gnathostomes (L. Z. Holland and Daza, 2018; Kuraku et al., 2008; Sacerdot et al., 2018; Smith and Keinath, 2015; Smith et al., 2018). Out of the nineteen lamprey SLRPs recovered, twelve of them branch within canonical clades. Out of those, ten have a standard LRRCE motif (Pm1-8, 10 and 12; not shown). *P. marinus* SLRPs contain a variable number of LRRs ranging from six to twenty-five. This number variation, however, is coherent with the results on the catshark sequences (i.e. closely related sequences have a similar number of LRRs). In the sea lamprey genome, several gene tandems are identified. First, SLRP5 is found with SLRP6 (respectively clade 1

and 3 genes); and SLRP3 with SLRP12 (respectively clade 2 and 4 genes) which supports a vertebrate-ancestral cluster of genes from clades 1 to 4. This suggests that the SLRP clusters identified in extant species are derived from a single four-gene ancestral cluster that predates the vertebrate 2R events.

### Evolutionary scenario for canonical SLRPs expansion in vertebrates

The clustering of gnathostome canonical SLRPs along four different loci and the branching of paralogs in clade phylogenies are all congruent with the expected pattern for cluster multiplications by the 2R whole genome duplications (Dehal and Boore, 2005) following a ((A,B),(C,D)) relation. For instance, within clade 2, *aspn* (cluster A) and *bgn* (cluster B) are closer to each other, while *dcn1* (cluster C) and *dcn2* (cluster D) are closer together (Fig 4.3). This pattern is also visible for clades 1 and 4 if we consider gene loss for the sister genes to *ecm2L* and *ogn* respectively (Fig 4.3 and Fig 4.6). However, this is not the observed topology for clade 3 that cannot be explained using only two rounds of WGD and gene loss. To explain the several additional paralogs of clade 3 observed in clusters C and D, we propose a most-parsimonious scenario where two serial events of tandem duplication occurred between the two rounds of WGD from a single ancestral copy (Fig 4.6). Before the second round of WGD, one ancestral cluster with four SLRPs duplicated into clusters A and B, and another ancestral cluster with six SLRP paralogs gave rise to clusters C and D (Fig 4.6). Subsequent gene loss (and additional WGD in teleosts) would explain the genomic organization of extant gnathostomes and the inferred phylogenies. It is relevant to mention that the synteny markers on either side of the SLRP clusters are annotated as genes belonging to the same gene families (namely *atp2b* and *btg*). A recent study showed a clear case of clustered, tandemly duplicated genes diversifying through the 2R-event in the zone neighboring SLRPs with similar ((A,B),(C,D)) relationship (Daza et al., 2022), making the excellent conservation of the gnathostome SLRP loci a wider characteristic of a whole chromosomal section.

In the sea lamprey, two tandems each made of two clade 3 genes can be found in two different loci: SLRPs 7 with 10, and SLRPs 8 with 9, also supporting a vertebrate ancestral tandem duplication of a clade 3 gene along one of the ancestral clusters. The lamprey sequences branch together in the phylogeny (Fig 4.3) as (7+8) and (9+10), which is the expected topology for the products of the cyclostome specific WGD. The cluster with internal clade 3 duplication therefore subsequently underwent the 2nd WGD (independently in the cyclostome and gnathostome lineages). The ancestral vertebrate four-gene cluster has therefore expanded first by means of clade 3 tandem duplications and then by a parallel 2nd WGD event in both gnathostomes and cyclostomes (Fig6).



**Figure 4.6:** Hypothesized scenario for canonical SLRP evolution in gnathostomes. Color code is the same as in Figure 4.1. X indicates gene loss.

### Gnathostome-conserved and *S. canicula* derived features of SLRP expression in skeletal tissues

Two main types of SLRP expression patterns can be observed in embryonic endoskeletal tissues: perichondrium or chondrocyte-associated expressions. For instance, *bgn*, *dcn1* and *lum* all show a perichondrium-associated expression pattern congruent with these genes being described to interact with collagen I in mice (Fig 4.5; S. Chen and Birk, 2013). The chondrocyte expression of *chad*, *chadL1*, *epyc*, *omd*, *ogn* and *prelp* (Fig 4.5) is a conserved feature between the catshark and bony fishes (Funderburgh et al., 1997; Grover and Roughley, 2001; Shinomura and Kimata, 1992; Tillgren et al., 2015; Wilda et al., 2000). A major exception to this conservation status is *aspn*, that was shown to interact with Collagen type I in mice (Kalamajski et al., 2009) and to be expressed in the perichondrium but not in differentiated cartilage (Henry et al., 2001). In the catshark however, *aspn* was only detected in chondrocytes and not in the perichondrium, suggesting divergent evolution of *aspn* expression patterns. In addition, the catshark Omd protein sequence is one of the most divergent SLRPs when comparing it to its bony fish ortholog as: it showed no GAG attachment sites but six N-glycosylation sites; there were no detected putative tyrosine sulfation sites while several are found in bony fishes; no acidic C-terminal region was found. Its expression was located in chondrocytes in the embryonic catshark while it has been detected in mice osteoblasts (Ninomiya et al., 2007) and in rat fetal femur bone with a function in binding hydroxyapatite (Wendel et al., 1998). Since chondrichthyans lost the ability to make bone and have novel modes of mineralization (Seidel et al., 2016), evolution of the *aspn* and *omd* expression patterns and sequences could be linked to this skeletal transition. The *bgn* and *lum* genes are expressed in mice chondrocytes and known to interact with Collagen II and Aggrecan (S. Chen and Birk, 2013; Wilda et al., 2000) but their expression was not detectable in the cartilage in the catshark (Fig 4.5). Expression in catshark embryonic exoskeletal tissues in this study was focused on dermal denticles that develop similarly to teeth (Debiais-Thibaud et al., 2015). The observed patterns in the catshark can be ameloblast-associated (*bgn*, *ogn*, *omd*), odontoblast-associated (*dcn1*) or both (*lum*; Fig 4.5). In mice teeth, *bgn* and *omd* were detected in both ameloblasts and odontoblasts while *aspn*, *dcn1*, *lum*, *ogn* and *fmod* were only found in odontoblasts (Buchaille et al., 2000; Hou et al., 2011; Houari et al., 2014; Matsuura et al., 2001; Randilini et al., 2020). Therefore, we can speculate on a conserved function of *bgn* and *omd* in enamel/enameloid formation and of *dcn1* and *lum* in dentin formation. Odontoblasts, the dentine-producing cells, show more differences in SLRP expression between bony and cartilaginous fishes than ameloblasts, which is surprising as dentine is considered a stable tissue in vertebrates, while the evolution of extant enamels and enameloids are considered more derived tissues (Kawasaki and Weiss, 2008). The odontoblasts are thought to be very similar to osteocytes in their differentiation and secretion pathways and again, the lack of *aspn* expression in catshark odontoblasts might be a signature of bone loss consequences.

### Specific SLRP evolutionary trends in cartilaginous species

From our phylogenetic data, *Dcn2* was kept only in chondrichthyans and lost in all other gnathostome lineages. Screening of the protein motifs showed high similarity with its paralog *Dcn1* which is itself very similar to its bony fish orthologs (Kalamajski and Oldberg, 2010; Low et al., 2021). However, *dcn2* shows very low TPM values which coupled with its sequence similarity with *dcn1* may support functional redundancy between these genes. However, *dcn2* expression enrichment in the catshark kidney is not found for *dcn1*, which may be a reason for its maintenance in chondrichthyan genomes. In contrast to other SLRPs, *ecm2L* and *kera* expression is enriched in early embryonic stages. However, *ecm2L* is unknown from functional studies since it was lost in amniotes and the catshark ortholog is very peculiar since we found eighteen putative N-glycosylation sites before the LRRs. Functional implications of these features remain unknown since there is no point of comparison. However, *kera* is quite similar to its ortholog in bony fishes and has been shown to be expressed very early during chick embryogenesis with possible roles in neural-crest cells migration (Conrad and Conrad, 2003). An enrichment of expression for *ecm2*, *bgn*, *lum*, *ogn*, *prelp* was detected in the ampullae of Lorenzini (AOL), which are a cartilaginous fish evolutionary innovation: they are sensory organs that can detect electromagnetic fields and temperature gradients, and are filled with a keratan sulfate rich gel. Previous studies on AOL did not identify the proteoglycans to which these keratan sulfate chains might be linked to (Melrose, 2019; X. Zhang et al., 2018), but specifically *lum*, *ogn*, *prelp* have been reported to carry keratan sulfate GAGs in mammals (Funderburgh et al., 1997; Hultgårdh-Nilsson et al., 2015; Iozzo and Schaefer, 2015). Our predictions of glycosylation sites in the catshark SLRP sequences recovered conserved positions, including those where keratan sulfate attachment was shown in mammals (Low et al., 2021; Suppl Mat on demand). Therefore, *lum*, *ogn* and *prelp* are excellent candidates for structural SLRPs involved in the secretion of the specialized gel of the AOL in chondrichthyan fishes. More surprising are *bgn* and *ecm2* expression in the AOL. The catshark *ecm2* is divergent when compared to its bony fish ortholog: it is predicted to have five possibly sulfated tyrosines in the N-terminal region (none predicted in mice), four putative GAG attachment sites (only one in mice) and three N-glycosylation sites (none in mice), all possible sites for keratan sulfate attachment. This SLRP might carry keratan sulfate chains in cartilaginous fishes, but these results are highly speculative since Ecm proteins have been very poorly studied in bony fishes, and our data are only predictions which are more permissive than the actual observation of GAG attachment (Kalamajski and Oldberg, 2010). A better characterization of GAG chain linkage of chondrichthyan SLRPs is still critical to better understand lineage specific similarities or differences associated with the function of these proteoglycans.

### 4.2.5 Concluding remarks

Our results support a major expansion of the canonical SLRP gene repertoire firstly by tandem duplications in early vertebrates predating the two events of whole genome duplication, then by the 2R events in extant lineages of vertebrates. This expansion co-occurred with the great diversification of connective tissues, including skeletal tissues, that characterize the vertebrate lineage. Additionally, specific SLRP expression in cartilaginous fish evolutionary innovations such as the ampullae of Lorenzini, highlights the potential modularity of these genes by being recruited for the function of new structures. More gene expression data from the cyclostome lineage might help understanding the deeper conservation of SLRP function and therefore the evolution connective tissue at the origins of vertebrates.

### Acknowledgements

We acknowledge the MRI platform (ANR-10-INBS-04) for imaging support, the GenSeq platform, Labex CEMEB (ANR-10-LABX-0004) and NUMEV (ANR-10-LABX-0020) for sequencing support and the RHEM facility (INCa.Inserm.DGOS.12553, FEDER-FSE 2014-2020 Languedoc Roussillon) for histology expertise. We thank H el ene Mayeur and Sylvie Mazan for discussions on RNAseq data analysis.

### Funding

Research was supported by the LabEx CeMEB, and ANR "Investissements d'avenir" program (ANR-10-LABX-04-01).

## 4.3 Expansion of Mmp proteins in cartilaginous fishes

*Nicolas Leurs, Amirhossein Karimizadeh, Mélanie Debiais-Thibaud and Camille Martinand-Mari*

Institut des Sciences de l'Evolution de Montpellier, ISEM, Univ Montpellier, CNRS, IRD, EPHE, Montpellier, France.

### 4.3.1 Introduction

Matrix metalloproteinase (Mmps) are a group of calcium-dependent zinc-containing endopeptidases that have been shown to cleave both extracellular matrix (ECM) and non-ECM proteins (Paiva and Granjeiro, 2014). Mmp proteins, found in all kingdoms of life, are involved in many vital biological processes such as matrix remodeling (Stamenkovic, 2003), cell differentiation (Brizzi et al., 2012; George, 2004), apoptosis (Si-Tayeb et al., 2006), cell migration (Gifford and Itoh, 2019), cell motility (Shiple et al., 1996), cell proliferation (Sans-Fons et al., 2010), cell cycle (Golubkov et al., 2005) and tissue repair (reviewed in Paiva and Granjeiro, 2014). In humans, twenty-three Mmp proteins have been identified and, based on their substrate affinity and tridimensional structure, are divided into six subgroups, which are collagenases (Mmp1, Mmp8, and Mmp13), gelatinases (Mmp2 and Mmp9), matrilysins (Mmp7 and Mmp26), stromelysins (Mmp3, Mmp10 and Mmp11), membrane type metalloproteinases (Mmp14-17, Mmp24-25), and others (Mmp12, Mmp19-21, Mmp23, Mmp27, and Mmp28) (Nagase et al., 2006).

Some Mmps have been shown to be involved in bone and cartilage development, ossification and bone remodeling (Paiva and Granjeiro, 2014). This is particularly relevant because these proteins may potentially contribute to lineage specific differences between bony and cartilaginous fishes. Bony fish cartilage and bone development involves a variety of processes intimately linked to degradation and regeneration of the extracellular matrices of these connective tissues. Cartilaginous fishes however, have lost the ability to make bone (Gillis, 2019); but show a derived form of mineralized cartilage, the tessellated cartilage, and a fibrous perichondrium in the neural arches, which has been described as a bone-like tissue (Berio et al., 2021). The genetic control of the development of these structures, however, is still poorly known.

Mmp proteins have also been shown to function in regeneration of injured limb in amphibians (Santosh et al., 2011; Vinarsky et al., 2005). Recently, in this lineage, a considerable expansion of a subclade of Mmps has been demonstrated (Baddar et al., 2021). Chondrichthyans display continuous growth throughout their life and recently the little skate *Leucoraja erinacea* has been shown to have the ability to regenerate its cartilage (Gillis, 2019; Marconi et al., 2020), which raises the question of the state of Mmp repertoire in chondrichthyans fishes.

Previous studies on Mmp family are biased towards bony fishes (Fanjul-Fernández et al., 2010), probably due to the lack of available sequence data on chondrichthyans. Here, we take advantage of newly available high-quality genomic data in the cartilaginous fish lineages to

assess the evolution of the *mmp* gene family at the vertebrate scale. In this context, phylogeny and synteny analyses have been performed to identify lineage-specific losses or duplications and RNAseq data was used to evaluate the divergence of expression patterns in a chondrichthyan: the small-spotted catshark *Scyliorhinus canicula*.

### 4.3.2 Material and Methods

#### Phylogeny

Mmp protein sequences from 6 species of tetrapods (*Homo sapiens*, *Xenopus tropicalis*, *Mus musculus*, *Gallus gallus*, *Cynopus pyrrhogaster*, *Amblystoma mexicanus*), 6 species of chondrichthyans (*Scyliorhinus canicula*, *Amblyraja radiata*, *Callorhinchus milii*, *Rhincodon typus*, *Carcharodon carcharias*, *Chiloscyllium plagiosum*) and 5 species of actinopterygians (*Erpetoichthys calabaricus*, *Lepisosteus oculatus*, *Danio rerio*, *Astyanax mexicanus*, *Oryzias latipes*), and one species of cyclostome (*Petromyzon marinus*) were downloaded from both NCBI (chondrichthyans and actinopterygians) and from Al Haj Baddar et al. 2021 (tetrapods). The searches on NCBI were performed using tBLASTN with *Mus musculus* Mmps as a query due to the confidence in the gene annotation and quality of data available on NCBI for this species. The identified Mmps were subsequently used to further retrieve lineage specific sequences (i.e. chondrichthyan Mmps blasted onto chondrichthyan genomes). Mmp protein sequences were also recovered by blasting on a *Raja clavata* jaw transcriptome generated in the lab (Debiais-Thibaud, 2019) using TBLASTN from the Blast+ software (Camacho et al., 2009). Reciprocal BLASTs were performed on *Amblyraja radiata*, to insure their belonging to the Mmp family. The CDS sequences were translated into amino acids using expasy translation tool (Gasteiger et al., 2003). We removed the human Mmp26 sequence because this sequence is human specific and did not align well with other Mmps.

To understand the evolutionary history of Mmp proteins in vertebrates, we performed a phylogenetic analysis using the sampled protein sequences. The 429 sequences were aligned using Mafft (Kato and Standley, 2013; `-auto -reorder`) and subsequently cleaned using Hmm-Cleaner (Franco et al., 2019; `-large`) to remove low similarity segments leading to an alignment with 1138 positions. The appropriate amino-acid evolution model for maximum likelihood analysis was found using ModelFinder (Kalyaanamoorthy et al., 2017). The best model was chosen using the lowest BIC (Bayesian information criterion). Phylogenetic reconstruction was performed using Maximum likelihood in IQ-TREE v1.6.1 (Nguyen et al., 2014; `-alrt 1000 -bb 5000 -bcor 0.9`). Node support was evaluated using 5000 ultra-fast-bootstraps (Hoang et al., 2017) and SH-alrt single branch tests (Guindon et al., 2010). Minimum correlation coefficient for UF bootstrap convergence criterion was set to 0.90.

## Synteny

In order to understand the homology relationships among several Mmp clades, we performed a synteny analysis. Syntenic data was retrieved from reference genomes of *Scyliorhinus canicula* (GCF\_902713615.1), *Carcharodon carcharias* (GCF\_017639515.1), *Amblyraja radiata* (GCF\_01-0909765.2), *Chiloscyllium plagiosum* (GCF\_004010195.1), *Callorhinchus milii* (GCF\_00016504-5.1), *Erpetoichthys calabaricus* (GCF\_900747795.1), *Lepisosteus oculatus* (GCF\_000242695.1), *Danio rerio* (GCF\_000002035.6), *Mus musculus* (GCF\_000001635.27), and *Homo sapiens* (GCF\_000001405.40). In some cases the genomes of *Rhincodon typus* (GCF\_021869965.1) and *Xenopus tropicalis* (GCF\_000004195.4) were also used.

## RNAseq

*S. canicula* tissue libraries from whole embryos at stages 12, 22, 24, 26, 30 and 31 (Ballard et al., 1993) and adult tissues sampled from three individuals were recovered from NCBI (BioSample accessions: 21694013 to 21694062) and are as following: “Tissue (biological replicates)” Ampullae of Lorenzini (2), Blood (2), Brain+Olfactory sac (3), Chondrocranium (3), Dermal denticles (3), Dental lamina (3), Esophagus (1), Eye (2), Gills (1), Heart (1), Kidney (1), Liver (1), Meckel’s cartilage (3), Muscle (2), Ovary (1), Pancreas (2), Rectal gland (2), Spinal cord (1), Spiral intestine (2), Spleen (1), Stomach (1), Seminal vesicle (1), Testis (2), Uterus (1) and Vertebrae (2). The paired end RNAseq reads from adult and embryonic *S. canicula* tissue samples were mapped onto the reference gene model (Mayeur et al., 2021) using BWA MEM (H. Li, 2013). After sorting and converting the resulting alignments to BAM files using samtools, counts for each library were extracted. Count data was then normalized as Transcripts Per Kilobase Millions (TPMs) and used as such for quantification of expression. When a tissue had several replicates (max 3 replicates), TPM values were averaged, leading to a final number of 31 tissues compared. For each gene, Z-scores were then calculated to evaluate tissue-biased expression as

$$Z = (x - \mu) \cdot \frac{\sqrt{n}}{\sigma} \quad (4.2)$$

Where for any sample:  $x$  is the observed TPM value in a given tissue,  $\mu$  is the mean of all tissues,  $\sigma$  is the standard deviation of all tissues and  $n$  is the number of tissues.

### 4.3.3 Results

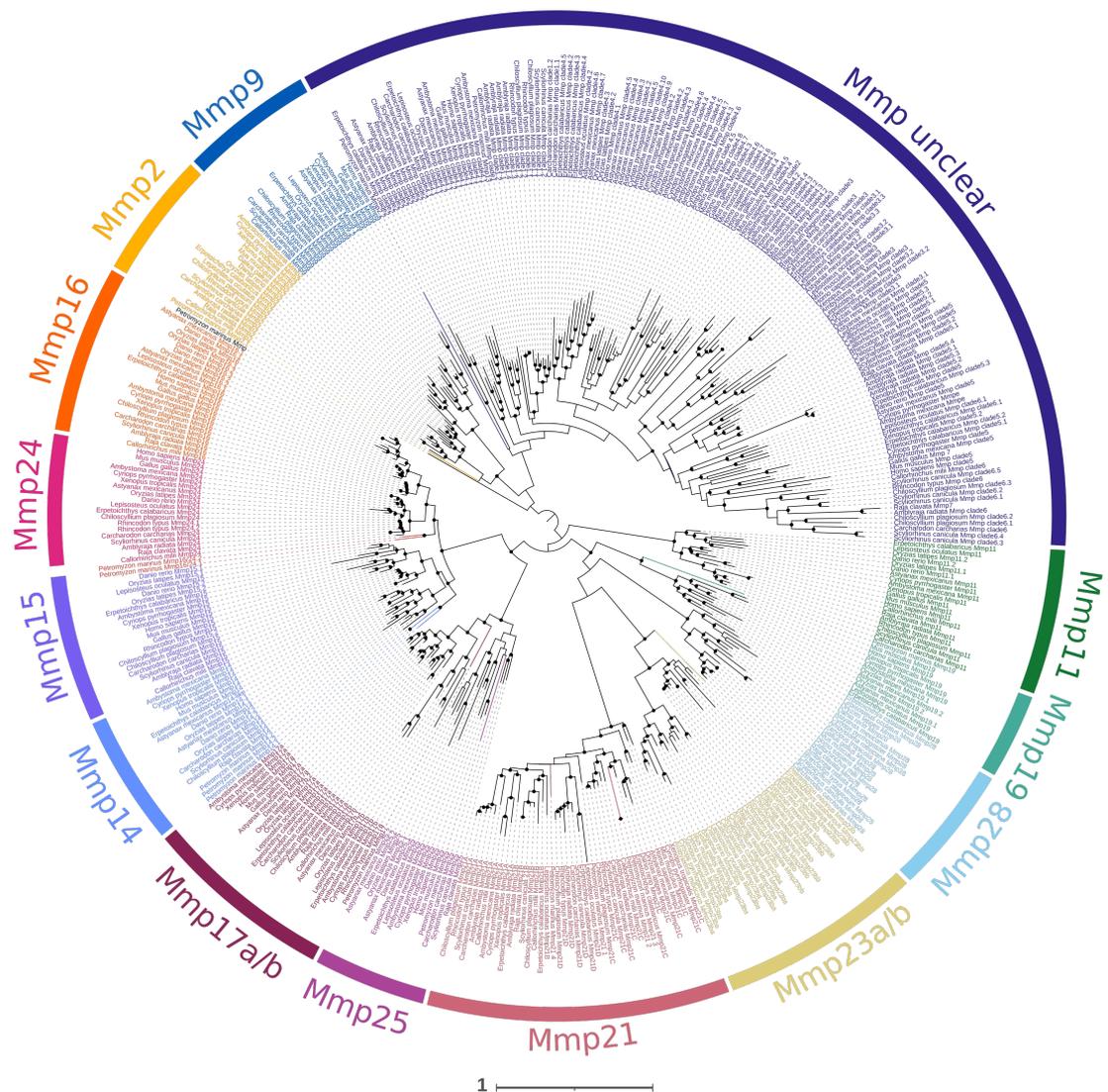
#### Evolution of Mmp genes in jawed vertebrates and their genomic organization

In total 429 sequences of Mmp proteins were used to make the complete phylogenetic tree (Fig 4.7, and Suppl Mat on demand). Since there is no identified outgroup for the complete tree, we used midpoint rooting (the root is at the midpoint of the most divergent taxonomic

group). This method is useful when an outgroup is not known (Hess and Russo, 2007). Overall, the phylogeny shows conserved monophyletic groups with 1:1 orthology relationships, at the gnathostome levels for most Mmp paralogs (e.g. Mmp2, Mmp9, Fig1) making the original/ancestral gnathostome Mmp repertoire of at least 14 paralogs. However, this is not true for Mmp21, where we find four chondrichthyan copies for only one in humans. Additionally, a second big clade of Mmps shows internal relationships with low robustness. Within this clade relationships between orthologs are quite complex, e.g. many specific duplicates at different taxonomical levels and a more detailed phylogeny was needed. In the following, this clade is called the “clustered” clade (Fig 4.7) because the human sequences present in this Mmp clade (denoted Mmp2, Mmp3, Mmp7, Mmp8, Mmp10, Mmp12, Mmp13, Mmp20, and Mmp27 in this species), show a clustered genomic organization. Additionally, Mmp17b and Mmp19 families do not have chondrichthyan orthologs suggesting that these genes have been lost in the chondrichthyan ancestor. Most teleosts have two copies for Mmp11, Mmp14, Mmp15, and Mmp16, probably resulting from their additional 3R of whole-genome duplication (Suppl Mat on demand).

