



Lamprey lecticans link new vertebrate genes to the origin and elaboration of vertebrate tissues

Zachary D. Root^a, David Jandzik^{a,b}, Cara Allen^a, Margaux Brewer^a, Marek Romášek^a, Tyler Square^{a,c}, Daniel M. Medeiros^{a,*}

^a Department of Ecology and Evolutionary Biology, University of Colorado, Boulder, CO, 80309, USA

^b Department of Zoology, Comenius University in Bratislava, Bratislava, 84215, Slovakia

^c Department of Molecular and Cellular Biology, University of California, Berkeley, CA, 94720, USA

ARTICLE INFO

Keywords:

Vertebrate
Lamprey
Evolution
Cartilage
Lectican

ABSTRACT

The evolution of vertebrates from an invertebrate chordate ancestor involved the evolution of new organs, tissues, and cell types. It was also marked by the origin and duplication of new gene families. If, and how, these morphological and genetic innovations are related is an unresolved question in vertebrate evolution. Hyaluronan is an extracellular matrix (ECM) polysaccharide important for water homeostasis and tissue structure. Vertebrates possess a novel family of hyaluronan binding proteins called Lecticans, and studies in jawed vertebrates (gnathostomes) have shown they function in many of the cells and tissues that are unique to vertebrates. This raises the possibility that the origin and/or expansion of this gene family helped drive the evolution of these vertebrate novelties. In order to better