We generated two new alignments and phylogenies by subsampling sequences from the Mmp21 and the clustered Mmp clades. The phylogeny for the Mmp21 clade (Fig 4.8) shows four gnathostome groups of orthology here named A, B, C, and D. Only Mmp21C includes members of all major gnathostome lineages. Mmp21A, however, was secondarily lost in amniotes while Mmp21B and D were lost in both tetrapods and teleosts (Fig 4.8). Syntenic analyses of these clades show conserved genetic markers between species and for each locus (Suppl Mat on demand). Only chondrichthyans and non-teleostean actinopterygians have kept all four Mmp21 paralogs.

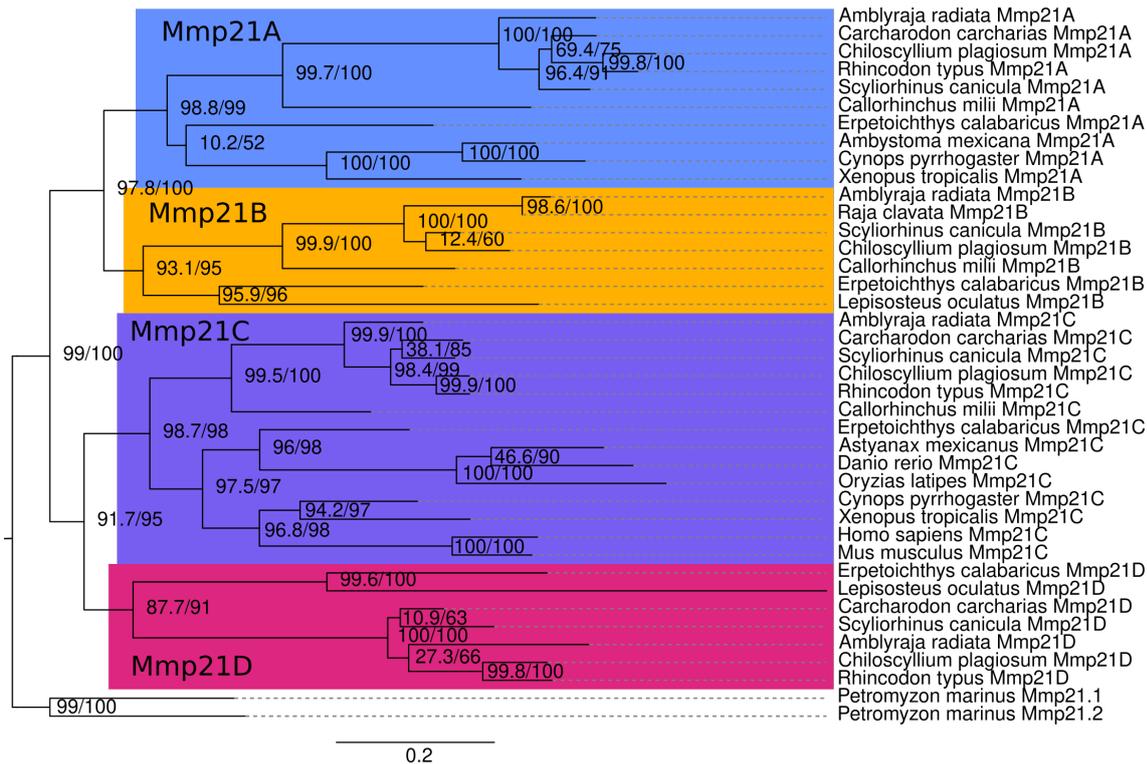
In the clustered Mmp phylogeny, we isolated six different gnathostome orthology groups (named clades 1 to 6; Fig 4.9). Mmp clade 1 is only made of chondrichthyan sequences (Fig 4.9). This group seems to have been lost in the last common ancestor of osteichthyans. It also includes recent duplication events specific to *Scyliorhinus canicula*, *Carcharodon carcharias* and *Amblyraja radiata*. Additionally, a duplication event specific to both *Rhincodon typus* and *Chiloscyllium plagiosum*, suggests that this event happened in the last common ancestor of orectolobiformes. Mmp clade 2 shows that this gene has been lost in sarcopterygians but kept in other lineages, albeit with an unexpected placement of *Callorhinchus milii*, with low support values. Mmp clade 3 consists of sequences of all major gnathostome clades. However, the topology suggests that a duplication of this gene happened in the ancestor of actinopterygians, and that one of the copies was lost in sarcopterygians (Fig 4.9). Mmp clade 4 lacks chondrichthyan orthologs, meaning that this clade has been completely lost in chondrichthyans. In contrast, this clade presents multiple duplications in bony fish and up to seven copies in humans (Fig 4.9). In particular, this is the clade where Baddar et al., 2021 found the expansion of Mmps in amphibians, and seems to be made up of several independent duplications more generally in actinopterygians. In Mmp clade 5, there is a specific duplication in *Scyliorhinus*



**Figure 4.7:** Vertebrate Mmp phylogeny. The alignment was 429 sequences with 1138 positions and the model for amino acid evolution used was LG+I+G4. Nodes with bootstraps  $\geq 90$  are represented by black dots. The phylogeny was rooted at the midpoint. Colored branches correspond to lamprey sequences

*canicula*, and in the ancestor of rajidae. Two additional duplications of the Mmp clade 5 sequence happened leading to four sequences in *Amblyraja radiata*, while *Callorhynchus milii* has also experienced two duplications (Fig 4.9). Mmp clade 6 consists of genes belonging to chondrichthyans and non-teleostean actinopterygians with multiple independent duplications (except in *Callorhynchus milii*). *Scyliorhinus canicula* has five copies, *Chiloscyllium plagiosum* has two and *Carcharodon carcharias* has only one (Fig 4.9).

In mice and humans, these *mmps* are all located between two genetic markers: *dcun1d5* (mouse) or *pdgfd* (human) and *cfap300* (Fig 4.10). These markers are also found in chondrichthyans although the chromosomal portion seem to have moved farther at some point in

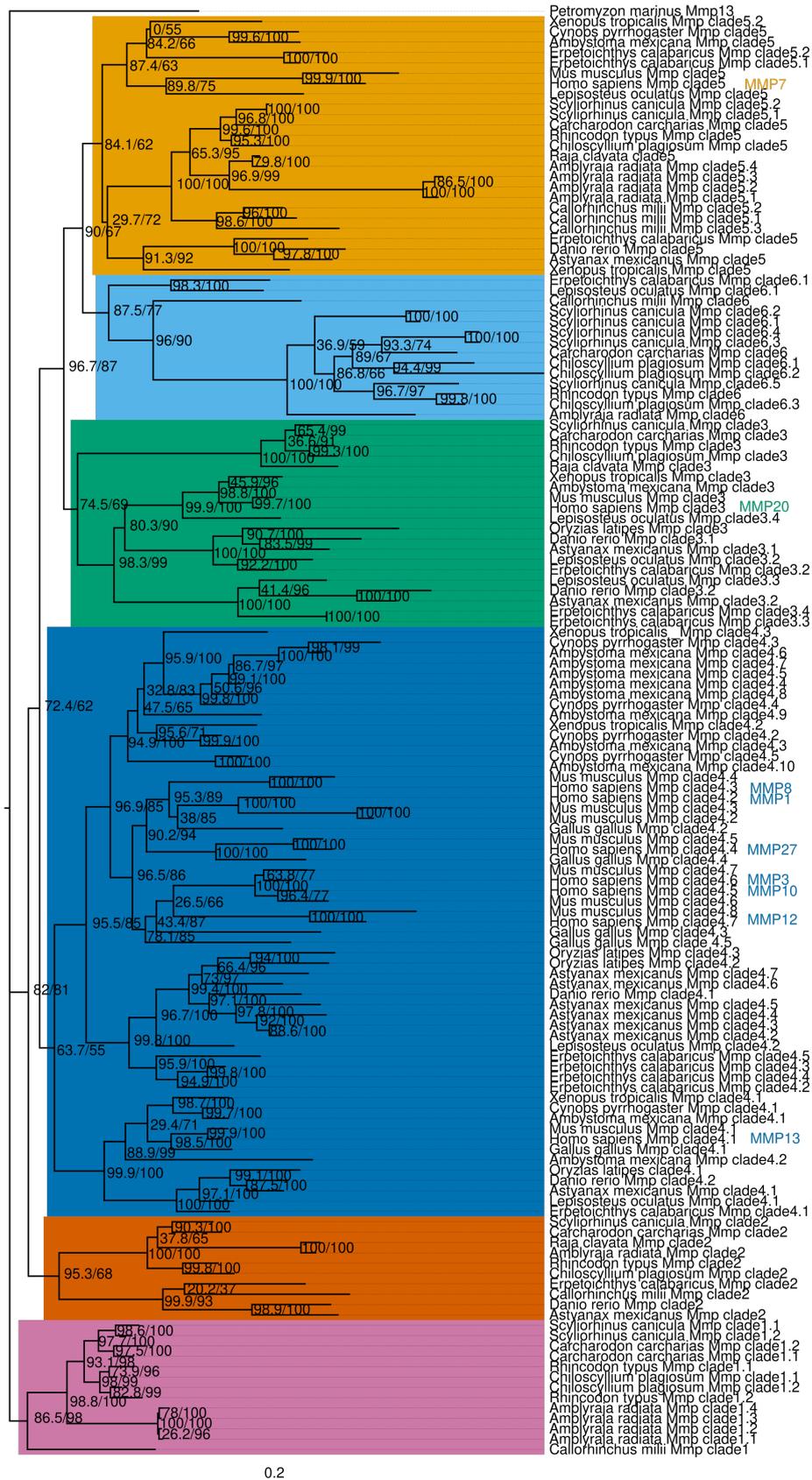


**Figure 4.8:** Mmp 21 phylogeny. The alignment contained 40 sequences and was 641 aa long. The model for amino acid evolution used was JTT+I+G4. The single Human *mmp21C* sequence corresponds to the actual human *mmp21*

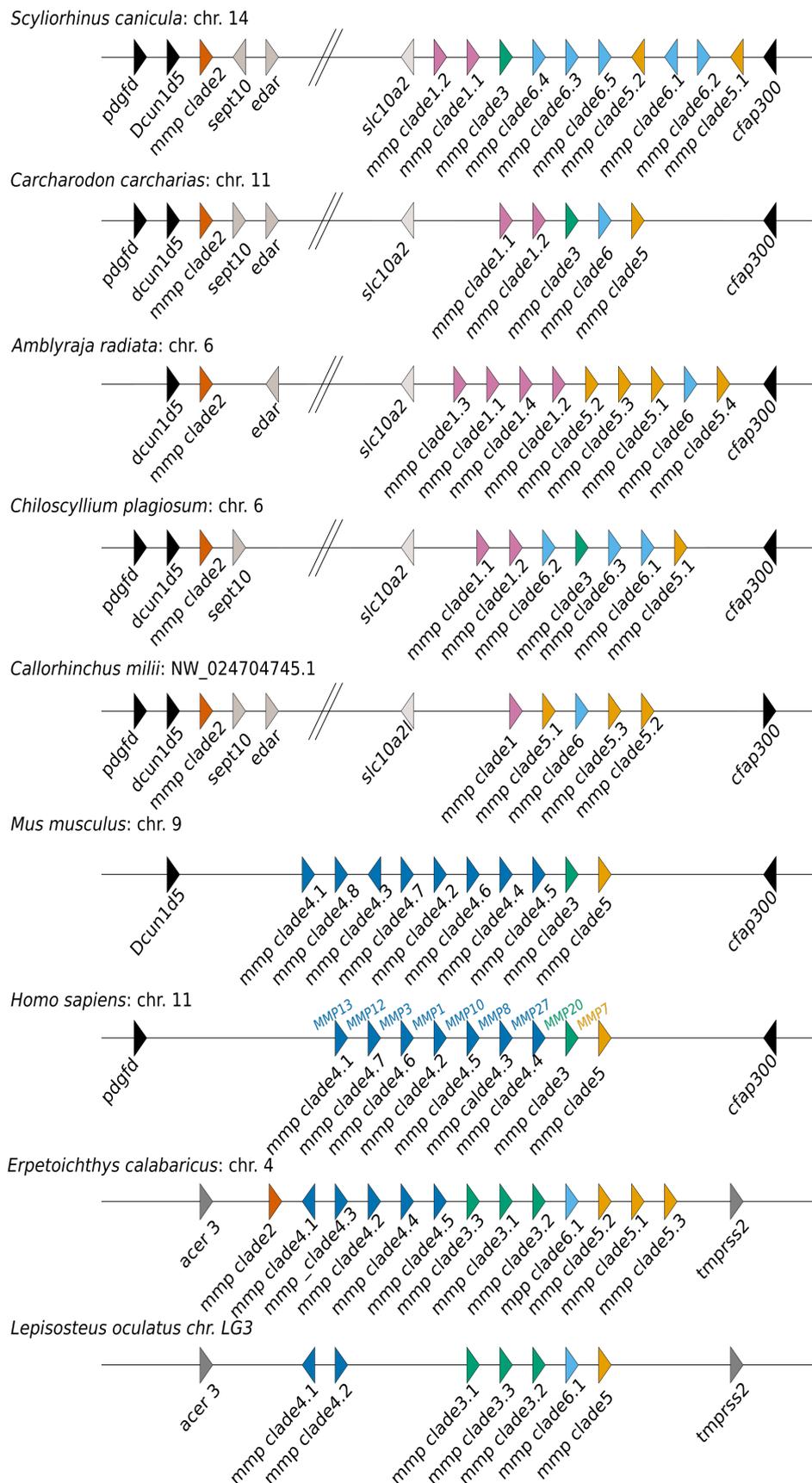
their evolution (Fig 4.10). Nonetheless, mmps are found near these markers, belonging to the clustered Mmp clade, and most of them (on the *cfap300* side) show a clustered organization similar to what is observed in mammals. In non-teleostean fishes a cluster is also observed but genetic markers are not conserved in these species as indicated by different genes instead: *acer3* and *tmprss2* (Fig 4.10).

### Tissue- and embryonic stage-specific expression of *mmp* genes in the small-spotted catshark

Tables 4.3 and 4.4 report RNAseq expression data for every one of the twenty-six identified *S. canicula mmp* genes. First, among the *S. canicula mmp* genes whose orthology relationship to other vertebrates is 1:1, *mmp2*, *mmp9* and *mmp14* are the most highly expressed ones having TPM values  $\geq 50$ . All three of these genes are over-represented in all skeletal tissues with the exception of *mmp9* who is specific to odontodes and the dental lamina. Additionally, *mmp2* and *mmp14* are highly expressed in sensory organs such as the eye (*mmp2*) and the ampullae of Lorenzini (*mmp2* and *mmp14*). These genes are also over-expressed in the circulatory system with *mmp2* in the gills and the heart while *mmp9* shows up in blood and the spleen along with *mmp14*. *Mmp2* and *mmp14* are both over represented in the esophagus while *mmp9* seems to be specifically expressed in testis. Only *mmp14* is found to be highly expressed and over represented in almost all embryonic stages. The other *mmps* with a 1:1 orthology, seem



**Figure 4.9:** Clustered Mmp phylogeny. The alignment contained 140 sequences and was 951 aa long. The model for amino acid evolution used was JTT+I+G4. Human *mmps* are also indicated by their usual names. Note that gene names ending by .1, .2, etc... strictly represent duplicates within a species, with an arbitrary numbering, and do not correspond to orthologues among vertebrates.



**Figure 4.10:** Clustered *mmp* synteny analysis. Genomic distances are not to scale. Color are as in the phylogeny Fig 4.9. Human *mmps* are also indicated by their usual names. Note that gene names ending by .1, .2, etc... strictly represent duplicates within a species, with an arbitrary numbering, and do not correspond to orthologues among vertebrates.

to show small expression values in most tissues, although many have expression values biased towards embryonic stages, skeletal tissues or the nervous system.

The *mmp21* genes are much more restricted in terms of expression, with *mmp21A* being solely over represented in endoskeletal tissues and *mmp21B* having a very specific expression in the seminal vesicle and to a lesser extent in the uterus. In contrast, *mmp21C* and *mmp21D* display low expression values with a maximum of  $\approx 15$  TPMs. These very low expression values suggest a possible degeneration of these extra duplicates in *S. canicula*.

Finally within the clustered *mmp* clade (in pink in Table 4.4) the different copies seem to have specific expression patterns. Several duplicates, such as *mmp clade 3*, *mmp clade 6.2* and *mmp clade 6.5*, only display expression values  $\geq 5$  in denticles and/or dental lamina. Similarly, *mmp clade 1.2*, *mmp clade 5.2* and *mmp clade 6.3*, display extremely low expression values (TPM  $\leq 15$ ) in all tissues, therefore questioning the function of these genes and indicating a possible degeneration of these duplicates. In addition, *mmp1.1* is specifically expressed in the rectal gland, while *mmp clade 2* is expressed in blood and the spleen, as well as in the testis. The *mmp clade 6.1* gene seems to be specifically expressed in dermal denticles similar to *mmp clade 6.4* who is additionally expressed in the Meckel's cartilage as well as the dental lamina.

#### 4.3.4 Discussion

##### Evolution of gnathostome *mmp* members

Novel Mmp orthologs have been previously found in different organisms such as amphibians (Baddar et al., 2021) or primates (Pan et al., 2022). In this study, we explored the evolution of Mmp proteins using phylogenetic, synteny, and transcriptomic data in chondrichthyans and more specifically in the small spotted catshark: *S. canicula*. Our analyses suggest that multiple duplication and gene loss events have happened at a high rate and independently in major lineages of jawed vertebrates especially in the clustered clade of Mmps. In chondrichthyans we also showed a loss of the Mmp17b, Mmp19, and Mmp clade 4 orthologs while also discovering an unknown diversity of *mmp* genes in this clade. For instance, our findings show that Mmp clade 1 is only found in chondrichthyans as a consequence of a secondary loss in osteichthyans (Fig 4.9). Similarly, chondrichthyans seem to have lost the orthologs corresponding to the clade 4 and possibly one of the clade 5 paralogs, although for the latter, our data cannot distinguish whether the duplication in this clade happened before gnathostome or before osteichthyan diversification. This is mostly due to the low support of the node linked to this duplication in our phylogeny (Fig 4.9; 29.7/72 and Fig 4.11). Most of the clustered *mmps* in bony fishes result from tandem duplications of an ancestral *mmp* of clade 4 while in cartilaginous fishes most tandem duplicates come from the ancestral genes of clade 1, 6 or 5 depending of the species (Fig 4.9 and 5.10).

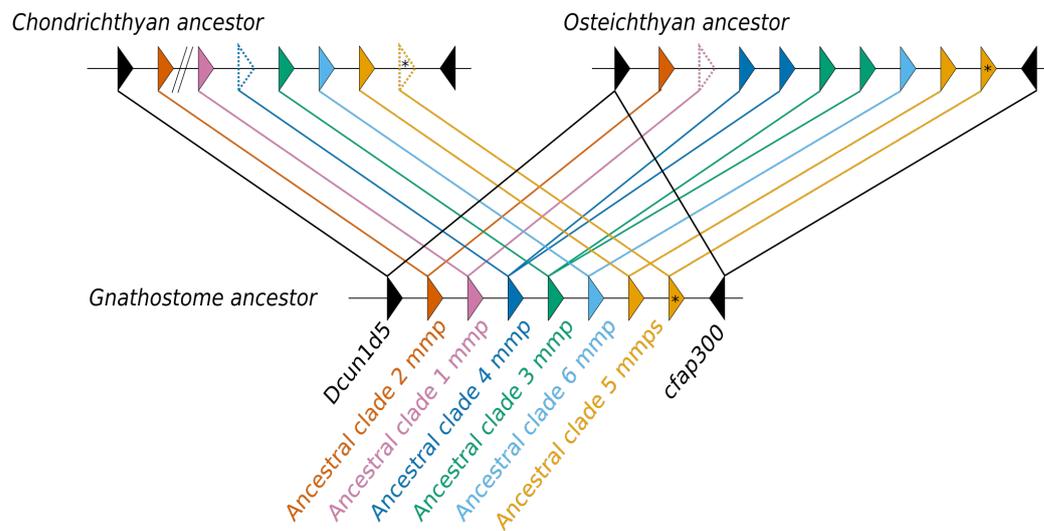
	<i>mmp2</i>	<i>mmp9</i>	<i>mmp11</i>	<i>mmp14</i>	<i>mmp15</i>	<i>mmp16</i>	<i>mmp17a</i>	<i>mmp21A</i>	<i>mmp21B</i>	<i>mmp21C</i>	<i>mmp21D</i>	<i>mmp23ba</i>	<i>mmp23bb</i>
Meckels	525,2	65,1	16,9	117,0	8,0	5,4	5,3	36,0	0,1	1,3	0,3	2,2	9,7
Vertebrae	385,3	8,5	3,4	81,7	9,0	2,9	5,4	69,8	0,1	1,6	0,3	1,6	7,0
Chondrocranium	426,7	14,8	6,6	68,0	13,2	7,7	7,9	18,5	2,7	1,2	0,9	3,8	8,1
Denticles	248,7	386,4	12,3	106,4	4,8	4,0	2,8	3,8	0,2	0,4	0,5	9,0	3,0
Dental lamina	516,1	170,1	29,6	137,4	9,9	3,9	4,2	4,6	0,8	0,6	0,5	3,5	14,7
Brain OS	29,5	4,0	6,2	24,1	4,0	6,2	6,8	5,6	0,8	1,1	0,6	0,9	0,5
Eye	209,1	10,4	10,1	48,9	8,5	7,7	10,1	9,8	0,3	1,4	1,0	1,4	1,3
Spinal chord	108,1	13,8	2,9	32,8	8,5	13,7	9,2	4,1	1,2	0,9	2,3	1,6	0,4
AOL	501,6	14,1	4,7	71,9	6,9	11,2	10,0	4,1	5,2	1,0	0,4	0,8	4,2
Gills	212,4	55,1	20,5	51,2	3,6	0,7	3,1	4,3	0,2	0,5	0,5	8,7	1,6
Blood	0,3	221,1	4,4	7,6	0,4	0,1	2,4	3,6	0,1	4,0	0,3	0,1	0,4
Spleen	19,7	136,0	4,7	74,6	3,5	2,2	3,2	5,6	5,6	0,6	0,4	0,6	0,4
Heart	142,9	30,9	2,8	42,4	24,3	0,5	3,1	3,3	4,5	0,7	0,3	11,1	0,4
Muscle	71,3	7,4	1,6	25,4	5,2	0,5	1,7	1,9	1,9	0,5	0,0	1,1	1,0
Esophagus	190,6	28,6	5,5	67,7	1,7	0,9	2,2	4,8	0,8	0,5	0,3	1,4	0,4
Stomach	8,7	1,3	2,4	44,5	0,4	0,1	0,7	1,0	3,7	0,2	0,1	0,3	0,1
Pancreas	6,2	5,6	0,6	9,2	0,2	0,1	0,4	0,6	1,8	0,1	0,1	0,2	2,6
Liver	4,1	14,9	1,4	62,4	0,5	2,2	1,7	1,6	0,5	0,1	0,2	0,0	0,1
Spiral intestine	3,9	0,7	4,2	48,3	0,4	0,2	2,1	2,5	0,4	0,4	0,3	0,5	0,2
Rectal gland	99,0	44,1	5,2	59,1	2,1	0,6	3,4	3,9	7,7	0,4	0,3	1,0	0,3
Kidney	47,3	8,5	19,7	25,3	4,2	8,6	16,3	4,2	4,3	0,8	0,7	2,7	0,3
Seminal vesicle	60,7	2,2	5,8	60,5	1,1	0,2	3,2	2,3	130,7	0,2	0,2	2,4	0,4
Testis	72,7	312,4	18,1	15,3	3,1	0,7	2,9	3,9	3,1	1,4	1,0	0,5	12,7
Ovary	5,8	0,2	0,4	4,3	0,2	0,1	0,7	0,5	0,2	0,0	0,1	0,1	1,9
Uterus	37,0	1,2	7,3	20,4	1,3	0,7	2,8	2,8	16,7	0,6	0,4	0,3	0,4
E-St-12	29,6	0,1	6,7	23,4	3,9	10,9	2,3	3,7	0,5	15,7	0,9	0,2	0,5
E-St-22	54,1	0,1	10,6	63,5	21,1	19,5	9,1	4,6	1,2	4,4	0,6	0,2	1,8
E-St-24	48,6	13,1	9,9	98,5	18,8	18,4	22,1	5,3	1,6	4,6	1,1	0,6	1,9
E-St-26	61,2	0,2	14,0	68,6	27,0	20,4	16,7	5,1	1,5	1,4	0,5	0,2	5,3
E-St-30	118,5	1,4	20,0	110,9	43,1	21,3	11,5	7,1	2,2	2,3	0,9	0,8	20,3
E-St-31	102,8	2,9	21,8	98,2	43,7	17,6	9,2	6,8	1,3	1,1	0,7	1,1	30,0

**Table 4.3:** TPM values of *S. canicula* Mmp genes. Samples with both Z-scores  $\geq 1$  and TPMs  $\geq 50$  are highlighted in orange, which indicate an enriched and strong expression in the sample. Samples with Z-scores  $\geq 1$  but TPMs  $\leq 50$  are highlighted in purple. AOL: ampullae of Lorenzini; E-St-X: embryo stage x (from Ballard et al 1993); OS: olfactory sac.

	<i>mmp24</i>	<i>mmp28</i>	<i>mmpc1.1</i>	<i>mmpc1.2</i>	<i>mmpc2</i>	<i>mmpc3</i>	<i>mmpc5.1</i>	<i>mmpc5.2</i>	<i>mmpc6.1</i>	<i>mmpc6.2</i>	<i>mmpc6.3</i>	<i>mmpc6.4</i>	<i>mmpc6.5</i>
Meckels	10,1	3,9	4,3	0,2	0,6	4,4	1,3	0,0	1,9	0,4	3,0	113,5	4,4
Vertebrae	8,1	3,0	0,2	0,3	1,1	0,2	1,0	0,0	0,1	0,0	1,3	0,4	0,8
Chondrocranium	12,4	8,6	2,1	0,2	6,4	0,2	5,9	0,1	0,2	0,0	1,7	1,7	1,3
Denticles	4,9	7,1	22,6	4,5	2,3	32,6	12,2	0,5	180,7	48,6	14,3	1245,5	46,5
Dental lamina	5,9	12,0	23,5	3,9	0,9	7,8	5,7	0,3	27,2	3,1	10,2	651,8	26,9
Brain OS	15,7	4,7	0,9	0,6	1,3	1,7	240,2	4,7	0,1	0,1	2,2	0,3	1,4
Eye	7,7	8,3	0,9	0,3	3,9	0,4	4,3	0,2	0,1	0,0	1,8	0,7	1,5
Spinal chord	13,0	3,0	1,2	0,6	5,8	0,0	2,0	0,2	0,2	0,1	2,0	0,2	1,2
AOL	8,2	5,5	2,0	0,6	6,7	0,7	8,2	0,4	0,8	0,0	2,3	26,6	2,4
Gills	4,3	12,2	4,1	0,4	16,3	0,3	961,8	4,6	0,1	0,0	1,9	0,1	2,8
Blood	2,8	4,0	18,8	0,6	124,1	0,1	21,9	1,2	0,1	0,1	2,1	0,0	3,1
Spleen	4,7	7,3	6,8	0,4	88,5	0,1	27,4	0,8	0,1	0,0	2,4	0,1	2,1
Heart	2,9	18,2	2,1	0,1	10,0	0,0	14,3	0,0	0,0	0,0	1,3	0,0	1,4
Muscle	1,3	1,6	0,9	0,2	1,8	0,0	1,3	0,1	0,4	0,1	0,5	2,9	0,5
Esophagus	3,4	5,2	1,3	0,2	9,1	0,0	6,1	0,0	0,0	0,1	1,5	0,0	1,2
Stomach	1,0	6,2	0,8	0,0	0,4	0,2	2,7	0,0	0,0	0,0	0,6	1,1	0,4
Pancreas	0,5	0,9	0,5	0,0	2,8	0,0	0,8	0,0	0,0	0,0	0,3	0,0	0,2
Liver	2,1	2,4	1,4	0,1	6,8	0,0	3,1	0,0	0,1	0,0	1,8	0,0	0,7
Spiral intestine	2,0	2,0	4,5	0,0	1,1	0,1	7,4	0,5	0,0	0,0	1,9	0,0	0,7
Rectal gland	3,5	11,8	136,3	0,2	35,0	0,1	10,5	1,1	0,1	0,0	1,9	0,7	1,6
Kidney	5,2	11,6	2,4	0,2	3,8	0,1	2,2	0,1	0,0	0,0	2,5	0,5	1,8
Seminal vesicle	2,3	5,0	4,9	0,1	0,6	0,1	1,5	0,0	0,0	0,0	1,7	0,0	0,8
Testis	3,0	4,9	12,7	0,1	106,4	0,2	28,2	0,5	0,1	0,1	5,1	0,2	2,6
Ovary	0,5	1,0	0,1	0,2	0,1	0,1	0,1	0,0	0,0	0,0	0,4	0,0	0,1
Uterus	2,8	13,9	0,6	0,1	1,0	0,2	1,2	0,1	0,0	0,0	1,9	0,0	1,4
E-St-12	3,4	2,4	4,1	0,1	0,2	0,0	0,4	0,0	0,0	0,0	1,4	0,0	1,2
E-St-22	5,6	4,3	0,5	0,3	0,1	0,1	1,3	0,0	0,0	0,0	1,9	0,0	1,1
E-St-24	7,5	7,9	0,7	0,4	0,3	0,1	63,7	4,0	0,3	0,1	2,5	0,1	1,8
E-St-26	9,3	7,7	0,4	0,4	0,1	0,1	2,5	0,0	0,1	0,0	1,9	0,0	1,0
E-St-30	12,6	10,4	0,8	0,7	3,2	0,5	3,5	0,1	0,1	0,0	2,4	0,1	1,9
E-St-31	10,3	9,1	1,2	0,5	6,1	0,5	5,6	0,5	0,4	0,1	2,0	1,1	1,9

**Table 4.4:** TPM values of *S. canicula* Mmp genes. Samples with both Z-scores  $\geq 1$  and TPMs  $\geq 50$  are highlighted in orange, which indicate an enriched and strong expression in the sample. Samples with Z-scores  $\geq 1$  but TPMs  $\leq 50$  are highlighted in purple. Mmps belonging to the cluster clade are highlighted in pink. AOL: ampullae of Lorenzini; E-St-X: embryo stage x (from Ballard et al 1993); OS: olfactory sac.

Integrating phylogenetic and synteny data of the clustered *mmps*, helped us to propose an evolutionary scenario for this clustered calde (Fig 4.11). This scenario points towards six or seven (depending on the timing of the clade 5 duplication) *mmp* genes already present in a cluster, in the last common ancestor of extant gnathostomes. However, we cannot reject the hypothesis that both osteichthyan clade 5 genes are sister groups, in which case they would result from a duplication in the ancestor of osteichthyans, and there would be only one clade 5 copy in the ancestor of gnathostomes. The clade 4 ancestral gene was duplicated in the osteichthyan ancestor while it was lost in chondrichthyans. Similarly, the clade 3 ancestral gene also would have duplicated in the osteichthyan ancestor.



**Figure 4.11:** Ancestral state of clustered *mmps*

### Chondrichthyan specific expression of novel *mmp* genes

Transcriptomic data in *S. canicula* show highly specific expression zones of some *mmp* genes. Particularly relevant are the *mmps* that are highly expressed and over-represented in chondrichthyan specific derived anatomical features. For instance Mmp clade 1.1 is highly expressed in the rectal gland, a specialized elasmobranch organ, important for osmoregulation. Moreover, we detected expression of *mmp2* and *mmp14* in the ampullae of Lorenzini, a chondrichthyan specific sensory organ specialized in electroreception. Our data suggest that these *mmps* have been recruited in these novel structures although their functional significance is still unknown.

Another relevant feature specific to chondrichthyans is their ability to regenerate teeth throughout their life (Rasch et al., 2016). Moreover, it has been shown that Mmp proteins are involved in tooth development, mineralisation and regeneration (Guirado and George, 2021). Here, we found five *mmps* that are highly expressed and over-represented in *S. canicula* dermal denticles and dental lamina (*mmp2*, *mmp9*, *mmp14*, *mmpc6.1* and *mmpc6.4*) suggesting a

potential role of these genes in the development, regeneration and maintenance of teeth and scales in *S. canicula*.

Lastly, several *mmp* genes were detected in endoskeletal tissues including *mmp2*, *mmp14*, *mmp21A* and *mmpc6.4*. These expression patterns could indicate a potential contribution of these genes in the evolution of the chondrichthyan skeleton, and explain some of the observed differences between them and bony fishes. For instance in *mmp14* null mice, strong skeletal defects can be observed due to delayed cartilage resorption during endochondral ossification (Holmbeck et al., 1999; Zhou et al., 2000). Similarly, *mmp2* null mice show disrupted bone development as well as decreased mineralization and osteoblast / osteoclast numbers (Mosig et al., 2007). Comparisons for the other two genes are more complex since both *mmp21A* and *mmpc6.4* orthologs have been lost in mammals.

## 4.4 Discussion: Evolutionary dynamics of clustered gene families

In this chapter I focused on genes that evolved into clusters through tandem duplications and the dynamics behind these genomic structures. In general a gene cluster is defined as two or more paralogs that encode similar proteins and are located next to each other on the same chromosome. A gene cluster can arise through several tandem duplications at a particular loci. This effect can be further exacerbated through non homologous recombination if the duplicated sequences are very similar to each other, increasing the number of duplicates (Hastings et al., 2009). In general, duplicated genes, whether they are generated through tandem or whole genome duplications, are first functionally redundant and therefore one of the duplicated go through either the acquisition of a new function (neofunctionalization), become degenerate (pseudogenization) or share the ancestral function with the other duplicate (subfunctionalization; DDC model Force et al., 1999). However, when many duplicates are beneficial (such as in dosage-sensitive genes), there can be positive selection for an increase in duplicate conservation, therefore resulting in a higher apparent duplication rate (Hahn, 2009; Margres et al., 2017; Schrider and Hahn, 2010).

These tandem duplications can be seen in many gene families studied in this thesis such as the *scpp* genes in bony fishes, the *spp2* genes in cartilaginous fishes and both the (clustered) *slrp* and the (clustered) *mmp* genes in gnathostomes. All these gene families code for structural (or enzymatic, for mmps) proteins of the ECM. In particular for the *slrps* and clustered *mmps*, we showed there was already an ancestral cluster of these genes in the ancestor of jawed vertebrates, suggesting an already diversified repertoire in this last common ancestor. However, how these ancestral clusters evolved thereafter varies greatly depending on the studied family.

On the one hand, *slrps* have retained a clear signature of the 2R of whole genome duplications before vertebrate diversification. Therefore, the ancestral four-gene cluster was duplicated

such as to display four *slrp* clusters in different chromosomes. Although not all paralogs were retained in every species, *slrps* do show a relatively static condition in terms of gene conservation. This static condition is uncommon following WGD events since they are hypothesized to be followed by massive and rapid genomic reorganizations and gene deletions (Hufton and Panopoulou, 2009; Sémon and Wolfe, 2007 but also see Berthelot et al., 2014).

On the other hand, the clustered *mmps* subsequently diversified through additional tandem duplication in extant gnathostomes. Surprisingly, the ancestral copy from which these recent tandem duplicated emerged are not the same depending in cartilaginous fishes when compared to bony fishes. This suggests that similar requirements for larger *mmp* repertoires may exist in these species. Compared to *slrps*, clustered *mmps* are not as conserved and show high turn-over and high duplication rates.

These differences between *slrps* and *mmps* highlight two types of cluster: one very well conserved at the gnathostome scale with little alterations, and another not so well conserved and seemingly very dynamic. Why these differences exist is still to be determined, however, *slrps* seem to have stabilized after vertebrate diversification while clustered *mmps* have continued duplicating and evolving afterwards. Within the studied genes in this manuscript, the condition of the *mmp* gene cluster seems to be the standard since similar dynamics can be seen in the bony fishes *scpps* as well as in the cartilaginous fishes *spp2* genes. The *slrp* condition is more intriguing and the constraints that force them to remain clustered should be studied further.



# Chapter 5

## Conclusion: Evolution of the biomineralization toolkit in gnathostomes

### 5.1 O tissue, who art thou

As an evolutionary developmental biologist, the ideal situation would be that all the homology relationships between structures are easily inferred from simple comparative data. However, complex organisms such as vertebrates are not too keen in letting their secrets be known. In particular, this comes from several levels of organization that are integrated into each other, such that homology relationships become blurry. For instance, homologous genes can be easily identified but their homology does not answer the question of the homology of genetic interactions. This problem is further exacerbated by long evolutionary times that tend to create derived features, at all levels of organization, which further complicates the retracing of homology. In the case of, but not restricted to, vertebrate skeletal biology this is an issue that must always be considered such that our deductions stay as true to the reality as possible. For instance, the homology of skeletal tissues among vertebrates has been the subject of intense debate. In odontodes, dentin is mostly considered to be a tissue with stable features, shared among wide taxonomical ranges so interpreted as homologous among vertebrates, but the extant enamels and enameloids are considered more derived tissues, with homology relationships that were long debated first based on tissular morphology (Janvier, 1996), then on protein content which depends on cell-specific gene expression (Kawasaki and Weiss, 2008). One of the main arguments against enamel/enameloid homology is the contribution of odontoblasts in enameloid formation with collagen I fibrils (Ørvig, 1967). However, we could argue that this was an ancestral condition that was subsequently lost in mammalian enamel (or earlier, see Qu et al., 2015). Already in odontodes, tissular homology is not clear and depends of morphology,

which is strongly influenced by protein content, which is itself directly linked to gene expression, controlled by cells. Therefore, it is essential to better understand these cells and how they relate to each other during evolution. However, it must be taken into account that, similar to the evolution of sequences and morphological features, working on tissues and cells requires a sufficiently big sampling and well chosen outgroup(s) that allows us to polarize the changes in characters.

In endoskeletal tissues, cartilage seems to be a very ancient tissue and is probably homologous among all of vertebrates and possibly bilaterians (Tarazona et al., 2016, but see below). Bones on the other hand are more complicated since e.g. endochondral and dermal bones do not develop through the same developmental processes (endochondral *vs* intramembranous ossification, respectively). It is possible to argue that because not all of these tissues receive the same contribution of different cellular types (i.e. absence of chondrocytes in dermal bones), they look different morphologically, but that they are in fact homologous since derived from an ancestral form of bone that used osteoblasts. To illustrate this, a 2016 study (Weigele and Franz-Odenaal, 2016) on the histology of adult zebrafish revealed that most of the zebrafish skull bones are either cellular or acellular compact bones (formed through intramembranous ossification). The remaining endochondral bones are formed by two different endochondral ossification processes named type I and II. Type I endochondral ossification is similar to mammalian endochondral ossification while Type II lacks both a degradation zone with osteoblasts and a mineralization zone adjacent to the hypertrophic chondrocytes. Bones formed through Type II endochondral ossification are called tubular bones, they lack a bony trabecular network, blood vessels and nerves and are instead filled with adipose tissue. The authors of the article hypothesize that these represent a derived skeletal feature specific to the zebrafish. In this case, the novel tubular bones of the zebrafish are most probably derived from the “classic” endochondral bone, and are simply morphologically different because they lack a contribution of osteoblasts in the innermost part of the bone.

Therefore, the problem of tissular homology seems to boil down to what are the cells that are participating in the formation of the tissue, and most importantly whether these cells are homologous or not. However, we will see in the following that cell homology is a quite challenging topic to tackle.

## 5.2 Cell homology conundrum

Classically there are two ways of classifying a cell: through overall morphological (most often histological) attributes (e.g. rod and cone cells in the retina) or through their function. Cellular function depends on many proteins that constitute the toolkit of the cell, and finding what is each cell’s toolkit is a common approach for defining a cell-type. Other studies have attempted to characterize cells through their gene expression profile: a cell’s “molecular fingerprint”, mostly fueled by high-throughput approaches such as single-cell RNAseq. How-

ever, these characterizations do have their limitations, in particular since cell-type similarity can evolve through several processes (such as true homology through inheritance, but also convergence or concerted evolution; Arendt et al., 2016), and current cell-type characterizations cannot differentiate between these scenarios easily. More recent definitions of cell-type are based on the concept of gene regulatory networks (GRNs), representing a regulatory signature specific to a single cell-type, influencing morphology and function and allowing the cell to act as a single evolutionary unit (Arendt et al., 2016). This concept, is also employed in deconvolution methods (cf. chapter 2). Still, this definition is not optimal since it is not clear on where the homology lies. This is well illustrated in Tarazona et al., 2016 where the authors explore different hypotheses for the evolution of cartilage and more specifically chondrocytes in bilaterians. Two main hypotheses compete in this paper: (i) the independent origin of the chondrocytes happened through the recruitment of a homologous GRN, by the same homologous progenitor cell-type, or (ii) independent evolution by activation of the same homologous GRN but in different, non-homologous progenitor cell-types. These results highlight that the homology of cell-types cannot be sharply defined because it integrates two independent classical arguments to identify homology: of cell provenance (embryonic origin) and of GRNs. Therefore, GRNs cannot alone be the defining character of a cell-type. However, it seems that homologous GRNs in different cell progenitors can trigger the differentiation of similar cell-types, resulting in morphologically and functionally similar tissues (Tarazona et al., 2016).

Following these lines of evidence, the best way to study the evolution of tissues through the lens of Evo-Devo is to understand how GRNs were co-opted in the different cell progenitors they are built from. We could argue that what we call bone in bony fishes is in fact any tissue formed by any cell progenitor containing a precise homologous GRN specialized in collagen secretion and calcium phosphate precipitation. This hypothesis implies a common GRN in all osteoblast-type cells. However, several elements must be taken into account such as the independent evolution of these GRNs in separate taxonomic (or organismal) lineages (e.g. shark vs mice chondrocytes) and/or independent evolution of a given GRN compared to others within a single organism (e.g. fibrous mineralization cells vs areolar mineralization cells). Among the many other cell-types that mineralize, including but not limited to odontoblasts, ameloblasts and chondrocytes, is there a common GRN between all of these? Several studies have already argued the existence of an ancient “skeletal” genetic network (reviewed in Fisher and Franz-Odenaal, 2012), however, they mainly focussed on early master regulators of cell differentiation. In the following section I will expand on this ancient skeletal genetic network, although I will be focusing on the set of genes I worked with: proximal actors of mineralization. Can we describe the mineralization part of an ancestral skeletal GRN (herein referred to as mineralization toolkit), specific to some or common to all mineralizing skeletal cell progenitors? Is our data sufficient to characterize skeletal cells as early as the chondrichtyan/osteichtyan split? Can we conclude on the loss of bone or not in cartilaginous fishes? And can we classify these cells based on their gene expression resemblance?

### 5.3 The biomineralization toolkit of gnathostomes

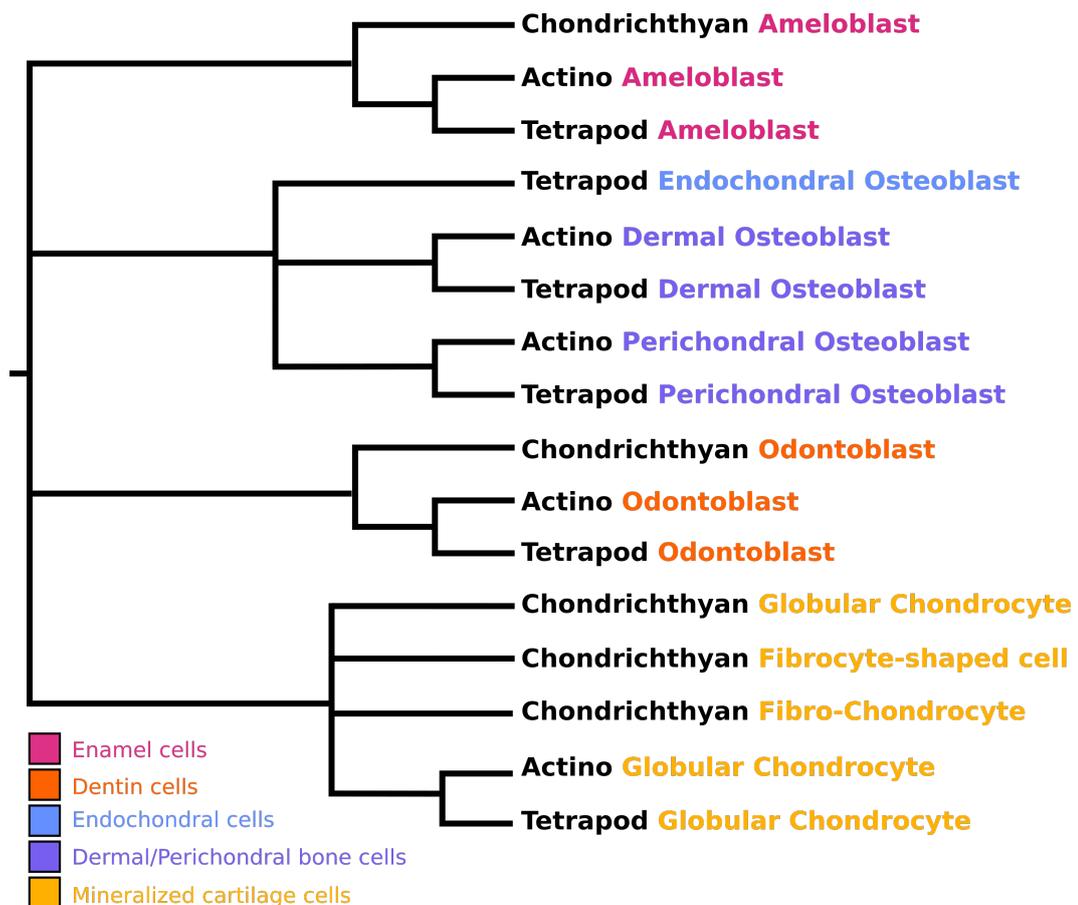
Along this manuscript a variety of genes, cells and tissues have been studied in order to understand their evolution, homology relationships and hopefully their origin. The focus was made on several structural proteins but also in secreted proteins directly linked to calcium phosphate precipitation. Table 5.1 summarizes the expression patterns of all the genes studied in this manuscript in three main lineages: chondrichthyans, actinopterygians and tetrapods. The presence or absence of expression was extracted either from data presented here or from the available literature. Additional unpublished data presented in chapter I for *alp*, *fam20c* and the chondrichthyan *scpp* have been added to complete the analysis. This data was then used to trace back the evolution of the characters. For the inference of the *mgp* ancestral state, it must be mentioned that as long as one of the duplicates (*mgp1* and *mgp2*) in chondrichthyans displays presence of expression in a given tissue, then we consider it as part of the ancestral expression before duplication in this clade. In addition, because the *scpps* independently inherited the expression pattern of *sparc-L*, for instance in ameloblasts, both in bony fishes and cartilaginous fishes, we considered that *sparc-L* displayed this expression pattern ancestrally despite losing it subsequently in bony fishes. Therefore, similar to *mgp1* and *mgp2* these genes will be considered as just one unit in subsequent analyses. Lastly, the *slrps* are not very well conserved in terms of expression among tissues and among lineages: different *slrps* seem to have been recruited individually in different tissues. Thus, they will not be further considered. In addition, since *spp2* is missing so much data that we also chose to remove it from subsequent analyses.

Actinopterygian endochondral bone data must be interpreted with care. This comes from several biases and gaps in knowledge concerning the main animal model in this clade: the zebrafish. Endochondral bones in the zebrafish (and more generally in teleostean species) are under-studied compared to dermal bones because they are less abundant, especially in the skull, the focus of most studies (see sources in Table 5.1). Moreover, since zebrafish embryos are so small, labelling of expression (through *in situ* hybridization for instance) is often done as whole mount, and therefore does not allow for a detailed separation between perichondral and endochondral compartments. Surprisingly, little is known about the adult zebrafish skeleton, thus most publications focus on early bone development, before endochondral bone is formed. In addition, given the fact that sharks do not have bones (dermal, perichondral and endochondral) comparisons for these tissues are limited to actinopterygians and tetrapods, therefore limiting inferences to the bony fish ancestor. Moreover, because bony fishes are the only extant species presenting bone and therefore osteoblasts, there can be a confounding effect between bony fish-specific and osteoblast-specific toolkits linked to mineralization. Therefore, I might not be able to differentiate whether these genes were recruited in the last common ancestor of bony fishes, or before that when osteoblasts first appeared (which is still a matter of debate).

	<i>collagen I</i>	<i>collagen II</i>	<i>collagen X</i>	<i>sparc</i>	<i>sparc-L / L1</i>	<i>scpp</i>	<i>spp2</i>
Ch. ameloblasts	- [1]	- [1]	+ [3]	- [9]	+ [9]	+	-
Ac. ameloblasts	+ [7,8]		+ [3]	+ [10]	(-) [10]	+ [8]	
Te. ameloblasts	- in mammals [6], + in others [4]	- [9]		- [9]		+ [8]	
Ch. odontoblasts	+ [1]	- [1]	+ [3]	+ [9]	- [9]	-	-
Ac. odontoblasts	+ [7,8]	- [9]	+ [3]	+ [10,14]	+ [10]	+ [8]	
Te. odontoblasts	+ [4,6]		+ [3]	+ [9]		+ [8]	
Ac. dermal osteoblasts	+ [12]	- [33]	+ [13]	+ [13]	(-) [16]	+ [8]	
Te. dermal osteoblasts	+ [30]		+ [34]	+ [14]	(-) [15]	+ [45]	
Ac. endochondral osteoblasts					(-) [16]		
Te. endochondral osteoblasts	+ [26]	- [31, 32]	+ [14]	+ [14]	(-) [15]	+	+
Ac. perichondral osteoblasts	+ [2,12]	+ [2]	+ [2]	+ [13]	(-) [16]	+ [8]	+
Te. perichondral osteoblasts	+ [26]	- [31,32]	+ [34]	+ [14]	(-) [15]	+	+
Ch. Fibrocyte-shaped cells (lamellar min.)	+ [1]	- [1]	+ [3]	+	-	-	+
Ch. Fibro-chondrocytes (areolar min.)	- [1]	- [1]	+ [3]	+	-	-	+
Ch. Globular chondrocytes (globular min.)	- [1]	- [1]	+ [3]	+	-	-	+
Ac. Globular chondrocytes (globular min.)	- [2]	+ [2]	+ [2]	+	-	- [8,11]	
Te. Globular chondrocytes (globular min.)	- [26]	- [5]	+ [5]	-	- [15,16]	+ [9]	
	<i>bgp / bgp1</i>	<i>mgp / mgp1 / mgp2</i>	<i>alp</i>	<i>fam20c</i>	<i>sp7</i>	<i>slrps</i>	
Ch. ameloblasts	-	- <i>mgp1 / + mgp2</i>	+	+	- [42*]	<i>bgn, lum, ogn, omd</i>	
Ac. ameloblasts	- [7,19]	+ [19]	+ [44]	+ [39]	+ [41]	<i>fmod, dcn1 [37], omd [38]</i>	
Te. ameloblasts	- [22]		+	+	+ [41]	<i>dcn1, lum</i>	
Ch. odontoblasts	-	- <i>mgp1 / + mgp2</i>			+ [42*,46]	<i>bgn, dcn1, lum [37], omd [38]</i>	
Ac. odontoblasts	+ [19]	+ [19]	+ [44]	+ [39]	+ [41]	( <i>aspn</i> ) [36]	
Te. odontoblasts	+ [22]		+ [18*]	+ [39]	+ [13]		
Ac. dermal osteoblasts	+ [18,27]	- [18]	+ [28]	+ [39]	+ [40]		
Te. dermal osteoblasts	+ [28]		+ [18*]				
Ac. endochondral osteoblasts		(-) [18]	+ [18*]				
Te. endochondral osteoblasts	- [25]	+ [24,23]	+ [29*]	+ [39]	+ [40]		
Ac. perichondral osteoblasts	+ [18,27]	- [18]	+ [18*]		+ [13,46]		
Te. perichondral osteoblasts	+ [21,25]	+ mammals [20] - frog [21]	+ [28,29*]	+ [39]	+ [40]	<i>aspn</i> [36]	
Ch. Fibrocyte-shaped cells (lamellar min.)	-	+ <i>mgp1 / - mgp2</i>	(+)	-			
Ch. Fibro-chondrocytes (areolar min.)	-	+ <i>mgp1 / - mgp2</i>	+	-			
Ac. Globular chondrocytes (globular min.)	-	+ <i>mgp1 / - mgp2</i>	+	-			
Te. Globular chondrocytes (globular min.)	+ [18]	+ [18]	+ [18*]				
	+ chick [47] - others [21,25]	+ [20,21]	+ [29*]	+ [39]	- [43]	<i>chad</i> [37]	

**Table 5.1:** Described expression of genes in bony fishes (literature) and the small-spotted catshark (this study) in selected tissues and selected species of jawed vertebrates. +: presence, -: absence, b: bony fish, Ch and c: cartilaginous fish, Te and t: tetrapods, Ac and a: actinopterygians. Sources: (1) Enault et al., 2015, (2) Eames et al., 2012, (3) Debais-Thibaud et al., 2019, (4) Assaraf-Weill et al., 2014, (5) Maes et al., 2002, (6) P. He et al., 2010, (7) Huyseune et al., 2008, (8) Kawasaki, 2009, (9) Enault et al., 2018, (10) Weigele et al., 2015, (11) Kawasaki et al., 2005, (12) Gistelink et al., 2016, (13) N. Li et al., 2009, (14) P. Holland et al., 1987, (15) McKimmon et al., 2000, (16) Bertrand et al., 2013, (17) Kawasaki and Weiss, 2008, (18) Gavaia et al., 2006, (19) Ortiz-Delgado et al., 2005, (20) Luo et al., 1997, (21) Espinoza et al., 2010, (22) Bleicher et al., 1999, (23) Dan et al., 2012, (24) Coen et al., 2009, (25) Sommer et al., 1996, (26) G. He et al., 2010, (27) Viegas et al., 2013, (28) Weinreb et al., 1990, (29) Miao and Scutt, 2002, (30) Yoshida et al., 2002, (31) Zelzer et al., 2002, (32) Zhao et al., 1997, (33) Åberg et al., 2005, (34) Aldea et al., 2013, (35) Rotllant et al., 2008, (36) Henry et al., 2001, (37) Wilda et al., 2000, (38) Buchaille et al., 2000, (39) Wang et al., 2010, (40) Gao et al., 2004, (41) S. Chen et al., 2009, (42) Kague et al., 2018, (43) Kaback et al., 2008, (44) Väkevä et al., 1990, (45) Candeliere et al., 2001, (46) Wiweger et al., 2012, (47) Neugebauer et al., 1995.

If we consider an evolutionary scenario where an ancestral GRN evolved freely in each diverging cell type (that evolved in an early agnathan vertebrate) and then again freely in each diverging taxonomic lineages of jawed vertebrates, then we would expect a tree topology as in Fig 5.1, where each cell-type (e.g. ameloblasts) forms a monophyletic clade. This is what I will consider as my null hypothesis. In this tree however, there are some missing elements, such as the internal relationships of mineralized cartilage cells, bone tissue types and deeper relationships between ancestral tissues. All major cell type branches multifurcate at their base because we hypothesize that all mineralizing cells share a mineralization toolkit that has a single origin.

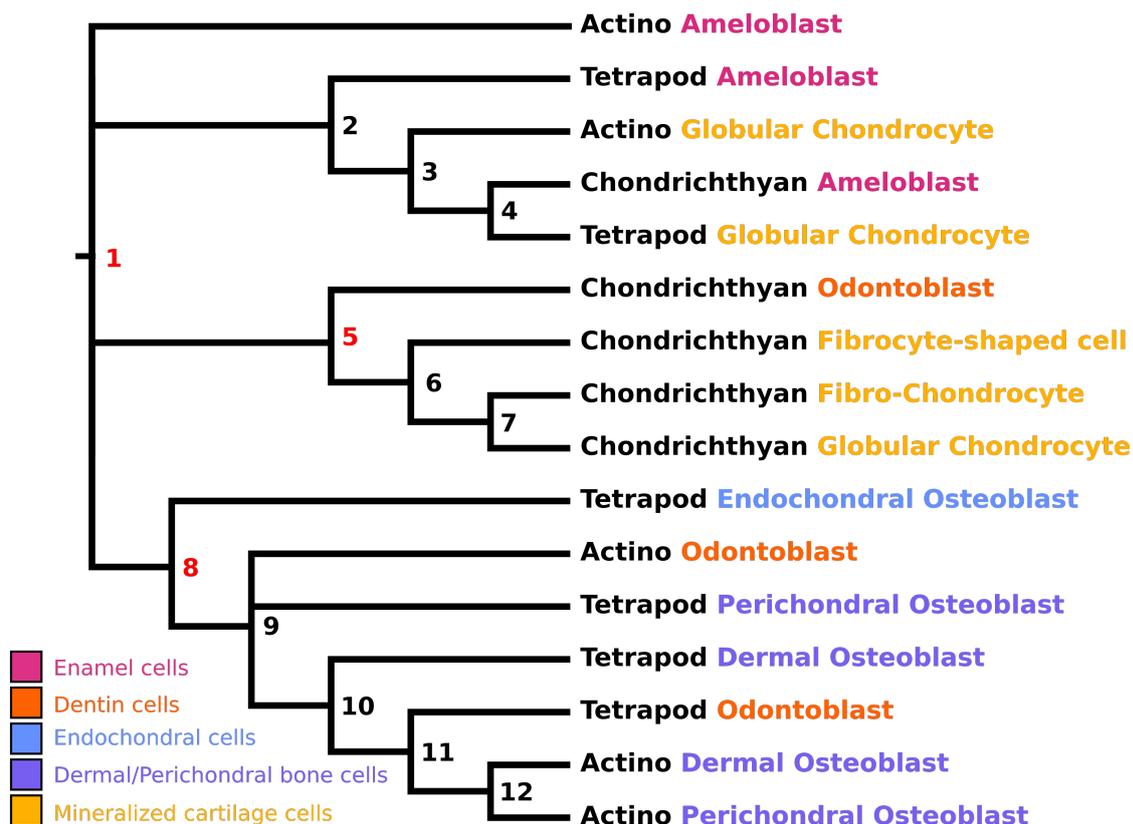


**Figure 5.1:** Null hypothesis of the evolution of mineralization-associated cells and tissues in jawed vertebrates

From the gene expression data presented above, a final list of only ten characters can be derived into a presence / absence matrix for only a list of sixteen tissues (e.g. Table 5.2). This matrix was used to produce several equally probable tree topologies through a maximum parsimony approach. In total 48 trees were recovered. From these tree topologies a consensus tree was built and is presented in Fig 5.2. This tree falls far from the null hypothesis and despite the very faint robustness of this result (due to the very few data available), I will discuss several relevant elements. First the topology suggests non-independent evolution of osteoblasts and odontoblasts (at least in bony fishes), a result in accordance with several other authors that

have already proposed that osteoblasts are derived from odontoblasts (as early as Klaatsch 1890, cited in Ørving, 1951). However, this clade is supported by just one evolutionary event: the hypothetical gain of *sp7* expression (node “8”, Fig 5.2 and Table 5.2), despite uncertainty of the ancestral state at the root of the tree (node “1”, Fig 5.2 and Table 5.2).

Second, this topology groups chondrichthyan mineralized cartilages with chondrichthyan odontoblasts breaking the putative monophyly of odontoblasts. However, this result is also only supported by one evolutionary transition: the loss of *sparc-L/scpp* expression (node “5”, Fig 5.2 and Table 5.2). If further supported, this suggests these tissues derived recently and/or that there was a concerted evolution of the GRNs in these cell types in the chondrichthyan lineage. The former hypothesis would imply that chondrichthyan odontoblasts are not homologous to bony fish odontoblasts, a hypothesis with no support in the literature. As a consequence we consider more credible that there is concerted evolution in this lineage. In this regard the hypothesis of chondrichthyan lamellar mineralization (by fibrocyte-shaped cells) being a derived version of bone (bone-like tissue; e.g. Atake et al., 2019) cannot even be tested through gene expression comparisons, as these would keep a stronger taxonomic lineage signal, than cell type signal.



**Figure 5.2:** Maximum parsimony reconstruction based of gene expression presence / absence in several cells and tissues among jawed vertebrates

The concerted evolution hypothesis relies on very little support, and will probably be challenged with the acquisition of new mineralization-associated gene expression data. For instance,

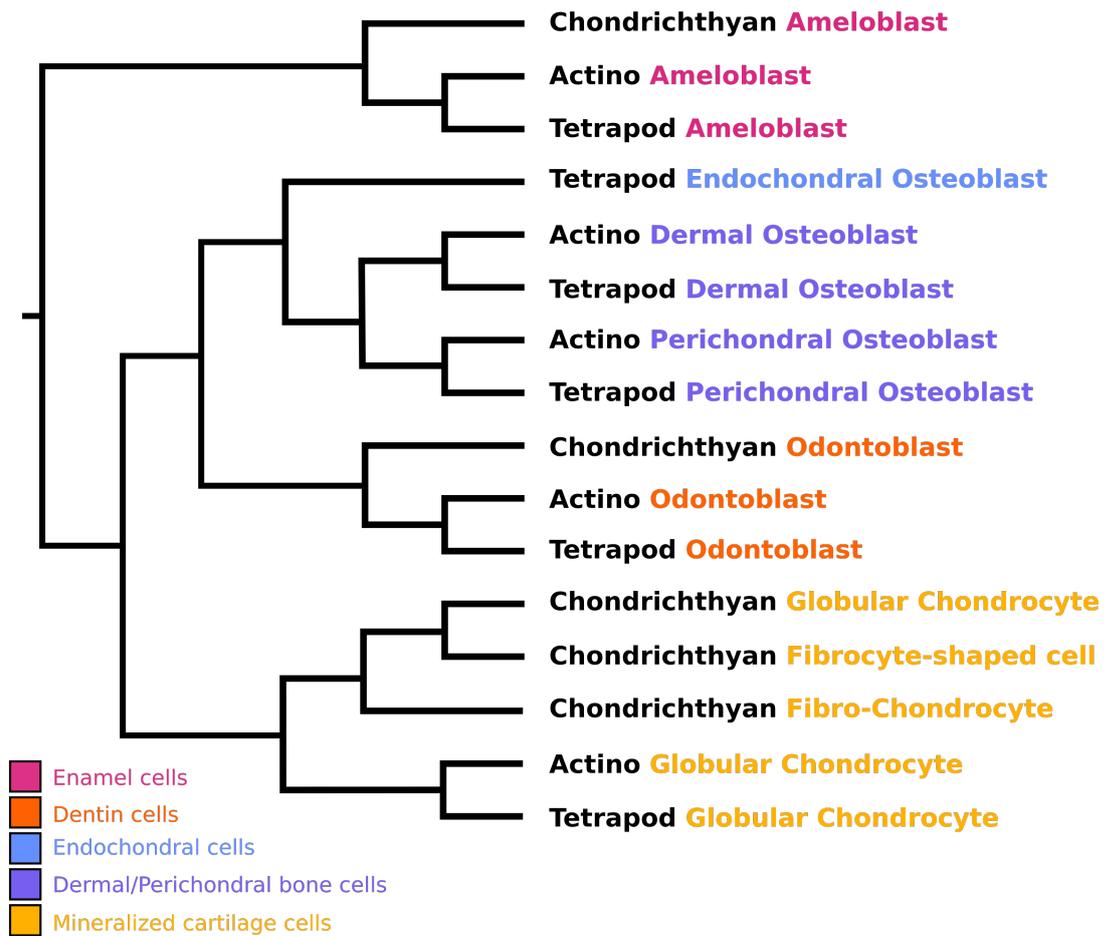
Gene/Node	1	2	3	4	5	6	7	8	9	10	11	12
<i>collagen I</i>	1	1	0	0	1	1	0	1	1	1	1	1
<i>collagen II</i>	0	0	0	0	0	0	0	0	0	0	0	0 1
<i>collagen X</i>	1	1	1	1	1	1	1	1	1	1	1	1
<i>sparc</i>	1	0	0	0	1	1	1	1	1	1	1	1
<i>sparc-L/scpp</i>	1	1	1	1	0	0	0	1	1	1	1	1
<i>bgp</i>	0	0	0 1	0 1	0	0	0	0	1	1	1	1
<i>mgp</i>	1	1	1	1	1	1	1	1	1	0 1	0 1	0
<i>alp</i>	1	1	1	1	1	1	1	1	1	1	1	1
<i>fam20c</i>	1	1	1	1	1	0	0	1	1	1	1	1
<i>sp7</i>	0 1	0 1	0 1	0 1	0 1	0 1	0 1	1	1	1	1	1

**Table 5.2:** Ancestral character reconstruction along the maximum parsimony reconstruction

expression of *spp2* genes, might be a specific feature common to all mineralized endoskeletal tissues of chondrichthyans (Table 5.1). Similarly, the expression *sp7*, a gene that is considered to be a master regulator of osteoblast differentiation in bony fishes, has no published data in cartilaginous fishes. The bias in the studied species for each of these two genes might give the wrong impression that their expression is specific to one lineage or another. Therefore, more work on *spp2* and *sp7* is needed to better understand whether their condition, in chondrichthyans and bony fishes respectively, is a derived or a conserved feature among vertebrates.

Lastly, ameloblasts are not monophyletic, some are grouped with bony fish globular chondrocytes (the chondrocytes found in the globular mineralization type of tissue). This result might be artefactual and probably due to the low number of characters taken into consideration for this analysis (only ten), also showing there is no synapomorphy specific to ameloblasts nor to globular chondrocytes. Ancestral character reconstruction based on this topology results in an ancestral state, before gnathostome diversification and common to all tissues and cells types (node “1”, Fig 5.2 and Table 5.2), made up of the co-expression of *collagen I*, *collagen X*, *sparc*, *sparc-L/scpp*, *mgp*, *alp*, *fam20c* and possibly *sp7*. This can be considered to be the ancestral mineralization toolkit of jawed vertebrates.

Modifying the null hypothesis with regards to the parsimony reconstruction, we propose a scenario of skeletal cell GRN evolution among jawed vertebrates as in Fig 5.3. In this figure, cell type-specific GRNs partly evolve independently from each other, although odontoblasts are considered the sister clade of osteoblasts. Mineralized cartilage cells are monophyletic, with chondrichthyan ones being lineage-specific innovations or following concerted evolution. Finally, we considered ameloblasts as monophyletic and diverging earlier than the other cell types. This scenario should be further refined using more gene expression data that may highlight cell-type interrelationships such as lineage concerted evolution, convergence and/or recent innovation, further strengthening our understanding of GRNs, cell and tissue evolution in jawed vertebrate skeletons.



**Figure 5.3:** Evolutionary scenario of mineralization-associated cells and tissues in jawed vertebrates



# Résumé étendu en français

## Introduction

Les poissons osseux présentent un squelette se développant principalement sous la forme de cartilage, localement remplacé par du tissu osseux endochondral. Les poissons cartilagineux, eux, présentent un squelette de nature cartilagineuse, localement minéralisé selon diverses modalités dont certaines sont probablement des innovations évolutives propres à cette lignée (le cartilage prismatique, par exemple). Les cellules associées à ces tissus sont capables de biominéralisation par précipitation de phosphate de calcium, une propriété qui semble limitée au groupe des vertébrés. Au cours de cette thèse, je focalise mon étude sur la génétique du développement et de la minéralisation du squelette chez un poisson cartilagineux, la petite roussette (*Scyliorhinus canicula*), pour décrire de manière la plus fine possible la boîte à outils génétique associée à l'ontogenèse du squelette et au processus de minéralisation. Ceci est ensuite complété par une étude comparative à l'échelle des vertébrés, pour inférer l'état ancestral de cette boîte à outils, et améliorer notre compréhension sur l'origine de la biominéralisation à cette échelle taxonomique.

## I - Objectifs de la thèse

Je me suis concentré sur la génétique du développement de la minéralisation du squelette chez un poisson cartilagineux : la petite roussette *Scyliorhinus canicula*. Je décris en détail l'évolution moléculaire et l'expression de facteurs génétiques impliqués dans le processus de minéralisation, et je le fais principalement grâce à deux approches : une approche par gènes candidats et une approche transcriptomique.

Dans une approche par gènes candidats, les gènes sont choisis en fonction de données fonctionnelles connues chez les poissons osseux, puis étudiés chez les poissons cartilagineux. Une approche par gène candidat peut être très efficace pour détecter les points communs ou les différences entre deux lignées, mais est biaisée vers l'étude de caractéristiques conservées ou propres aux ostéichthyens. Par conséquent, nous pouvons passer à côté d'innovations importantes spécifiques aux chondrichthyens, ou de caractéristiques ancestrales qui ont été perdues secondairement chez les poissons osseux.

La deuxième approche utilise des données “omiques”, telles que des données génomiques, transcriptomiques ou protéomiques. Celles-ci peuvent fournir des éléments intrinsèques des chondrichthyens liés à la minéralisation, qui pourraient ensuite être comparés à la littérature existante sur l’ossification. Ces techniques très attrayantes n’avaient cependant pas été utilisées jusqu’à présent en raison d’un répertoire très pauvre de génomes et de transcriptomes bien assemblés chez les chondrichthyens. Ces dernières années, de nouveaux assemblages de génomes au niveau chromosomique chez des requins et des raies ont été publiés, permettant de mieux identifier la boîte à outils moléculaire de la minéralisation des chondrichthyens, la comparer à celle des poissons osseux, et déduire l’état ancestral des gènes de la minéralisation chez les vertébrés à mâchoire.

Ces approches m’ont permis de mieux caractériser les bases génétiques de la minéralisation chez une espèce de chondrichthyens (la petite roussette), et grâce à une analyse comparative, de proposer un état ancestral pour les vertébrés à mâchoire. Notre première approche par gènes candidats nous a donné un aperçu du niveau de conservation (ou de non-conservation) de plusieurs gènes connus et associés à la minéralisation. Ensuite, j’ai pu analyser des données génomiques et transcriptomiques récemment générées chez la petite roussette pour étudier les gènes intrinsèques exprimés spécifiquement dans les tissus minéralisants. Ces deux approches m’ont permis de détecter des gènes intéressants qui ont ensuite été étudiés à travers des analyses phylogénétiques, de quantification / localisation de leur expression dans les tissus, suivis d’analyses comparatives avec les poissons osseux (lorsque a été possible). Dans une troisième partie découlant des deux premières, je présente des travaux sur deux familles de gènes qui se sont orientés vers un cadre d’évolution moléculaire plutôt qu’un cadre d’Evo-devo.

## II - Des gènes de la minéralisation chez les poissons cartilagineux

L’approche par gènes candidats s’est révélée puissante pour décrire les répertoires de gènes et les patrons d’expression ont pu être comparés dans les principales lignées de gnathostomes qui pourraient expliquer les différentes caractéristiques du squelette. Ainsi, mes trois premiers articles illustrent l’évolution substantielle du répertoire de gènes dans certaines familles, mais aussi la divergence du patron d’expression des gènes homologues.

Article 1: nous avons reconstruit l’histoire évolutive de la famille des gènes *mgp/bgp*: chaque paralogue, ancestral aux gnathostomes, a subi une duplication spécifique dans des lignées différentes : *mgp* chez les chondrichthyens et *bgp* chez les ostéichthyens. Dans la littérature, le gène *mgp* des poissons osseux est associé à l’inhibition de la minéralisation tandis que le gène *bgp1* est associé au dépôt d’hydroxyapatite. En outre, il n’existe aucune description de l’expression du gène *bgp2* dans la littérature. Chez les chondrichthyens, nous avons montré que le gène *mgp1* présente un profil d’expression conservé par rapport au gène *mgp* des pois-

sous osseux, tandis que le paralogue *mgp2* présente un profil d'expression dans les tissus non minéralisés, probablement en raison d'une néofonctionnalisation. En ce qui concerne le *bgp* des chondrichthyens, nous n'avons trouvé aucune expression congruente avec l'hypothèse de l'activation de la minéralisation. Ces résultats suggèrent qu'un gène *mgp* ancestral de gnathostome avait un rôle dans l'inhibition de la minéralisation, y compris dans le processus de biominéralisation du squelette, tandis que le gène ancestral *bgp* a été soit recruté chez les poissons osseux, soit perdu chez les chondrichthyens.

Articles 2 et 3: Nous avons montré que les chondrichthyens, les téléostéens et les tétrapodes diffèrent dans les profils d'expression des gènes *sparc*, *sparc-L* et *scpps*. Tout d'abord, *sparc* est exprimé uniquement dans les tissus en cours de minéralisation de l'endosquelette des chondrichthyens. Par ailleurs, il est exprimé dans les ostéoblastes et les odontoblastes des téléostéens et des tétrapodes (P. Holland et al., 1987; N. Li et al., 2009). Ceci suggère qu'un gène ancestral *sparc* était déjà exprimé dans tous ces tissus avant la diversification des gnathostomes. A l'inverse, les *scpps* présentent un profil d'expression commun dans les améloblastes des élastombranches et des ostéichthyens. Les *scpps* des élastombranches et des ostéichthyens proviennent de duplications parallèles (ou indépendantes) du gène *sparc-L*. Ces gènes sont donc convergents et n'ont pas été hérités de l'ancêtre commun des vertébrés à mâchoire. De plus, étant donné que les gènes *scpps* des élastombranches et des ostéichthyens sont issus de duplications indépendantes de leur *sparc-L* respectif et qu'ils ont conservé des fonctions similaires, cela suggère que le gène ancestral *sparc-L* jouait déjà un rôle ancestral dans les améloblastes. Par conséquent, les gènes sont convergents, mais leurs patrons d'expression sont ancestraux car hérités de *sparc-L*.

Ces articles mettent en évidence les variations de répertoires de gènes et de patrons d'expression que l'on peut observer en comparant les lignées de gnathostomes. Nous avons observé des gènes qui ont conservé des caractéristiques ancestrales d'expression (*mgp1* et *sparc*) et des gènes qui ont acquis de nouvelles caractéristiques (*bgp1* et *mgp2*). De manière surprenante, nous avons même observé des gènes issus de duplications indépendantes du même gène qui présentent encore des profils d'expression ancestraux (*scpps*). La compréhension de cette dynamique est essentielle pour comprendre les paramètres initiaux à l'origine des processus de minéralisation chez les vertébrés.

### III – Une approche transcriptomique

Dans ce chapitre, j'explique les différentes étapes, les défis et les résultats de cette approche transcriptomique. Pour isoler les gènes potentiellement liés au processus de minéralisation chez la petite roussette, nous avons utilisé des analyses d'expression différentielle, et plusieurs méthodes de sélection supplémentaires (Z-scores issus d'autres données transcriptomiques). Un ensemble de gènes a été sélectionné et nous avons poursuivi l'étude de ces gènes par des tests HiS ainsi que par des phylogénies (et synténie si nécessaire) et des analyses comparatives avec

des poissons osseux, afin de comprendre les fonctions putatives de ces gènes et leur histoire évolutive.

Ce projet est la première tentative d’exploration du processus de minéralisation chez les chondrichthyens par une approche transcriptomique. Cependant, comme les gènes sont sélectionnés par expression différentielle, de nombreux processus biologiques peuvent influencer cette analyse. Par exemple, même dans des tissus “purs” comme le cartilage de chondrichthyen, les marqueurs de la différenciation cellulaire autre que la minéralisation peuvent brouiller le signal recherché. Enfin, même un tissu “pur” n’est pas composé d’une seule lignée cellulaire, il contient plusieurs signatures moléculaires, par exemple dans le cartilage on peut trouver des fibroblastes et des chondrocytes.

Nous avons démontré que malgré un échantillonnage sous-optimal, les gènes associés à la minéralisation, tels que le *scpp* des chondrichthyens, peuvent être identifiés par notre approche transcriptomique, ceci permet de nous affranchir des a priori issus de la situation très dérivée des poissons osseux, sur laquelle nous nous sommes fortement appuyée dans le chapitre précédent. En revanche, d’autres gènes précédemment associés à la minéralisation ne sont pas ressortis par cette approche (e.g. *collagène X* pas présent dans les tissus endosquelettiques, ou *mgp* qui ne montre pas de variation significative entre les deux stades de développement étudiés).

Finalement, nous avons identifié plusieurs gènes spécifiquement exprimés dans le cartilage de la petite roussette, en particulier la famille de gènes *spp2* qui semble être associée au processus de minéralisation, avec le nombre de copies qui semble être positivement corrélé au degré de minéralisation chez les espèces de chondrichthyens.

## IV - Les gènes qui ne veulent pas tomber dans les rangs

Dans ce chapitre, je me suis concentré sur les gènes qui évoluent en complexes de gènes, par duplications en tandem et sur la dynamique évolutive qui sous-tend ces structures génomiques. En général, un complexe de gènes est défini comme plusieurs paralogues qui codent pour des protéines similaires et sont situés l’un à côté de l’autre sur le même chromosome. Un complexe de gènes peut résulter de plusieurs duplications en tandem sur un locus particulier, de manière exacerbée par la recombinaison non homologue si les séquences dupliquées sont très similaires (Hastings et al., 2009). En général, les gènes dupliqués, qu’ils soient générés par des duplications en tandem ou par des duplications complètes de génome, sont initialement fonctionnellement redondants et, par conséquent, l’un des duplicats acquiert une nouvelle fonction (néofonctionnalisation), dégénère (pseudogénéisation) ou partage la fonction ancestrale avec l’autre duplicat (sous-fonctionnalisation ; modèle DDC, Force et al., 1999). Cependant, lorsque de nombreuses duplications sont bénéfiques (comme dans le cas de gènes dose-dépendants), il peut y avoir une sélection positive pour la conservation des duplicats, qui entraîne un taux de duplication apparent plus élevé (Hahn, 2009; Margres et al., 2017; Schrider and Hahn, 2010).

Les duplications en tandem peuvent être observées dans de nombreuses familles de gènes

étudiées dans ce manuscrit, comme les gènes *scpps* chez les poissons osseux, les gènes *spp2* chez les poissons cartilagineux et les gènes *slrps* et *mmps* chez les gnathostomes. Toutes ces familles de gènes codent pour des protéines structurales (ou enzymatiques, pour les *mmps*) de la matrice extracellulaire. En particulier pour les *slrps* et les *mmps*, il existait déjà un groupe ancestral de ces gènes avant la diversification des vertébrés, ce qui suggère un répertoire déjà diversifié chez leur dernier ancêtre commun. Cependant, la manière dont ces complexes ancestraux ont évolué par la suite varie fortement en fonction de la famille étudiée.

Les *slrps* ont conservé une signature claire des deux événements de duplications du génome entier qui ont eu lieu avant la diversification des vertébrés à mâchoire. Par conséquent, le complexe ancestral de quatre gènes a été dupliqué de manière à produire quatre complexes de *slrps* dans différents chromosomes. Bien que tous les paralogues n'aient pas été conservés chez toutes les espèces, les *slrps* présentent une certaine stabilité en termes de conservation des duplicats. Cette stabilité est rare après les événements de duplication du génome entier, car on suppose qu'ils sont suivis de réorganisations génomiques et de suppressions de gènes massives et rapides (Hufton and Panopoulou, 2009; Sémon and Wolfe, 2007 mais aussi voir Berthelot et al., 2014).

De manière opposée, le complexe de *mmps* s'est largement diversifié dans les lignées de gnathostomes par plusieurs séries de duplications en tandem. La copie ancestrale ayant généré ces récentes duplications, n'est pas la même chez les poissons cartilagineux et chez les poissons osseux. Cela suggère que des exigences similaires pour des répertoires *mmp* plus grands ont pu exister chez ces espèces, produisant une amplification convergente du nombre de copies dans les deux lignées. Par rapport aux *slrps*, les *mmps* ne sont pas aussi conservés et présentent des taux de renouvellement et de duplication élevés.

Ces différences entre *slrps* et *mmps* illustrent deux types de complexes : un très bien conservé à l'échelle des gnathostomes, et un autre moins bien conservé et apparemment très dynamique. L'origine de ces différences de dynamique évolutive reste à déterminer, cependant, les *slrps* semblent s'être stabilisés après la diversification des vertébrés alors que les *mmps* ont continué à se dupliquer et à évoluer par la suite. Parmi les gènes étudiés dans ce manuscrit, la situation des *mmps* semble être plus fréquente puisque des dynamiques similaires peuvent être observées chez les *scpps* de poissons osseux ainsi que chez les *spp2* de poissons cartilagineux. La condition des *slrps* est plus intrigante et les contraintes ayant mené à leur meilleure conservation devraient être étudiées plus en profondeur.

## V - Discussion générale

### Homologie de tissu

Pour un biologiste Évo-Devo, la situation idéale est celle où toutes les relations d'homologie entre les structures sont connues et comprises. Cependant, les organismes complexes, tels que

les vertébrés, ne sont pas très enclins à laisser percer leurs secrets. Cela vient notamment du fait que plusieurs niveaux d'organisation sont intégrés les uns aux autres, de sorte que les relations d'homologie deviennent floues. Ce problème est encore exacerbé dans le cas de processus s'étendant sur de longues périodes évolutives, qui ont tendance à créer des structures dérivées qui compliquent encore plus le retraçage de l'homologie. Dans le cas, entre autres, du développement du squelette des vertébrés, il s'agit d'un problème qui doit toujours être pris en compte afin que nos déductions restent aussi fidèles que possible à la réalité. Par exemple, l'homologie des tissus du squelette chez les vertébrés a fait l'objet d'un débat intense. Dans les odontodes, la dentine est généralement considérée comme un tissu stable, homologue au sein des vertébrés, mais l'émail et l'émailloïde sont considérés comme des tissus dérivés, avec des relations d'homologie qui ont été longtemps débattues (et le sont encore), d'abord sur la base de la morphologie tissulaire (Janvier, 1996), puis sur la composition protéique, qui dépend elle-même de l'expression génique spécifique à la cellule (Kawasaki and Weiss, 2008). L'un des principaux arguments contre l'homologie entre l'émail et l'émailloïde est la contribution des odontoblastes dans la formation des fibres de collagène I dans l'émailloïde (Ørvig, 1967). Cependant, ce pourrait être une condition ancestrale qui a été perdue par la suite dans l'émail des mammifères (ou plus tôt Qu et al., 2015). Ainsi, dans les odontodes, l'homologie tissulaire n'est pas claire et dépend de l'histologie, qui est fortement influencée par le contenu en protéines, lui-même directement lié à l'expression des gènes, elle-même contrôlée par les cellules. Il est donc essentiel de mieux connaître ces cellules à différents niveaux d'organisation pour mieux comprendre leurs relations au cours de l'évolution. Cependant, il faut tenir compte du fait que, comme pour l'évolution des séquences et des caractéristiques morphologiques, travailler sur les tissus et les cellules nécessite un échantillonnage suffisamment important et un outgroup bien choisi qui permet de polariser les changements de caractères.

Dans les tissus squelettiques, le cartilage semble être un tissu très ancien et est probablement homologue chez tous les vertébrés et peut-être même chez les bilatériens (Tarazona et al., 2016, mais voir ci-dessous). L'os, par contre, est un cas plus complexe puisque trois types d'ossification (endochondrale, péri-chondrale, dermique) existent, qui ne se développent pas de la même manière. On peut considérer que ces tissus sont morphologiquement différents parce qu'ils ne reçoivent pas la même contribution de différents types cellulaires (par exemple les chondrocytes sont absents dans l'ossification dermique) mais qu'ils sont en fait homologues à condition qu'ils dérivent d'une forme d'os ancestrale. Pour illustrer cela, une étude de 2016 (Weigele and Franz-Odenaal, 2016) sur l'histologie du poisson zèbre adulte a révélé que la plupart des os du crâne du poisson zèbre sont des os compacts cellulaires ou acellulaires (formés par ossification dermique). Les autres os sont formés par deux processus d'ossification endochondrale différents, appelés type I et II. L'ossification endochondrale de type I est similaire à l'ossification endochondrale des mammifères, tandis que le type II est dépourvu à la fois d'une zone de dégradation et d'une zone de minéralisation adjacente aux chondrocytes hypertrophiques. Les os formés par l'ossification endochondrale de type II sont appelés os tubulaires.

Ils sont dépourvus de réseau trabéculaire osseux, de vaisseaux sanguins et de nerfs et sont remplis de tissu adipeux. Les auteurs de l'article émettent l'hypothèse qu'ils représentent une caractéristique squelettique dérivée spécifique au poisson zèbre. Dans ce cas, les nouveaux os tubulaires du poisson-zèbre sont très probablement dérivés de l'os endochondral "classique", et sont simplement différents sur le plan morphologique parce qu'il leur manque une contribution des ostéoblastes dans la partie la plus interne de l'os.

Par conséquent, le problème de l'homologie tissulaire semble se résumer à savoir quelles sont les cellules qui participent à la formation du tissu, et à reporter la question de l'homologie des tissus vers celle de l'homologie des cellules.

## Homologie de cellule

Classiquement, il existe deux façons de typer une cellule : par ses attributs morphologiques (par exemple, les bâtonnets et les cônes de la rétine) ou par sa fonction. La fonction cellulaire dépend de nombreuses protéines sécrétées par la cellule, et trouver cette signature moléculaire est une approche courante pour déterminer un type de cellule. D'autres études ont tenté de caractériser les cellules par leur profil d'expression génétique, principalement grâce à des approches à haut débit telles que le *single-cell* RNAseq. D'un point de vue évolutif, ces caractérisations ne sont pas optimales et ont leurs limites. La similarité des types cellulaires entre groupes taxonomiques, et la similarité entre cellules d'un même organisme, peut évoluer par le biais de plusieurs processus (tels que la véritable homologie par héritage, mais aussi la convergence ou l'évolution concertée ; Arendt et al., 2016) et les caractérisations actuelles des types cellulaires ne peuvent pas différencier facilement ces scénarios. Des définitions plus récentes des types cellulaires sont basées sur le concept de réseaux de régulation des gènes (*gene regulatory network*; GRN), représentant une signature moléculaire spécifique à un seul type de cellule, influençant la morphologie et la fonction et permettant à la cellule d'agir comme une seule unité évolutive (Arendt et al., 2016). Ce concept est illustré dans Tarazona et al., 2016 où les auteurs explorent différentes hypothèses pour l'évolution des chondrocytes chez les bilatériens. Deux hypothèses principales s'affrontent dans cet article : (i) l'origine indépendante des chondrocytes s'est produite par le recrutement d'un GRN homologue, dans un même type de progéniteur, lui-même homologue, ou (ii) une évolution indépendante par l'activation du même GRN homologue mais dans différents types de progéniteurs, non homologues. Ces résultats soulignent que l'homologie des types cellulaires ne peut pas toujours être définie car elle intègre deux niveaux indépendants d'homologie : l'origine développementale des cellules et les GRNs. Par conséquent, les GRNs ne peuvent pas constituer à eux seuls le caractère déterminant d'un type de cellule. Cependant, il semble que des GRN homologues dans des progéniteurs cellulaires différents peuvent déclencher la différenciation de types cellulaires similaires, donnant lieu à des tissus morphologiquement et fonctionnellement similaires.

Par conséquent, la meilleure façon d'étudier l'évolution des tissus à travers le prisme de

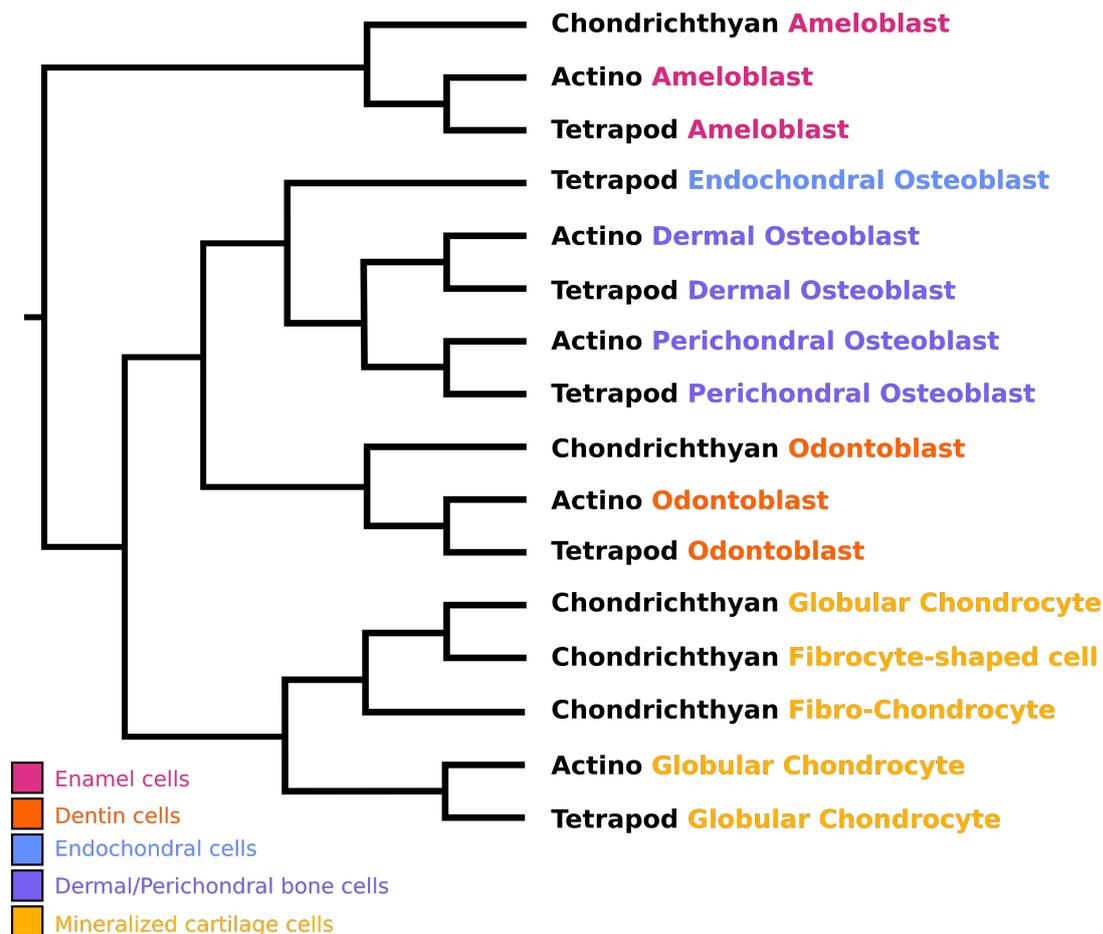
l'évolution est de comprendre comment les GRN ont été co-optés dans les différents types cellulaires. Ainsi, ce que nous appelons os chez les poissons osseux serait un tissu formé par un type cellulaire contenant un GRN homologue commun spécialisé dans la sécrétion de collagène et la précipitation du phosphate de calcium. Cette hypothèse implique l'existence d'un GRN commun à toutes les cellules de type ostéoblaste. Cependant, de nombreux autres types cellulaires minéralisent eux aussi leur matrice, notamment les odontoblastes, les améloblastes et les chondrocytes. Existe-t-il un GRN commun à tous ces types cellulaires ? Plusieurs études ont déjà soutenu l'existence d'un ancien réseau génétique "squelettique" (revue dans Fisher and Franz-Odenaal, 2012), cependant, elles se concentrent principalement sur les gènes maîtres régulateurs de la différenciation cellulaire. Dans cette discussion, j'étendrai la question de cet ancien GRN du squelette, à l'ensemble des gènes avec lesquels j'ai travaillé : les acteurs proximaux de la minéralisation. Pouvons-nous décrire un GRN ancestral (appelé ici toolkit de minéralisation), spécifique à certains ou commun à tous les types de cellules squelettiques minéralisantes ? Nos données sont-elles suffisantes pour caractériser les cellules squelettiques dès la séparation entre chondrichthyens et osteichthyens ? Peut-on conclure à la perte ou non d'os chez les poissons cartilagineux ? Et peut-on classer ces cellules en fonction de leur ressemblance en termes d'expression génétique ?

## **Le toolkit de minéralisation des gnathostomes**

Au cours de cette thèse, une variété de gènes, de cellules et de tissus a été étudiée afin de discuter de leur évolution, de leurs relations d'homologie et, éventuellement de leur origine. Ces gènes sont impliqués dans le développement du squelette et plus particulièrement dans le processus de minéralisation. L'objectif était de décrire ces gènes chez un représentant des chondrichthyens afin de comparer leur expression chez les vertébrés à mâchoires. Cette analyse comparative permet d'inférer une partie d'un GRN squelettique ancestral des gnathostomes : la boîte à outils de la minéralisation. Dans ce manuscrit, l'accent a été mis sur plusieurs protéines structurelles mais aussi sur des protéines sécrétées directement liées à la précipitation du phosphate de calcium.

Notre hypothèse "nulle" considère l'évolution indépendante des lignées cellulaires que nous avons confrontée aux données acquises chez la petite roussette et aux données déjà publiées chez d'autres espèces. La reconstruction d'un arbre sur la base de ces caractères, sous un modèle d'évolution par maximum de parcimonie, nous permet de proposer un scénario d'évolution du GRN des cellules squelettiques chez les vertébrés à mâchoire, comme le montre la figure ci dessous. Dans cette figure, certaines relations supposent que le GRN de certaines cellules évoluent de manière indépendante (améloblastes). Les odontoblastes sont le groupe frère des ostéoblastes, mais nos résultats peuvent aussi soutenir une évolution concertée des odontoblastes et ostéoblastes dans la lignée des poissons osseux. Les cellules cartilagineuses minéralisées sont monophylétiques, celles des chondrichthyens formant un groupe monophylétique, produit soit

parce qu'elles sont des innovations spécifiques à la lignée, soit parce que leur GRN a évolué de manière concertée dans cette lignée. Ce scénario devrait être affiné en utilisant davantage de données sur l'expression des gènes qui pourraient mettre en évidence les interrelations entre les types cellulaires, comme l'évolution concertée des lignées, la convergence et/ou l'innovation récente, renforçant ainsi notre compréhension des GRN, de l'évolution des cellules et des tissus dans les squelettes des vertébrés à mâchoire.



Scénario évolutif des cellules et tissus associés à la minéralisation chez les vertébrés à mâchoire

## VI - Conclusion

Dans cette thèse, j'ai étudié une variété de gènes, connus pour être associés à la minéralisation. J'ai démontré que certains de ces gènes sont impliqués dans la minéralisation depuis plus de 400 millions d'années (avant la séparation entre ostéichthyens / chondrichthyens) alors que d'autres sont des acquisitions plus récentes selon la lignée, mettant en évidence la dynamique de l'évolution de ces gènes au cours de l'évolution des vertébrés. Ces résultats font des chondrichthyens, et de la petite roussette en particulier, un groupe taxonomique précieux pour les analyses comparatives et l'inférence des états ancestraux chez les vertébrés à mâchoire. Ceci

est vrai pour l'objet principal de ma thèse (la boîte à outils de minéralisation) mais aussi pour de nombreux autres processus et structures. Bien que les chercheurs commencent à se rendre compte de la valeur de ces espèces en tant que modèles biologiques, il est encore assez difficile de travailler sur elles, en particulier dans le domaine de l'Evo-Devo, en raison du manque de matériel embryonnaire, de leur long cycle de vie, de leurs exigences écologiques, ou encore de leur grande taille. La poursuite des recherches sur ces espèces permettra sans aucun doute de renforcer notre compréhension de leur histoire évolutive et de celle du groupe des vertébrés.

# Bibliography

- Åberg, T., Rice, R., Rice, D., Thesleff, I., & Waltimo-Sirén, J. (2005). Chondrogenic potential of mouse calvarial mesenchyme. *Journal of Histochemistry & Cytochemistry*, 53(5), 653–663.
- Abzhanov, A. (2013). Von baer’s law for the ages: Lost and found principles of developmental evolution. *Trends in Genetics*, 29(12), 712–722.
- Abzhanov, A., Rodda, S. J., McMahon, A. P., & Tabin, C. J. (2007). Regulation of skeletogenic differentiation in cranial dermal bone.
- Agassiz, L. (1844). *Monographie des poissons fossiles du vieux grès rouge ou système dévonien (old red sandstone) des îles britanniques et de russie ouvrage rédigé a la demande de l’association britannique... par lui agassiz:° testo* (Vol. 1). Soleure, Jent & Gassmann.
- Ai-Aql, Z., Alagl, A. S., Graves, D. T., Gerstenfeld, L. C., & Einhorn, T. A. (2008). Molecular mechanisms controlling bone formation during fracture healing and distraction osteogenesis. *Journal of dental research*, 87(2), 107–118.
- Albertson, R. C., Yan, Y.-L., Titus, T. A., Pisano, E., Vacchi, M., Yelick, P. C., Detrich, H. W., & Postlethwait, J. H. (2010). Molecular pedomorphism underlies craniofacial skeletal evolution in antarctic notothenioid fishes. *BMC evolutionary biology*, 10(1), 1–12.
- Aldea, D., Hanna, P., Munoz, D., Espinoza, J., Torrejon, M., Sachs, L., Buisine, N., Oulion, S., Escriva, H., & Marcellini, S. (2013). Evolution of the vertebrate bone matrix: An expression analysis of the network forming collagen paralogues in amphibian osteoblasts. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, 320(6), 375–384.
- Aleksandra, M., Hancock-Ronemus, A., & Andrew, G. J. (2020). Adult chondrogenesis and spontaneous cartilage repair in the skate, *leucoraja erinacea*. *eLife*, 9.
- Arendt, D., Musser, J. M., Baker, C. V., Bergman, A., Cepko, C., Erwin, D. H., Pavlicev, M., Schlosser, G., Widder, S., Laubichler, M. D., et al. (2016). The origin and evolution of cell types. *Nature Reviews Genetics*, 17(12), 744–757.
- Arratia, G., Schultze, H.-P., & Casciotta, J. (2001). Vertebral column and associated elements in dipnoans and comparison with other fishes: Development and homology. *Journal of Morphology*, 250(2), 101–172.

- Assaraf-Weill, N., Gasse, B., Silvent, J., Bardet, C., Sire, J.-Y., & Davit-Béal, T. (2014). Ameloblasts express type I collagen during amelogenesis. *Journal of dental research*, *93*(5), 502–507.
- Atake, O. J., Cooper, D. M., & Eames, B. F. (2019). Bone-like features in skate suggest a novel elasmobranch synapomorphy and deep homology of trabecular mineralization patterns. *Acta biomaterialia*, *84*, 424–436.
- Atake, O. J., & Eames, B. F. (2021). Mineralized cartilage and bone-like tissues in chondrichthyans offer potential insights into the evolution and development of mineralized tissues in the vertebrate endoskeleton. *Frontiers in Genetics*, 2615.
- Avila Cobos, F., Alquicira-Hernandez, J., Powell, J. E., Mestdagh, P., & De Preter, K. (2020). Benchmarking of cell type deconvolution pipelines for transcriptomics data. *Nature communications*, *11*(1), 1–14.
- Baanannou, A., Rastegar, S., Bouzid, A., Takamiya, M., Gerber, V., Souissi, A., Beil, T., Jrad, O., Strähle, U., & Masmoudi, S. (2020). Gene duplication and functional divergence of the zebrafish otospiralin genes. *Development Genes and Evolution*, *230*(1), 27–36.
- Baddar, N. A. H., Timoshevskaya, N., Smith, J. J., Guo, H., & Voss, S. R. (2021). Novel expansion of matrix metalloproteases in the laboratory axolotl (*Ambystoma mexicanum*) and other salamander species. *Frontiers in Ecology and Evolution*, *9*. <https://doi.org/10.3389/fevo.2021.786263>
- Ballard, W. W., Mellinger, J., & Lechenault, H. (1993). A series of normal stages for development of *Scyliorhinus canicula*, the lesser spotted dogfish (chondrichthyes: Scyliorhinidae). *Journal of Experimental Zoology*, *267*(3), 318–336.
- Bej, A., Sahoo, B. R., Swain, B., Basu, M., Jayasankar, P., & Samanta, M. (2014). LRRsearch: An asynchronous server-based application for the prediction of leucine-rich repeat motifs and an integrative database of NOD-like receptors. *Computers in Biology and Medicine*, *53*, 164–170. <https://doi.org/10.1016/j.compbiomed.2014.07.016>
- Bellido, T. (2014). Osteocyte-driven bone remodeling. *Calcified tissue international*, *94*(1), 25–34.
- Benton, M. J. (2014). *Vertebrate palaeontology*. John Wiley & Sons.
- Berendsen, A. D., & Olsen, B. R. (2015). Bone development. *Bone*, *80*, 14–18.
- Berio, F., Broyon, M., Enault, S., Pirot, N., López-Romero, F. A., & Debais-Thibaud, M. (2021). Diversity and evolution of mineralized skeletal tissues in chondrichthyans. *Frontiers in Ecology and Evolution*, *9*, 660767.
- Berthelot, C., Brunet, F., Chalopin, D., Juanchich, A., Bernard, M., Noël, B., Bento, P., Da Silva, C., Labadie, K., Alberti, A., et al. (2014). The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature communications*, *5*(1), 1–10.
- 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. *Proceedings of the Royal Society B: Biological Sciences*, 280(1757), 20122963.
- Bleicher, F., Couble, M., Farges, J., Couble, P., & Magloire, H. (1999). Sequential expression of matrix protein genes in developing rat teeth. *Matrix biology*, 18(2), 133–143.
- Bornstein, P., & Sage, E. H. (2002). Matricellular proteins: Extracellular modulators of cell function. *Current opinion in cell biology*, 14(5), 608–616.
- Boskey, A. L. (2010). The biochemistry of bone. *Osteoporosis in men* (pp. 3–13). Elsevier. <https://doi.org/10.1016/b978-0-12-374602-3.00001-8>
- Brazeau, M. D., Giles, S., Dearden, R. P., Jerve, A., Ariunchimeg, Y., Zorig, E., Sansom, R., Guillerme, T., & Castiello, M. (2020). Endochondral bone in an early devonian ‘placoderm’ from mongolia. *Nature Ecology & Evolution*, 4(11), 1477–1484.
- Brizzi, M. F., Tarone, G., & Defilippi, P. (2012). Extracellular matrix, integrins, and growth factors as tailors of the stem cell niche. *Current Opinion in Cell Biology*, 24(5), 645–651. <https://doi.org/10.1016/j.ceb.2012.07.001>
- Buchaille, R., Couble, M., Magloire, H., & Bleicher, F. (2000). Expression of the small leucine-rich proteoglycan osteoadherin/osteomodulin in human dental pulp and developing rat teeth. *Bone*, 27(2), 265–270.
- Burke, A. C., Dial, K., Shubin, N., & Brainerd, E. (2015). Origin of the turtle body plan. *Great transformations in vertebrate evolution*, 77–89.
- Burra, S., Nicoletta, D. P., Francis, W. L., Freitas, C. J., Mueschke, N. J., Poole, K., & Jiang, J. X. (2010). Dendritic processes of osteocytes are mechanotransducers that induce the opening of hemichannels. *Proceedings of the National Academy of Sciences*, 107(31), 13648–13653.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST: Architecture and applications. *BMC Bioinformatics*, 10(1). <https://doi.org/10.1186/1471-2105-10-421>
- Candeliere, G., Liu, F., & Aubin, J. (2001). Individual osteoblasts in the developing calvaria express different gene repertoires. *Bone*, 28(4), 351–361.
- Catón, J., & Tucker, A. S. (2009). Current knowledge of tooth development: Patterning and mineralization of the murine dentition. *Journal of Anatomy*, 214(4), 502–515.
- Chaumel, J., Schotte, M., Bizzarro, J. J., Zaslansky, P., Fratzl, P., Baum, D., & Dean, M. N. (2020). Co-aligned chondrocytes: Zonal morphological variation and structured arrangement of cell lacunae in tessellated cartilage. *Bone*, 134, 115264.
- Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., Clark, N. R., & Ma’ayan, A. (2013). Enrichr: Interactive and collaborative html5 gene list enrichment analysis tool. *BMC bioinformatics*, 14(1), 1–14.
- Chen, S., Gluhak-Heinrich, J., Wang, Y., Wu, Y., Chuang, H., Chen, L., Yuan, G., Dong, J., Gay, I., & MacDougall, M. (2009). Runx2, osx, and dspp in tooth development. *Journal of dental research*, 88(10), 904–909.

- Chen, S., & Birk, D. E. (2013). The regulatory roles of small leucine-rich proteoglycans in extracellular matrix assembly. *The FEBS journal*, *280*(10), 2120–2137.
- Chen, X., Ji, Y., Feng, F., Liu, Z., Qian, L., Shen, H., & Lao, L. (2022). C-type lectin domain-containing protein clec3a regulates proliferation, regeneration and maintenance of nucleus pulposus cells. *Cellular and Molecular Life Sciences*, *79*(8), 1–11.
- Chevrel, R. (1889). *Sur l'anatomie du système nerveux grand sympathique des elasmobranches et des poissons osseux*.
- Chevrisais, M., Johanson, Z., Trinajstić, K., Long, J., Morel, C., Renaud, C. B., & Cloutier, R. (2018). Evolution of vertebrate postcranial complexity: Axial skeleton regionalization and paired appendages in a devonian jawless fish. *Palaeontology*, *61*(6), 949–961.
- Clause, K. C., & Barker, T. H. (2013). Extracellular matrix signaling in morphogenesis and repair. *Current opinion in biotechnology*, *24*(5), 830–833.
- Coen, G., Ballanti, P., Silvestrini, G., Mantella, D., Manni, M., Di Giulio, S., Pisanò, S., Leopizzi, M., Di Lullo, G., & Bonucci, E. (2009). Immunohistochemical localization and mrna expression of matrix gla protein and fetuin-a in bone biopsies of hemodialysis patients. *Virchows Archiv*, *454*(3), 263–271.
- Cohen, S. M., & Jürgens, G. (1989). Proximal—distal pattern formation in drosophila: Cell autonomous requirement for distal-less gene activity in limb development. *The EMBO journal*, *8*(7), 2045–2055.
- Conrad, A. H., & Conrad, G. W. (2003). The keratocan gene is expressed in both ocular and non-ocular tissues during early chick development. *Matrix Biology*, *22*(4), 323–337. [https://doi.org/10.1016/s0945-053x\(03\)00039-8](https://doi.org/10.1016/s0945-053x(03)00039-8)
- Costa, R. A., Martins, R. S. T., Capilla, E., Anjos, L., & Power, D. M. (2018). Vertebrate SLRP family evolution and the subfunctionalization of osteoglycin gene duplicates in teleost fish. *BMC Evolutionary Biology*, *18*(1). <https://doi.org/10.1186/s12862-018-1310-2>
- Criswell, K. E., Coates, M. I., & Gillis, J. A. (2017). Embryonic development of the axial column in the little skate, *leucoraja erinacea*. *Journal of Morphology*, *278*(3), 300–320.
- Dan, H., Simsa-Maziel, S., Reich, A., Sela-Donenfeld, D., & Monsonego-Ornan, E. (2012). The role of matrix gla protein in ossification and recovery of the avian growth plate. *Frontiers in Endocrinology*, *3*, 79.
- Davesne, D., Meunier, F. J., Schmitt, A. D., Friedman, M., Otero, O., & Benson, R. B. (2019). The phylogenetic origin and evolution of acellular bone in teleost fishes: Insights into osteocyte function in bone metabolism. *Biological Reviews*, *94*(4), 1338–1363.
- Daza, D. O., Bergqvist, C. A., & Larhammar, D. (2022). The evolution of oxytocin and vasotocin receptor genes in jawed vertebrates: A clear case for gene duplications through ancestral whole-genome duplications. *Frontiers in Endocrinology*, *12*. <https://doi.org/10.3389/fendo.2021.792644>

- Dean, M., Mull, C. G., Gorb, S. N., & Summers, A. P. (2009). Ontogeny of the tessellated skeleton: Insight from the skeletal growth of the round stingray *urobatis halleri*. *Journal of Anatomy*, *215*(3), 227–239.
- Dean, M., & Summers, A. (2006). Mineralized cartilage in the skeleton of chondrichthyan fishes. *Zoology*, *109*(2), 164–168.
- Dean, M., Socha, J., Hall, B., & Summers, A. (2010). Canaliculi in the tessellated skeleton of cartilaginous fishes. *Journal of Applied Ichthyology*, *26*(2), 263–267.
- Debiais-Thibaud, M. (2019). The evolution of endoskeletal mineralisation in chondrichthyan fish. *Evolution and development of fishes*, 110–125.
- Debiais-Thibaud, M., Chiori, R., Enault, S., Oulion, S., Germon, I., Martinand-Mari, C., Casane, D., & Borday-Birraux, V. (2015). Tooth and scale morphogenesis in shark: An alternative process to the mammalian enamel knot system. *BMC Evolutionary Biology*, *15*(1), 1–17.
- Debiais-Thibaud, M., Oulion, S., Bourrat, F., Laurenti, P., Casane, D., & Borday-Birraux, V. (2011). The homology of odontodes in gnathostomes: Insights from *dlx* gene expression in the dogfish, *scyliorhinus canicula*. *BMC evolutionary biology*, *11*(1), 1–14.
- Debiais-Thibaud, M., Simion, P., Ventéo, S., Muñoz, D., Marcellini, S., Mazan, S., & Haitina, T. (2019). Skeletal mineralization in association with type x collagen expression is an ancestral feature for jawed vertebrates. *Molecular biology and evolution*, *36*(10), 2265–2276.
- de Castro, E., Sigrist, C. J. A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., Bairoch, A., & Hulo, N. (2006). ScanProsite: Detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Research*, *34*(Web Server), W362–W365. <https://doi.org/10.1093/nar/gkl124>
- Dehal, P., & Boore, J. L. (2005). Two rounds of whole genome duplication in the ancestral vertebrate (P. Holland, Ed.). *PLoS Biology*, *3*(10), e314. <https://doi.org/10.1371/journal.pbio.0030314>
- Delprat, B., Ruel, J., Guitton, M. J., Hamard, G., Lenoir, M., Pujol, R., Puel, J.-L., Brabet, P., & Hamel, C. P. (2005). Deafness and cochlear fibrocyte alterations in mice deficient for the inner ear protein otospiralin. *Molecular and cellular biology*, *25*(2), 847–853.
- Denison, R. H. (1947). The exoskeleton of tremataspis. *American Journal of Science*, *245*(6), 337–365.
- Denison, R. H. (1967). Ordovician vertebrates from western united states.
- Denison, R. H. (1978). *Placodermi* (Vol. 2). München [Germany].
- Dhouailly, D., Godefroit, P., Martin, T., Nonchev, S., Caraguel, F., & Oftedal, O. (2019). Getting to the root of scales, feather and hair: As deep as odontodes? *Experimental dermatology*, *28*(4), 503–508.

- 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 Evolutionary Biology*, *19*(1), 1–17.
- Donoghue, P. C., & Sansom, I. J. (2002). Origin and early evolution of vertebrate skeletonization. *Microscopy research and technique*, *59*(5), 352–372.
- Dornburg, A., & Near, T. J. (2021). The emerging phylogenetic perspective on the evolution of actinopterygian fishes. *Annual Review of Ecology, Evolution, and Systematics*, *52*, 427–452.
- Doyle, J. (1968). Ageing changes in cartilage from squalus acanthias l. *Comparative biochemistry and physiology*, *25*(1), 201–206.
- Eames, B. F., Allen, N., Young, J., Kaplan, A., Helms, J. A., & Schneider, R. A. (2007). Skeletogenesis in the swell shark cephaloscyllium ventriosum. *Journal of anatomy*, *210*(5), 542–554.
- Eames, B. F., Amores, A., Yan, Y.-L., & Postlethwait, J. H. (2012). Evolution of the osteoblast: Skeletogenesis in gar and zebrafish. *BMC evolutionary biology*, *12*(1), 1–13.
- Eliaz, N., & Metoki, N. (2017). Calcium phosphate bioceramics: A review of their history, structure, properties, coating technologies and biomedical applications. *Materials*, *10*(4), 334.
- Ellis, J., & Shackley, S. (1997). The reproductive biology of scyliorhinus canicula in the bristol channel, uk. *Journal of Fish Biology*, *51*(2), 361–372.
- Enault, S., Adnet, S., & Debiais-Thibaud, M. (2016). Skeletogenesis during the late embryonic development of the catshark scyliorhinus canicula (chondrichthyes; neoselachii). *MorphoMuseum*, *1*(4).
- Enault, S., Muñoz, D., Simion, P., Ventéo, S., Sire, J.-Y., Marcellini, S., & Debiais-Thibaud, M. (2018). Evolution of dental tissue mineralization: An analysis of the jawed vertebrate sparac and sparac-l families. *BMC evolutionary biology*, *18*(1), 1–12.
- Enault, S., Muñoz, D. N., Silva, W. T., Borday-Birraux, V., Bonade, M., Oulion, S., Ventéo, S., Marcellini, S., & Debiais-Thibaud, M. (2015). Molecular footprinting of skeletal tissues in the catshark scyliorhinus canicula and the clawed frog xenopus tropicalis identifies conserved and derived features of vertebrate calcification. *Frontiers in genetics*, *6*, 283.
- Ernst, K., & Serres, R. A. (1828). Karl Ernst von Baer ' s Laws of Embryology, 17–20.
- Espinoza, J., Sanchez, M., Sanchez, A., Hanna, P., Torrejon, M., Buisine, N., Sachs, L., & Marcellini, S. (2010). Two families of xenopus tropicalis skeletal genes display well-conserved expression patterns with mammals in spite of their highly divergent regulatory regions. *Evolution & development*, *12*(6), 541–551.
- Fanjul-Fernández, M., Folgueras, A. R., Cabrera, S., & López-Otin, C. (2010). Matrix metalloproteinases: Evolution, gene regulation and functional analysis in mouse models. *Biochimica et Biophysica Acta (BBA) - Molecular Cell Research*, *1803*(1), 3–19. <https://doi.org/10.1016/j.bbamcr.2009.07.004>

- Fisher, S., & Franz-Odenaal, T. (2012). Evolution of the bone gene regulatory network. *Current opinion in genetics & development*, 22(4), 390–397.
- Force, A., Lynch, M., Pickett, F. B., Amores, A., Yan, Y.-l., & Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics*, 151(4), 1531–1545.
- Franco, A. D., Poujol, R., Baurain, D., & Philippe, H. (2019). Evaluating the usefulness of alignment filtering methods to reduce the impact of errors on evolutionary inferences. *BMC Evolutionary Biology*, 19(1). <https://doi.org/10.1186/s12862-019-1350-2>
- Funderburgh, J. L., Corpuz, L. M., Roth, M. R., Funderburgh, M. L., Tasheva, E. S., & Conrad, G. W. (1997). Mimecan, the 25-kDa corneal keratan sulfate proteoglycan, is a product of the gene producing osteoglycin. *Journal of Biological Chemistry*, 272(44), 28089–28095. <https://doi.org/10.1074/jbc.272.44.28089>
- Gao, Y., Jheon, A., Nourkeyhani, H., Kobayashi, H., & Ganss, B. (2004). Molecular cloning, structure, expression, and chromosomal localization of the human osterix (sp7) gene. *Gene*, 341, 101–110.
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). ExPASy: The proteomics server for in-depth protein knowledge and analysis. *Nucleic acids research*, 31(13), 3784–3788.
- Gavaia, P. J., Simes, D. C., Ortiz-Delgado, J., Viegas, C. S., Pinto, J. P., Kelsh, R. N., Sarasquete, M. C., & Cancela, M. L. (2006). Osteocalcin and matrix gla protein in zebrafish (*Danio rerio*) and senegal sole (*Solea senegalensis*): Comparative gene and protein expression during larval development through adulthood. *Gene Expression Patterns*, 6(6), 637–652.
- George, S. (2004). MMPs, cadherins, and cell proliferation. *Trends in Cardiovascular Medicine*, 14(3), 100–105. <https://doi.org/10.1016/j.tcm.2003.12.008>
- Gifford, V., & Itoh, Y. (2019). MT1-MMP-dependent cell migration: Proteolytic and non-proteolytic mechanisms. *Biochemical Society Transactions*, 47(3), 811–826. <https://doi.org/10.1042/bst20180363>
- Gilbert, S. F. (2003). The morphogenesis of evolutionary developmental biology. *International Journal of Developmental Biology*, 47(7-8), 467.
- Giles, S., Rücklin, M., & Donoghue, P. C. (2013). Histology of “placoderm” dermal skeletons: Implications for the nature of the ancestral gnathostome. *Journal of Morphology*, 274(6), 627–644.
- Gillis, J. A. (2019). The development and evolution of cartilage. *Reference module in life sciences*. Elsevier. <https://doi.org/10.1016/b978-0-12-809633-8.90770-2>
- Gistelincq, C., Gioia, R., Gagliardi, A., Tonelli, F., Marchese, L., Bianchi, L., Landi, C., Bini, L., Huysseune, A., Witten, P., et al. (2016). Zebrafish collagen type i: Molecular and biochemical characterization of the major structural protein in bone and skin. *Scientific reports*, 6(1), 1–14.

- Golubkov, V. S., Boyd, S., Savinov, A. Y., Chekanov, A. V., Osterman, A. L., Remacle, A., Rozanov, D. V., Doxsey, S. J., & Strongin, A. Y. (2005). Membrane type-1 matrix metalloproteinase (MT1-MMP) exhibits an important intracellular cleavage function and causes chromosome instability. *Journal of Biological Chemistry*, *280*(26), 25079–25086. <https://doi.org/10.1074/jbc.m502779200>
- Gordon, M. K., & Hahn, R. A. (2010). Collagens. *Cell and tissue research*, *339*(1), 247–257.
- Gould, S. (1977). Ontogeny and phylogeny.
- Grogan, E. D., Lund, R., & Greenfest-Allen, E. (2012). The origin and relationships of early chondrichthyans. *Biology of sharks and their relatives*, *2*, 3–29.
- Grover, J., & Roughley, P. J. (2001). Characterization and expression of murine PRELP. *Matrix Biology*, *20*(8), 555–564. [https://doi.org/10.1016/s0945-053x\(01\)00165-2](https://doi.org/10.1016/s0945-053x(01)00165-2)
- Grynopas, M. D., & Omelon, S. (2007). Transient precursor strategy or very small biological apatite crystals?
- Guindon, S., Dufayard, J.-F., 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. *Systematic Biology*, *59*(3), 307–321. <https://doi.org/10.1093/sysbio/syq010>
- Guirado, E., & George, A. (2021). Dentine matrix metalloproteinases as potential mediators of dentine regeneration. *European Cells and Materials*, *42*, 392–400. <https://doi.org/10.22203/ecm.v042a24>
- Gürich, G. (1891). Ueber placodermen und andere devonische fischreste im breslauer mineralogischen museum. *Zeitschrift der Deutschen Geologischen Gesellschaft*, 902–913.
- Haeckel, E. (1866). *Generelle morphologie der organismen. allgemeine grundzüge der organischen formen-wissenschaft, mechanisch begründet durch die von c. darwin reformirte descendenz-theorie, etc* (Vol. 1).
- Haeckel, E. (1892). *Evolution of man: A popular exposition of the principal points of human ontogeny & phylogeny* (Vol. 1). D. Appleton.
- Hahn, M. W. (2009). Distinguishing among evolutionary models for the maintenance of gene duplicates. *Journal of Heredity*, *100*(5), 605–617.
- Hall, B. (2014a). Endoskeleton/exo (dermal) skeleton—mesoderm/neural crest: Two pair of problems and a shifting paradigm. *Journal of Applied Ichthyology*, *30*(4), 608–615.
- Hall, B. (2014b). *Bones and cartilage: Developmental and evolutionary skeletal biology*. Academic Press.
- Hall, B. K. (1999). Agnathans. *The neural crest in development and evolution* (pp. 46–57). Springer.
- Hall, B. K. (2008). The neural crest and neural crest cells: Discovery and significance for theories of embryonic organization. *Journal of biosciences*, *33*(5), 781–793.
- Hara, Y., Yamaguchi, K., Onimaru, K., Kadota, M., Koyanagi, M., Keeley, S. D., Tatsumi, K., Tanaka, K., Motone, F., Kageyama, Y., et al. (2018). Shark genomes provide insights

- into elasmobranch evolution and the origin of vertebrates. *Nature ecology & evolution*, 2(11), 1761–1771.
- Hardingham, T. (2006). Proteoglycans and glycosaminoglycans. *Dynamics of bone and cartilage metabolism*, 2, 85–98.
- Hastings, P. J., Lupski, J. R., Rosenberg, S. M., & Ira, G. (2009). Mechanisms of change in gene copy number. *Nature Reviews Genetics*, 10(8), 551–564.
- He, G., Tavella, S., Hanley, K. P., Self, M., Oliver, G., Grifone, R., Hanley, N., Ward, C., & Bobola, N. (2010). Inactivation of *six2* in mouse identifies a novel genetic mechanism controlling development and growth of the cranial base. *Developmental biology*, 344(2), 720–730.
- He, P., Zhang, Y., Kim, S. O., Radlanski, R. J., Butcher, K., Schneider, R. A., & DenBesten, P. K. (2010). Ameloblast differentiation in the human developing tooth: Effects of extracellular matrices. *Matrix biology*, 29(5), 411–419.
- Henry, S. P., Takanosu, M., Boyd, T. C., Mayne, P. M., Eberspaecher, H., Zhou, W., De Crombrughe, B., Höök, M., & Mayne, R. (2001). Expression pattern and gene characterization of asporin: A newly discovered member of the leucine-rich repeat protein family. *Journal of Biological Chemistry*, 276(15), 12212–12221.
- Hess, P. N., & Russo, C. A. D. M. (2007). An empirical test of the midpoint rooting method. *Biological Journal of the Linnean Society*, 92(4), 669–674. <https://doi.org/10.1111/j.1095-8312.2007.00864.x>
- Hill, R. V. (2006). Comparative anatomy and histology of xenarthran osteoderms. *Journal of Morphology*, 267(12), 1441–1460.
- Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2017). UFBoot2: Improving the ultrafast bootstrap approximation. *Molecular Biology and Evolution*, 35(2), 518–522. <https://doi.org/10.1093/molbev/msx281>
- Holland, L. Z., & Daza, D. O. (2018). A new look at an old question: When did the second whole genome duplication occur in vertebrate evolution? *Genome Biology*, 19(1). <https://doi.org/10.1186/s13059-018-1592-0>
- Holland, P., Harper, S. J., McVey, J. H., & Hogan, B. (1987). In vivo expression of mrna for the  $ca^{++}$ -binding protein *sparc* (osteonectin) revealed by in situ hybridization. *The Journal of cell biology*, 105(1), 473–482.
- Holmbeck, K., Bianco, P., Caterina, J., Yamada, S., Kromer, M., Kuznetsov, S. A., Mankani, M., Robey, P. G., Poole, A. R., Pidoux, I., et al. (1999). *Mt1-mmp*-deficient mice develop dwarfism, osteopenia, arthritis, and connective tissue disease due to inadequate collagen turnover. *Cell*, 99(1), 81–92.
- Hoshi, K., Kemmotsu, S., Takeuchi, Y., Amizuka, N., & Ozawa, H. (1999). The primary calcification in bones follows removal of decorin and fusion of collagen fibrils. *Journal of bone and mineral research*, 14(2), 273–280.

- Hou, C., Liu, Z. X., Tang, K. L., Wang, M. G., Sun, J., Wang, J., & Li, S. (2011). Developmental changes and regional localization of dspp, mepe, mimecan and versican in postnatal developing mouse teeth. *Journal of Molecular Histology*, *43*(1), 9–16. <https://doi.org/10.1007/s10735-011-9368-9>
- Houari, S., Wurtz, T., Ferbus, D., Chateau, D., Dessombz, A., Berdal, A., & Babajko, S. (2014). Asporin and the mineralization process in fluoride-treated rats. *Journal of Bone and Mineral Research*, *29*(6), 1446–1455. <https://doi.org/10.1002/jbmr.2153>
- Hu, B., Coulson, L., Moyer, B., & Price, P. A. (1995). Isolation and molecular cloning of a novel bone phosphoprotein related in sequence to the cystatin family of thiol protease inhibitors. *Journal of Biological Chemistry*, *270*(1), 431–436.
- Hufton, A. L., & Panopoulou, G. (2009). Polyploidy and genome restructuring: A variety of outcomes. *Current opinion in genetics & development*, *19*(6), 600–606.
- Hultgårdh-Nilsson, A., Borén, J., & Chakravarti, S. (2015). The small leucine-rich repeat proteoglycans in tissue repair and atherosclerosis. *Journal of Internal Medicine*, *278*(5), 447–461. <https://doi.org/10.1111/joim.12400>
- Huysseune, A., & Sire, J.-Y. (1998). Evolution of patterns and processes in teeth and tooth-related tissues in non-mammalian vertebrates. *European journal of oral sciences*, *106*(S1), 437–481.
- Huysseune, A., Takle, H., Soenens, M., Taerwe, K., & Witten, P. E. (2008). Unique and shared gene expression patterns in atlantic salmon (*salmo salar*) tooth development. *Development genes and evolution*, *218*(8), 427–437.
- Iozzo, R. V., & Schaefer, L. (2015). Proteoglycan form and function: A comprehensive nomenclature of proteoglycans. *Matrix Biology*, *42*, 11–55. <https://doi.org/10.1016/j.matbio.2015.02.003>
- Irie, N., & Kuratani, S. (2011). Comparative transcriptome analysis reveals vertebrate phylogenetic period during organogenesis. *Nature communications*, *2*(1), 1–8.
- Jaillon, O., Aury, J.-M., Brunet, F., Petit, J.-L., Stange-Thomann, N., Mauceli, E., Bouneau, L., Fischer, C., Ozouf-Costaz, C., Bernot, A., Nicaud, S., Jaffe, D., Fisher, S., Lutfalla, G., Dossat, C., Segurens, B., Dasilva, C., Salanoubat, M., Levy, M., . . . Crollius, H. R. (2004). Genome duplication in the teleost fish *tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature*, *431*(7011), 946–957. <https://doi.org/10.1038/nature03025>
- Janvier, P. (1990). La structure de l'exosquelette des galeaspida (vertebrata). *Comptes rendus de l'Académie des sciences. Série 2, Mécanique, Physique, Chimie, Sciences de l'univers, Sciences de la Terre*, *310*(5), 655–659.
- Janvier, P. (1996). *Early vertebrates*.
- Janvier, P., & Arsénault, M. (2002). Calcification of early vertebrate cartilage. *Nature*, *417*(6889), 609–609.

- Janvier, P., Arsenault, M., & Desbiens, S. (2004). Calcified cartilage in the paired fins of the osteostracan *escuminaspis laticeps* (traquair 1880), from the late devonian of miguasha (québec, canada), with a consideration of the early evolution of the pectoral fin endoskeleton in vertebrates. *Journal of Vertebrate Paleontology*, *24*(4), 773–779.
- Jones, J. H. .-, Robertson, D. L., & Boot-Handford, R. P. (2007). On the origins of the extracellular matrix in vertebrates. *Matrix Biology*, *26*(1), 2–11. <https://doi.org/10.1016/j.matbio.2006.09.008>
- Kaback, L. A., Soung, D. Y., Naik, A., Smith, N., Schwarz, E. M., O’Keefe, R. J., & Drissi, H. (2008). Osterix/sp7 regulates mesenchymal stem cell mediated endochondral ossification. *Journal of cellular physiology*, *214*(1), 173–182.
- Kague, E., Witten, P. E., Soenens, M., Campos, C. L., Lubiana, T., Fisher, S., Hammond, C., Brown, K. R., Passos-Bueno, M., & Huyseune, A. (2018). Zebrafish sp7 mutants show tooth cycling independent of attachment, eruption and poor differentiation of teeth. *Developmental biology*, *435*(2), 176–184.
- Kalamajski, S., Aspberg, A., Lindblom, K., Heinegård, D., & Oldberg, Å. (2009). Asporin competes with decorin for collagen binding, binds calcium and promotes osteoblast collagen mineralization. *Biochemical Journal*, *423*(1), 53–59. <https://doi.org/10.1042/bj20090542>
- Kalamajski, S., & Oldberg, Å. (2010). The role of small leucine-rich proteoglycans in collagen fibrillogenesis. *Matrix Biology*, *29*(4), 248–253. <https://doi.org/10.1016/j.matbio.2010.01.001>
- Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., & Jermini, L. S. (2017). Modelfinder: Fast model selection for accurate phylogenetic estimates. *Nature methods*, *14*(6), 587–589.
- Kang, K., Meng, Q., Shats, I., Umbach, D. M., Li, M., Li, Y., Li, X., & Li, L. (2019). Cdseq: A novel complete deconvolution method for dissecting heterogeneous samples using gene expression data. *PLoS computational biology*, *15*(12), e1007510.
- Karamanos, N. K., Piperigkou, Z., Theocharis, A. D., Watanabe, H., Franchi, M., Baud, S., Brézillon, S., Götte, M., Passi, A., Vigetti, D., Ricard-Blum, S., Sanderson, R. D., Neill, T., & Iozzo, R. V. (2018). Proteoglycan chemical diversity drives multifunctional cell regulation and therapeutics. *Chemical Reviews*, *118*(18), 9152–9232. <https://doi.org/10.1021/acs.chemrev.8b00354>
- Kardong, K. V. (1997). *Vertebrates: Comparative anatomy, function, evolution*. Heinle; Heinle Publishers.
- Karlsson, C., Dehne, T., Lindahl, A., Brittberg, M., Pruss, A., Sittinger, M., & Ringe, J. (2010). Genome-wide expression profiling reveals new candidate genes associated with osteoarthritis. *Osteoarthritis and cartilage*, *18*(4), 581–592.
- Karsenty, G. (2003). The complexities of skeletal biology. *Nature*, *423*(6937), 316–318.

- Kasahara, M., Naruse, K., Sasaki, S., Nakatani, Y., Qu, W., Ahsan, B., Yamada, T., Nagayasu, Y., Doi, K., Kasai, Y., Jindo, T., Kobayashi, D., Shimada, A., Toyoda, A., Kuroki, Y., Fujiyama, A., Sasaki, T., Shimizu, A., Asakawa, S., . . . Kohara, Y. (2007). The medaka draft genome and insights into vertebrate genome evolution. *Nature*, *447*(7145), 714–719. <https://doi.org/10.1038/nature05846>
- Katoh, K., & Standley, D. M. (2013). Mafft multiple sequence alignment software version 7: Improvements in performance and usability. *Molecular biology and evolution*, *30*(4), 772–780.
- Kawasaki, K., & Weiss, K. (2008). SCPP gene evolution and the dental mineralization continuum. *Journal of Dental Research*, *87*(6), 520–531. <https://doi.org/10.1177/154405910808700608>
- Kawasaki, K. (2009). The scpp gene repertoire in bony vertebrates and graded differences in mineralized tissues. *Development genes and evolution*, *219*(3), 147–157.
- Kawasaki, K., & Amemiya, C. T. (2014). Scpp genes in the coelacanth: Tissue mineralization genes shared by sarcopterygians. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *322*(6), 390–402.
- Kawasaki, K., Buchanan, A. V., & Weiss, K. M. (2007). Gene duplication and the evolution of vertebrate skeletal mineralization. *Cells Tissues Organs*, *186*(1), 7–24.
- Kawasaki, K., Mikami, M., Nakatomi, M., Braasch, I., Batzel, P., H. Postlethwait, J., Sato, A., Sasagawa, I., & Ishiyama, M. (2017). Scpp genes and their relatives in gar: Rapid expansion of mineralization genes in osteichthyans. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *328*(7), 645–665.
- Kawasaki, K., Suzuki, T., & Weiss, K. M. (2005). Phenogenetic drift in evolution: The changing genetic basis of vertebrate teeth. *Proceedings of the National Academy of Sciences*, *102*(50), 18063–18068.
- Kemp, N. E., & Westrin, S. K. (1979). Ultrastructure of calcified cartilage in the endoskeletal tesserae of sharks. *Journal of Morphology*, *160*(1), 75–101.
- Kessels, M. Y., Huitema, L. F., Boeren, S., Kranenbarg, S., Schulte-Merker, S., van Leeuwen, J. L., & de Vries, S. C. (2014). Proteomics analysis of the zebrafish skeletal extracellular matrix. *PLoS one*, *9*(3), e90568.
- King, B., Qiao, T., Lee, M. S., Zhu, M., & Long, J. A. (2017). Bayesian morphological clock methods resurrect placoderm monophyly and reveal rapid early evolution in jawed vertebrates. *Systematic Biology*, *66*(4), 499–516.
- Kirsch, T., & Wuthier, R. E. (1994). Stimulation of calcification of growth plate cartilage matrix vesicles by binding to type ii and x collagens. *Journal of Biological Chemistry*, *269*(15), 11462–11469.
- Kolliker, A. V. (1859). II. on the different types in the microscopic structure of the skeleton of osseous fishes. *Proceedings of the Royal Society of London*, (9), 656–668.
- Kovalevskij, A. O. (1866). *Entwicklungsgeschichte der einfachen ascidien* (Vol. 10). Kaiserliche Akad. der Wiss.

- Kumar, M., Gouw, M., Michael, S., Sámano-Sánchez, H., Pancsa, R., Glavina, J., Diakogianni, A., Valverde, J. A., Bukirova, D., Čalyševa, J., Palopoli, N., Davey, N. E., Chemes, L. B., & Gibson, T. J. (2019). ELM—the eukaryotic linear motif resource in 2020. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkz1030>
- Kumar, S., Stecher, G., Suleski, M., & Hedges, S. B. (2017). Timetree: A resource for timelines, timetrees, and divergence times. *Molecular biology and evolution*, *34*(7), 1812–1819.
- Kuraku, S., Meyer, A., & Kuratani, S. (2008). Timing of genome duplications relative to the origin of the vertebrates: Did cyclostomes diverge before or after? *Molecular Biology and Evolution*, *26*(1), 47–59. <https://doi.org/10.1093/molbev/msn222>
- Lamarck, J. B. (1794). *Recherches sur les causes des principaux faits physiques. Paris: Tome Second.*
- Laue, K., Jänicke, M., Plaster, N., Sonntag, C., & Hammerschmidt, M. (2008). Restriction of retinoic acid activity by *cyp26b1* is required for proper timing and patterning of osteogenesis during zebrafish development.
- Leprevost, A., & Sire, J.-Y. (2014). Architecture, mineralization and development of the axial skeleton in acipenseriformes, and occurrences of axial anomalies in rearing conditions; can current knowledge in teleost fish help? *Journal of Applied Ichthyology*, *30*(4), 767–776.
- Leurs, N., Martinand-Mari, C., Ventéo, S., Haitina, T., & Debiais-Thibaud, M. (2021). Evolution of matrix gla and bone gla protein genes in jawed vertebrates. *Frontiers in Genetics*, *12*. <https://doi.org/10.3389/fgene.2021.620659>
- Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with bwa-mem. *arXiv preprint arXiv:1303.3997*.
- Li, N., Felber, K., Elks, P., Croucher, P., & Roehl, H. H. (2009). Tracking gene expression during zebrafish osteoblast differentiation. *Developmental dynamics*, *238*(2), 459–466.
- Li, Q., Zhu, Y.-a., Lu, J., Chen, Y., Wang, J., Peng, L., Wei, G., & Zhu, M. (2021). A new silurian fish close to the common ancestor of modern gnathostomes. *Current Biology*, *31*(16), 3613–3620.
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for rna-seq data with *DESeq2*. *Genome biology*, *15*(12), 1–21.
- Low, S. W. Y., Connor, T. B., Kassem, I. S., Costakos, D. M., & Chaurasia, S. S. (2021). Small leucine-rich proteoglycans (SLRPs) in the retina. *International Journal of Molecular Sciences*, *22*(14), 7293. <https://doi.org/10.3390/ijms22147293>
- Lowenstam, H. A., Weiner, S. et al. (1989). *On biomineralization*. Oxford University Press on Demand.
- Luo, G., Ducky, P., McKee, M. D., Pinero, G. J., Loyer, E., Behringer, R. R., & Karsenty, G. (1997). Spontaneous calcification of arteries and cartilage in mice lacking matrix gla protein. *Nature*, *386*(6620), 78–81.

- Mackie, E., Ahmed, Y., Tatarczuch, L., Chen, K.-S., & Mirams, M. (2008). Endochondral ossification: How cartilage is converted into bone in the developing skeleton. *The international journal of biochemistry & cell biology*, *40*(1), 46–62.
- Maes, C., Carmeliet, P., Moermans, K., Stockmans, I., Smets, N., Collen, D., Bouillon, R., & Carmeliet, G. (2002). Impaired angiogenesis and endochondral bone formation in mice lacking the vascular endothelial growth factor isoforms vegf164 and vegf188. *Mechanisms of development*, *111*(1-2), 61–73.
- Maes, C., Kobayashi, T., Selig, M. K., Torrekens, S., Roth, S. I., Mackem, S., Carmeliet, G., & Kronenberg, H. M. (2010). Osteoblast precursors, but not mature osteoblasts, move into developing and fractured bones along with invading blood vessels. *Developmental cell*, *19*(2), 329–344.
- Marconi, A., Hancock-Ronemus, A., & Gillis, J. A. (2020). Adult chondrogenesis and spontaneous cartilage repair in the skate, leucoraja erinacea. *eLife*, *9*. <https://doi.org/10.7554/elife.53414>
- Margres, M. J., Bigelow, A. T., Lemmon, E. M., Lemmon, A. R., & Rokyta, D. R. (2017). Selection to increase expression, not sequence diversity, precedes gene family origin and expansion in rattlesnake venom. *Genetics*, *206*(3), 1569–1580.
- Matsushima, N., Miyashita, H., & Kretsinger, R. H. (2021). Sequence features, structure, ligand interaction, and diseases in small leucine rich repeat proteoglycans. *Journal of Cell Communication and Signaling*, *15*(4), 519–531. <https://doi.org/10.1007/s12079-021-00616-4>
- Matsuura, T., Duarte, W. R., Cheng, H., Uzawa, K., & Yamauchi, M. (2001). Differential expression of decorin and biglycan genes during mouse tooth development. *Matrix Biology*, *20*(5-6), 367–373. [https://doi.org/10.1016/s0945-053x\(01\)00142-1](https://doi.org/10.1016/s0945-053x(01)00142-1)
- Mayeur, H., Lanoizelet, M., Quillien, A., Menuet, A., Michel, L., Martin, K. J., Dejean, S., Blader, P., Mazan, S., & Lagadec, R. (2021). When bigger is better: 3d rna profiling of the developing head in the catshark scyliorhinus canicula. *Frontiers in cell and developmental biology*, 2944.
- McGinnis, W., Levine, M. S., Hafen, E., Kuroiwa, A., & Gehring, W. J. (1984). A conserved dna sequence in homoeotic genes of the drosophila antennapedia and bithorax complexes. *Nature*, *308*(5958), 428–433.
- McKinnon, P. J., McLaughlin, S. K., Kapsetaki, M., & Margolskee, R. F. (2000). Extracellular matrix-associated protein sc1 is not essential for mouse development. *Molecular and Cellular Biology*, *20*(2), 656–660.
- Melrose, J. (2019). Functional consequences of keratan sulfate sulfation in electrosensory tissues and in neuronal regulation. *Advanced Biosystems*, *3*(4), 1800327. <https://doi.org/10.1002/adbi.201800327>

- Miao, D., & Scutt, A. (2002). Histochemical localization of alkaline phosphatase activity in decalcified bone and cartilage. *Journal of Histochemistry & Cytochemistry*, *50*(3), 333–340.
- Miles, R. S., & Westoll, T. S. (1968). Ix.—the placoderm fish *Coccosteus cuspidatus* miller ex agassiz from the middle old red sandstone of scotland. part i. descriptive morphology. *Earth and Environmental Science Transactions of the Royal Society of Edinburgh*, *67*(9), 373–476.
- Min, Z., & Janvier, P. (1998). The histological structure of the endoskeleton in galeaspids (galeaspida, vertebrata). *Journal of Vertebrate Paleontology*, *18*(3), 650–654.
- Miyashita, T., Coates, M. I., Farrar, R., Larson, P., Manning, P. L., Wogelius, R. A., Edwards, N. P., Anné, J., Bergmann, U., Palmer, A. R., et al. (2019). Hagfish from the cretaceous tethys sea and a reconciliation of the morphological–molecular conflict in early vertebrate phylogeny. *Proceedings of the National Academy of Sciences*, *116*(6), 2146–2151.
- Miyashita, T., Gess, R. W., Tietjen, K., & Coates, M. I. (2021). Non-ammocoete larvae of palaeozoic stem lampreys. *Nature*, *591*(7850), 408–412.
- Mochida, Y., Parisuthiman, D., Kaku, M., Hanai, J.-i., Sukhatme, V. P., & Yamauchi, M. (2006). Nephrocan, a novel member of the small leucine-rich repeat protein family, is an inhibitor of transforming growth factor- signaling. *Journal of Biological Chemistry*, *281*(47), 36044–36051. <https://doi.org/10.1074/jbc.m604787200>
- Monigatti, F., Gasteiger, E., Bairoch, A., & Jung, E. (2002). The sulfinator: Predicting tyrosine sulfation sites in protein sequences. *Bioinformatics*, *18*(5), 769–770. <https://doi.org/10.1093/bioinformatics/18.5.769>
- Morse, P. E., Stock, M. K., James, K. C., Natanson, L. J., & Stock, S. R. (2022). Shark centra microanatomy and mineral density variation studied with laboratory microcomputed tomography. *Journal of Structural Biology*, *214*(1), 107831.
- Mosig, R. A., Dowling, O., DiFeo, A., Ramirez, M. C. M., Parker, I. C., Abe, E., Diouri, J., Aqeel, A. A., Wylie, J. D., Oblander, S. A., et al. (2007). Loss of mmp-2 disrupts skeletal and craniofacial development and results in decreased bone mineralization, joint erosion and defects in osteoblast and osteoclast growth. *Human molecular genetics*, *16*(9), 1113–1123.
- Moss, M. L. (1977). Skeletal tissues in sharks. *American Zoologist*, *17*(2), 335–342.
- Müller, G. B. (2017). Why an extended evolutionary synthesis is necessary. *Interface focus*, *7*(5), 20170015.
- Murray, S. S., Wang, J. C., Duarte, M. E. L., Zhao, K.-W., Tian, H., Francis, T., & Brochmann Murray, E. J. (2015). The bone matrix protein secreted phosphoprotein 24 kd (spp24): Bone metabolism regulator and starting material for biotherapeutic materials.
- Murshed, M. (2018). Mechanism of bone mineralization. *Cold Spring Harbor perspectives in medicine*, *8*(12), a031229.

- Nagase, H., VISSE, R., & MURPHY, G. (2006). Structure and function of matrix metalloproteinases and TIMPs. *Cardiovascular Research*, *69*(3), 562–573. <https://doi.org/10.1016/j.cardiores.2005.12.002>
- Neugebauer, B. M., Moore, M. A., Broess, M., Gerstenfeld, L., & Hauschka, P. (1995). Characterization of structural sequences in the chicken osteocalcin gene: Expression of osteocalcin by maturing osteoblasts and by hypertrophic chondrocytes in vitro. *Journal of Bone and Mineral Research*, *10*(1), 157–163.
- Nguyen, L.-T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2014). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. <https://doi.org/10.1093/molbev/msu300>
- Nikitovic, D., Aggelidakis, J., Young, M. F., Iozzo, R. V., Karamanos, N. K., & Tzanakakis, G. N. (2012). The biology of small leucine-rich proteoglycans in bone pathophysiology. *Journal of Biological Chemistry*, *287*(41), 33926–33933. <https://doi.org/10.1074/jbc.r112.379602>
- Ninomiya, K., Miyamoto, T., Imai, J.-i., Fujita, N., Suzuki, T., Iwasaki, R., Yagi, M., Watanabe, S., Toyama, Y., & Suda, T. (2007). Osteoclastic activity induces osteomodulin expression in osteoblasts. *Biochemical and Biophysical Research Communications*, *362*(2), 460–466. <https://doi.org/10.1016/j.bbrc.2007.07.193>
- Ohno, S. (1970). The enormous diversity in genome sizes of fish as a reflection of nature's extensive experiments with gene duplication. *Transactions of the American Fisheries Society*, *99*(1), 120–130. [https://doi.org/10.1577/1548-8659\(1970\)99<120:tedigs>2.0.co;2](https://doi.org/10.1577/1548-8659(1970)99<120:tedigs>2.0.co;2)
- Ohno, S. (1999). Gene duplication and the uniqueness of vertebrate genomes circa 1970–1999. *Seminars in cell & developmental biology*, *10*(5), 517–522.
- Omelon, S., Georgiou, J., Henneman, Z. J., Wise, L. M., Sukhu, B., Hunt, T., Wynnyckyj, C., Holmyard, D., Bielecki, R., & Grynypas, M. D. (2009). Control of vertebrate skeletal mineralization by polyphosphates. *PLoS One*, *4*(5), e5634.
- Omelon, S., Georgiou, J., Variola, F., & Dean, M. N. (2014). Colocation and role of polyphosphates and alkaline phosphatase in apatite biomineralization of elasmobranch tesserae. *Acta biomaterialia*, *10*(9), 3899–3910.
- Ortiz-Delgado, J., Simes, D., Gavaia, P., Sarasquete, C., & Cancela, M. (2005). Osteocalcin and matrix gla protein in developing teleost teeth: Identification of sites of mrna and protein accumulation at single cell resolution. *Histochemistry and cell biology*, *124*(2), 123–130.
- Ørvig, T. (1951). *Histologic studies of placoderms and fossil elasmobranchs*.
- Ørvig, T. (1967). Phylogeny of tooth tissues: Evolution of some calcified tissues: Evolution of some calcified tissues in early vertebrates. *Structural and chemical organization of teeth*, 45–110.

- O'Shea, J., Keating, J. N., & Donoghue, P. C. (2019). The dermal skeleton of the jawless vertebrate *tremataspis mammillata* (osteostraci, stem-gnathostomata). *Journal of morphology*, *280*(7), 999–1025.
- Ota, K. G., Fujimoto, S., Oisi, Y., & Kuratani, S. (2013). Late development of hagfish vertebral elements. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *320*(3), 129–139. <https://doi.org/10.1002/jez.b.22489>
- Ota, K. G., Fujimoto, S., Oisi, Y., & Kuratani, S. (2011). Identification of vertebra-like elements and their possible differentiation from sclerotomes in the hagfish. *Nature communications*, *2*(1), 1–6.
- Ota, K. G., Oisi, Y., Fujimoto, S., & Kuratani, S. (2014). The origin of developmental mechanisms underlying vertebral elements: Implications from hagfish evo-devo. *Zoology*, *117*(1), 77–80.
- Paiva, K. B. S., & Granjeiro, J. M. (2014). Bone tissue remodeling and development: Focus on matrix metalloproteinase functions. *Archives of Biochemistry and Biophysics*, *561*, 74–87. <https://doi.org/10.1016/j.abb.2014.07.034>
- Palasca, O., Santos, A., Stolte, C., Gorodkin, J., & Jensen, L. J. (2018). Tissues 2.0: An integrative web resource on mammalian tissue expression. *Database*, *2018*.
- Pan, Y., Fan, Y., Lu, Y., Peng, S., Lin, H., & Deng, Q. (2022). Molecular characterization of matrix metalloproteinase gene family across primates. *Aging*, *14*(8), 3425–3445. <https://doi.org/10.18632/aging.204021>
- Panganiban, G., Irvine, S. M., Lowe, C., Roehl, H., Corley, L. S., Sherbon, B., Grenier, J. K., Fallon, J. F., Kimble, J., Walker, M., et al. (1997). The origin and evolution of animal appendages. *Proceedings of the National Academy of Sciences*, *94*(10), 5162–5166.
- Pantalacci, S., & Sémon, M. (2015). Transcriptomics of developing embryos and organs: A raising tool for evo-devo. *Journal of Experimental Zoology Part B: Molecular and Developmental Evolution*, *324*(4), 363–371.
- Park, H., Huxley-Jones, J., Boot-Handford, R. P., Bishop, P. N., Attwood, T. K., & Bella, J. (2008). LRRCE: A leucine-rich repeat cysteine capping motif unique to the chordate lineage. *BMC Genomics*, *9*(1). <https://doi.org/10.1186/1471-2164-9-599>
- Pasteris, J. D., Wopenka, B., & Valsami-Jones, E. (2008). Bone and tooth mineralization: Why apatite? *Elements*, *4*(2), 97–104.
- Peignoux-Deville, J., Lallier, F., & Vidal, B. (1982). Evidence for the presence of osseous tissue in dogfish vertebrae. *Cell and tissue research*, *222*(3), 605–614.
- Qin, C., Baba, O., & Butler, W. (2004). Post-translational modifications of sibling proteins and their roles in osteogenesis and dentinogenesis. *Critical Reviews in Oral Biology & Medicine*, *15*(3), 126–136.
- Qu, Q., Haitina, T., Zhu, M., & Ahlberg, P. E. (2015). New genomic and fossil data illuminate the origin of enamel. *Nature*, *526*(7571), 108–111.

- Randilini, A., Fujikawa, K., & Shibata, S. (2020). Expression, localization and synthesis of small leucine-rich proteoglycans in developing mouse molar tooth germ. *European Journal of Histochemistry*, *64*(1). <https://doi.org/10.4081/ejh.2020.3092>
- Rasch, L. J., Martin, K. J., Cooper, R. L., Metscher, B. D., Underwood, C. J., & Fraser, G. J. (2016). An ancient dental gene set governs development and continuous regeneration of teeth in sharks. *Developmental Biology*, *415*(2), 347–370. <https://doi.org/10.1016/j.ydbio.2016.01.038>
- Ren, C., Pan, R., Hou, L., Wu, H., Sun, J., Zhang, W., Tian, X., & Chen, H. (2020). Suppression of clec3a inhibits osteosarcoma cell proliferation and promotes their chemosensitivity through the akt1/mTOR/hif1 $\alpha$  signaling pathway. *Molecular Medicine Reports*, *21*(4), 1739–1748.
- Richtsmeier, J. T. (2018). A century of development. *American journal of physical anthropology*, *165*(4), 726.
- Ridewood, W., & MacBride, E. W. (1921). Viii.—on the calcification of the vertebral centra in sharks and rays. *Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character*, *210*(372-381), 311–407.
- Robinson, R. A., & Watson, M. L. (1955). Crystal-collagen relationships in bone as observed in the electron microscope. iii. crystal and collagen morphology as a function of age. *Annals of the New York Academy of Sciences*, *60*(5), 596–630.
- Romer, A. S., Parsons, T. S. et al. (1978). *Vertebrate body*. Saunders.
- Romero, A., Leurs, N., Muñoz, D., Debais-Thibaud, M., & Marcellini, S. (2021). Divergent expression of SPARC, SPARC-1, and SCPP genes during jawed vertebrate cartilage mineralization. *Frontiers in Genetics*, *12*. <https://doi.org/10.3389/fgene.2021.788346>
- Root, Z. D., Allen, C., Gould, C., Brewer, M., Jandzik, D., & Medeiros, D. M. (2022). A comprehensive analysis of fibrillar collagens in lamprey suggests a conserved role in vertebrate musculoskeletal evolution. *Frontiers in Cell and Developmental Biology*, *10*. <https://doi.org/10.3389/fcell.2022.809979>
- Root, Z. D., Jandzik, D., Allen, C., Brewer, M., Romášek, M., Square, T., & Medeiros, D. M. (2021). Lamprey lecticans link new vertebrate genes to the origin and elaboration of vertebrate tissues. *Developmental Biology*, *476*, 282–293. <https://doi.org/10.1016/j.ydbio.2021.03.020>
- Rotllant, J., Liu, D., Yan, Y.-L., Postlethwait, J. H., Westerfield, M., & Du, S.-J. (2008). Sparc (osteonectin) functions in morphogenesis of the pharyngeal skeleton and inner ear. *Matrix Biology*, *27*(6), 561–572.
- Ruzicka, L., Howe, D. G., Ramachandran, S., Toro, S., Van Slyke, C. E., Bradford, Y. M., Eagle, A., Fashena, D., Frazer, K., Kalita, P., Mani, P., Martin, R., Moxon, S. T., Paddock, H., Pich, C., Schaper, K., Shao, X., Singer, A., & Westerfield, M. (2018). The zebrafish information network: New support for non-coding genes, richer gene ontology annota-

- tions and the alliance of genome resources. *Nucleic Acids Research*, 47(D1), D867–D873. <https://doi.org/10.1093/nar/gky1090>
- Ryll, B., Sanchez, S., Haitina, T., Tafforeau, P., & Ahlberg, P. E. (2014). The genome of *Calorhynchus* and the fossil record: A new perspective on scpp gene evolution in gnathostomes. *Evolution & Development*, 16(3), 123.
- Sacerdot, C., Louis, A., Bon, C., Berthelot, C., & Crollius, H. R. (2018). Chromosome evolution at the origin of the ancestral vertebrate genome. *Genome Biology*, 19(1). <https://doi.org/10.1186/s13059-018-1559-1>
- Sans-Fons, M. G., Sole, S., Sanfeliu, C., & Planas, A. M. (2010). Matrix metalloproteinase-9 and cell division in neuroblastoma cells and bone marrow macrophages. *The American Journal of Pathology*, 177(6), 2870–2885. <https://doi.org/10.2353/ajpath.2010.090050>
- Santosh, N., Windsor, L. J., Mahmoudi, B. S., Li, B., Zhang, W., Chernoff, E. A., Rao, N., Stocum, D. L., & Song, F. (2011). Matrix metalloproteinase expression during blastema formation in regeneration-competent versus regeneration-deficient amphibian limbs. *Developmental Dynamics*, 240(5), 1127–1141.
- Schaefer, L., & Iozzo, R. V. (2008). Biological functions of the small leucine-rich proteoglycans: From genetics to signal transduction. *Journal of Biological Chemistry*, 283(31), 21305–21309. <https://doi.org/10.1074/jbc.r800020200>
- Schreiweis, M. A., Butler, J. P., Kulkarni, N. H., Knierman, M. D., Higgs, R. E., Halladay, D. L., Onyia, J. E., & Hale, J. E. (2007). A proteomic analysis of adult rat bone reveals the presence of cartilage/chondrocyte markers. *Journal of Cellular Biochemistry*, 101(2), 466–476.
- Schrider, D. R., & Hahn, M. W. (2010). Gene copy-number polymorphism in nature. *Proceedings of the Royal Society B: Biological Sciences*, 277(1698), 3213–3221.
- Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 461–464.
- Scott, J. E., & Orford, C. R. (1981). Dermatan sulphate-rich proteoglycan associates with rat tail-tendon collagen at the d band in the gap region. *Biochemical Journal*, 197(1), 213–216.
- Scott, J. E., Orford, C. R., & Hughes, E. W. (1981). Proteoglycan-collagen arrangements in developing rat tail tendon. an electron microscopical and biochemical investigation. *Biochemical Journal*, 195(3), 573–581.
- Seidel, R., Blumer, M., Chaumel, J., Amini, S., & Dean, M. N. (2020). Endoskeletal mineralization in chimaera and a comparative guide to tessellated cartilage in chondrichthyan fishes (sharks, rays and chimaera). *Journal of the Royal Society Interface*, 17(171), 20200474.
- Seidel, R., Blumer, M., Pechriggl, E.-J., Lyons, K., Hall, B. K., Fratzl, P., Weaver, J. C., & Dean, M. N. (2017). Calcified cartilage or bone? collagens in the tessellated endoskeletons of cartilaginous fish (sharks and rays). *Journal of structural biology*, 200(1), 54–71.
- Seidel, R., Blumer, M., Zaslansky, P., Knötel, D., Huber, D. R., Weaver, J. C., Fratzl, P., Omelon, S., Bertinetti, L., & Dean, M. N. (2017). Ultrastructural, material and crys-

- tallographic description of endophytic masses—a possible damage response in shark and ray tessellated calcified cartilage. *Journal of Structural Biology*, 198(1), 5–18.
- Seidel, R., Lyons, K., Blumer, M., Zaslansky, P., Fratzl, P., Weaver, J. C., & Dean, M. N. (2016). Ultrastructural and developmental features of the tessellated endoskeleton of elasmobranchs (sharks and rays). *Journal of Anatomy*, 229(5), 681–702.
- Sémon, M., & Wolfe, K. H. (2007). Consequences of genome duplication. *Current opinion in genetics & development*, 17(6), 505–512.
- Shimeld, S. M., & Holland, P. W. (2000). Vertebrate innovations. *Proceedings of the National Academy of Sciences*, 97(9), 4449–4452.
- Shinomura, T., & Kimata, K. (1992). Proteoglycan-lb, a small dermatan sulfate proteoglycan expressed in embryonic chick epiphyseal cartilage, is structurally related to osteoinductive factor. *Journal of Biological Chemistry*, 267(2), 1265–1270. [https://doi.org/10.1016/s0021-9258\(18\)48424-4](https://doi.org/10.1016/s0021-9258(18)48424-4)
- Shipley, J. M., Wesselschmidt, R. L., Kobayashi, D. K., Ley, T. J., & Shapiro, S. D. (1996). Metalloelastase is required for macrophage-mediated proteolysis and matrix invasion in mice. *Proceedings of the National Academy of Sciences*, 93(9), 3942–3946. <https://doi.org/10.1073/pnas.93.9.3942>
- Simões-Costa, M., & Bronner, M. E. (2015). Establishing neural crest identity: A gene regulatory recipe. *Development*, 142(2), 242–257.
- Sintuu, C., Murray, S. S., Behnam, K., Simon, R., Jawien, J., Silva, J. D. P., Duarte, M. E. L., & Brochmann, E. J. (2008). Full-length bovine spp24 [spp24 (24-203)] inhibits bmp-2 induced bone formation. *Journal of orthopaedic research*, 26(6), 753–758.
- Sire, J.-Y., Donoghue, P. C., & Vickaryous, M. K. (2009). Origin and evolution of the integumentary skeleton in non-tetrapod vertebrates. *Journal of Anatomy*, 214(4), 409–440.
- Si-Tayeb, K., Monvoisin, A., Mazzocco, C., Lepreux, S., Decossas, M., Cubel, G., Taras, D., Blanc, J.-F., Robinson, D. R., & Rosenbaum, J. (2006). Matrix metalloproteinase 3 is present in the cell nucleus and is involved in apoptosis. *The American Journal of Pathology*, 169(4), 1390–1401. <https://doi.org/10.2353/ajpath.2006.060005>
- Smith, J. J., & Keinath, M. C. (2015). The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. *Genome Research*, 25(8), 1081–1090. <https://doi.org/10.1101/gr.184135.114>
- Smith, J. J., Timoshevskaya, N., Ye, C., Holt, C., Keinath, M. C., Parker, H. J., Cook, M. E., Hess, J. E., Narum, S. R., Lamanna, F., Kaessmann, H., Timoshevskiy, V. A., Waterbury, C. K. M., Saraceno, C., Wiedemann, L. M., Robb, S. M. C., Baker, C., Eichler, E. E., Hockman, D., ... Amemiya, C. T. (2018). The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nature Genetics*, 50(2), 270–277. <https://doi.org/10.1038/s41588-017-0036-1>
- Sommer, B., Bickel, M., Hofstetter, W., & Wetterwald, A. (1996). Expression of matrix proteins during the development of mineralized tissues. *Bone*, 19(4), 371–380.

- Stamenkovic, I. (2003). Extracellular matrix remodelling: The role of matrix metalloproteinases. *The Journal of Pathology*, *200*(4), 448–464. <https://doi.org/10.1002/path.1400>
- Stensiö, E. A. et al. (1927). The downtonian and devonian vertebrates of spitsbergen. i, family cephalaspidae.
- Stephens, M. (2017). False discovery rates: A new deal. *Biostatistics*, *18*(2), 275–294.
- Tarazona, O. A., Slota, L. A., Lopez, D. H., Zhang, G., & Cohn, M. J. (2016). The genetic program for cartilage development has deep homology within bilateria. *Nature*, *533*(7601), 86–89.
- Thesleff, I. (2014). Current understanding of the process of tooth formation: Transfer from the laboratory to the clinic. *Australian dental journal*, *59*, 48–54.
- Tillgren, V., Ho, J. C., Önerfjord, P., & Kalamajski, S. (2015). The novel small leucine-rich protein chondroadherin-like (CHADL) is expressed in cartilage and modulates chondrocyte differentiation. *Journal of Biological Chemistry*, *290*(2), 918–925. <https://doi.org/10.1074/jbc.m114.593541>
- Trott, J., & Lucow, R. (1964). A macroscopic and microscopic investigation of the teeth of two cyclostomes—petromyzon marinus and polistotema stoutii. *Cells Tissues Organs*, *56*(3), 201–215.
- Turner, M. E., White, C. A., Taylor, S. M., Neville, K., Rees-Milton, K., Hopman, W. M., Adams, M. A., Anastassiades, T., & Holden, R. M. (2021). Secreted phosphoprotein 24 is a biomarker of mineral metabolism. *Calcified Tissue International*, *108*(3), 354–363.
- Ustriyana, P., Schulte, F., Gombedza, F., Gil-Bona, A., Paruchuri, S., Bidlack, F. B., Hardt, M., Landis, W. J., & Sahai, N. (2021). Spatial survey of non-collagenous proteins in mineralizing and non-mineralizing vertebrate tissues ex vivo. *Bone reports*, *14*, 100754.
- Väkevä, L., Mackie, E., Kantomaa, T., & Thesleff, I. (1990). Comparison of the distribution patterns of tenascin and alkaline phosphatase in developing teeth, cartilage, and bone of rats and mice. *The Anatomical Record*, *228*(1), 69–76.
- Venkatesh, B., Lee, A. P., Ravi, V., Maurya, A. K., Korzh, V., Lim, Z. W., Ingham, P. W., Boehm, T., Brenner, S., & Warren, W. C. (2014). On the origin of scpp genes. *Evolution & Development*, *16*(3), 125–126.
- Venkatesh, B., Lee, A. P., Ravi, V., Maurya, A. K., Lian, M. M., Swann, J. B., Ohta, Y., Flajnik, M. F., Sutoh, Y., Kasahara, M., et al. (2014). Elephant shark genome provides unique insights into gnathostome evolution. *Nature*, *505*(7482), 174–179.
- Vickaryous, M. K., Gilbert, E. A., & Delorme, S. L. (2016). Fortified frogs and skeletogenesis in the skin: Osteoderm development and structure across tetrapods. *The FASEB Journal*, *30*, 15–1.
- Vickaryous, M. K., & Hall, B. K. (2008). Development of the dermal skeleton in alligator mississippiensis (archosauria, crocodylia) with comments on the homology of osteoderms. *Journal of morphology*, *269*(4), 398–422.

- Viegas, C. S., Simes, D. C., Williamson, M. K., Cavaco, S., Laizé, V., Price, P. A., & Cancela, M. L. (2013). Sturgeon osteocalcin shares structural features with matrix gla protein: Evolutionary relationship and functional implications. *Journal of Biological Chemistry*, *288*(39), 27801–27811.
- Vinarsky, V., Atkinson, D. L., Stevenson, T. J., Keating, M. T., & Odelberg, S. J. (2005). Normal newt limb regeneration requires matrix metalloproteinase function. *Developmental biology*, *279*(1), 86–98.
- Walker, W. F., Liem, K. F. et al. (1994). *Functional anatomy of the vertebratesan evolutionary perspective*.
- Wang, X., Hao, J., Xie, Y., Sun, Y., Hernandez, B., Yamoah, A. K., Prasad, M., Zhu, Q., Feng, J. Q., & Qin, C. (2010). Expression of fam20c in the osteogenesis and odontogenesis of mouse. *Journal of Histochemistry & Cytochemistry*, *58*(11), 957–967.
- Wazer, J. R. V., & Holst, K. A. (1950). Structure and properties of the condensed phosphates. i. some general considerations about phosphoric acids<sup>1</sup>. *Journal of the American Chemical Society*, *72*(2), 639–644.
- Weigle, J., & Franz-Odenaal, T. A. (2016). Functional bone histology of zebrafish reveals two types of endochondral ossification, different types of osteoblast clusters and a new bone type. *Journal of anatomy*, *229*(1), 92–103.
- Weigle, J., Franz-Odenaal, T. A., & Hilbig, R. (2015). Expression of sparc and the osteopontin-like protein during skeletal development in the cichlid fish oreochromis mossambicus. *Developmental Dynamics*, *244*(8), 955–972.
- Weiner, S., & Addadi, L. (1997). Design strategies in mineralized biological materials. *Journal of Materials Chemistry*, *7*(5), 689–702.
- Weiner, S., & Traub, W. (1992). Bone structure: From ångstroms to microns. *The FASEB journal*, *6*(3), 879–885.
- Weiner, S. (2006). Transient precursor strategy in mineral formation of bone.
- Weinreb, M., Shinar, D., & Rodan, G. A. (1990). Different pattern of alkaline phosphatase, osteopontin, and osteocalcin expression in developing rat bone visualized by in situ hybridization. *Journal of Bone and Mineral Research*, *5*(8), 831–842.
- Wen, L., Chen, J., Duan, L., & Li, S. (2018). Vitamin k-dependent proteins involved in bone and cardiovascular health. *Molecular medicine reports*, *18*(1), 3–15.
- Wendel, M., Sommarin, Y., & Heinegård, D. (1998). Bone matrix proteins: Isolation and characterization of a novel cell-binding keratan sulfate proteoglycan (osteoaderin) from bovine bone. *Journal of Cell Biology*, *141*(3), 839–847. <https://doi.org/10.1083/jcb.141.3.839>
- Wilda, M., Bächner, D., Just, W., Geerkens, C., Kraus, P., Vogel, W., & Hameister, H. (2000). A comparison of the expression pattern of five genes of the family of small leucine-rich proteoglycans during mouse development. *Journal of Bone and Mineral Research*, *15*(11), 2187–2196.

- Wiweger, M. I., Zhao, Z., van Merkesteyn, R. J., Roehl, H. H., & Hogendoorn, P. C. (2012). Hspg-deficient zebrafish uncovers dental aspect of multiple osteochondromas. *PLoS One*, 7(1), e29734.
- Wright, G. M., Keeley, F. W., & Robson, P. (2001). The unusual cartilaginous tissues of jawless craniates, cephalochordates and invertebrates. *Cell and tissue research*, 304(2), 165–174.
- Yáñez, M., Gil-Longo, J., & Campos-Toimil, M. (2012). Calcium binding proteins. *Calcium signaling*, 461–482.
- Yoshida, C. A., Furuichi, T., Fujita, T., Fukuyama, R., Kanatani, N., Kobayashi, S., Satake, M., Takada, K., & Komori, T. (2002). Core-binding factor  $\beta$  interacts with runx2 and is required for skeletal development. *Nature genetics*, 32(4), 633–638.
- Zelzer, E., McLean, W., Ng, Y.-S., Fukai, N., Reginato, A. M., Lovejoy, S., D'Amore, P. A., & Olsen, B. R. (2002). Skeletal defects in vegf120/120 mice reveal multiple roles for vegf in skeletogenesis.
- Zhang, X., Xia, K., Lin, L., Zhang, F., Yu, Y., Ange, K. S., Han, X., Edsinger, E., Sohn, J., & Linhardt, R. J. (2018). Structural and functional components of the skate sensory organ ampullae of lorenzini. *ACS Chemical Biology*, 13(6), 1677–1685. <https://doi.org/10.1021/acscchembio.8b00335>
- Zhang, Z. (2015). Chondrons and the pericellular matrix of chondrocytes. *Tissue Engineering Part B: Reviews*, 21(3), 267–277.
- Zhao, Q., Eberspaecher, H., Lefebvre, V., & de Crombrughe, B. (1997). Parallel expression of sox9 and col2a1 in cells undergoing chondrogenesis. *Developmental dynamics: an official publication of the American Association of Anatomists*, 209(4), 377–386.
- Zhou, Z., Apte, S. S., Soininen, R., Cao, R., Baaklini, G. Y., Rauser, R. W., Wang, J., Cao, Y., & Tryggvason, K. (2000). Impaired endochondral ossification and angiogenesis in mice deficient in membrane-type matrix metalloproteinase i. *Proceedings of the National Academy of Sciences*, 97(8), 4052–4057.
- Zhu, G., Zhang, T., Chen, M., Yao, K., Huang, X., Zhang, B., Li, Y., Liu, J., Wang, Y., & Zhao, Z. (2021). Bone physiological microenvironment and healing mechanism: Basis for future bone-tissue engineering scaffolds. *Bioactive materials*, 6(11), 4110–4140.
- Zhu, Y.-a., Giles, S., Young, G. C., Hu, Y., Bazzi, M., Ahlberg, P. E., Zhu, M., & Lu, J. (2021). Endocast and bony labyrinth of a devonian “placoderm” challenges stem gnathostome phylogeny. *Current Biology*, 31(5), 1112–1118.

## Abstract

Bony fish display a skeleton that develops mainly as cartilage, locally replaced by endochondral bone tissue. Cartilaginous fish, however, have a cartilaginous skeleton, locally mineralized in various ways, some of which are evolutionary innovations specific to this lineage (prismatic cartilage, for instance). The cells associated with these tissues are thus capable of biomineralization by precipitation of calcium phosphate, a property that seems to be limited to the vertebrate clade. In this thesis, I focus on the genetics of skeletal mineralization development in a cartilaginous fish: the small-spotted catshark (*Scyliorhinus canicula*). I describe in detail the putative genetic factors underlying the mineralization process, and I do so through two approaches: through candidate gene studies and through transcriptomic studies. This allows me to better characterize the set of genetic mineralization tools in chondrichthyans, so that through comparative analysis, an ancestral state for jawed vertebrates can be inferred, at least partially. First, through the candidate gene approach, I test the level of conservation regarding genes known to be associated with mineralization in bony fish. Second, I describe how I used recently generated genomic and transcriptomic data in the small-spotted catshark to study intrinsic genes specifically expressed in mineralizing tissues. These two approaches allow me to detect interesting genes that are then studied through phylogenetic analyses, quantification/localization of expression in tissues, as well as comparative analyses with bony fish (when possible). In the last part, I present two papers in preparation that result from the work we have done with the help of two master students. These papers stem from a candidate gene approach but involve large gene families and in an effort to include all paralogs, the papers move towards a molecular evolutionary framework rather than an Evo-Devo framework.

## Résumé

Les poissons osseux présentent un squelette se développant principalement sous la forme de cartilage, localement remplacé par du tissu osseux endochondral. Les poissons cartilagineux, eux, présentent un squelette de nature cartilagineuse, localement minéralisé selon diverses modalités dont certaines sont des innovations évolutives propres à cette lignée (le cartilage prismatique, par exemple). Les cellules associées à ces tissus sont donc capables de biominéralisation par précipitation de phosphate de calcium, une propriété qui semble limitée au groupe des vertébrés. Au cours de cette thèse, je me concentre sur la génétique du développement de la minéralisation du squelette chez un poisson cartilagineux : la petite roussette (*Scyliorhinus canicula*). Je décris en détail les facteurs génétiques putatifs qui sous-tendent le processus de minéralisation, et je le fais par deux approches : par l'étude de gènes candidats et par la transcriptomique. Cela me permet de mieux caractériser l'ensemble des outils de minéralisation chez les chondrichthyens, de sorte que par une analyse comparative, un état ancestral pour les vertébrés à mâchoire peut être déduit, au moins partiellement. Tout d'abord, à travers l'approche par gènes candidats, je teste le niveau de conservation concernant des gènes connus par leur association à la minéralisation chez les poissons osseux. Ensuite, je décris comment j'ai utilisé les données génomiques et transcriptomiques récemment générées chez la petite roussette pour étudier les gènes intrinsèques exprimés spécifiquement dans les tissus minéralisants. Ces deux approches me permettent de détecter des gènes intéressants qui sont ensuite étudiés à travers des analyses phylogénétiques, de quantification/localisation de l'expression dans les tissus, ainsi que des analyses comparatives avec les poissons osseux (lorsque cela est possible). Dans la dernière partie, je présente deux articles en préparation qui résultent du travail que nous avons réalisé avec l'aide de deux étudiants de master. Ces articles sont issus d'une approche gène candidat mais concernent des familles de gènes étendues et dans un effort d'inclure tous les paralogues, les articles s'orientent vers un cadre d'évolution moléculaire plutôt qu'un cadre Evo-Devo.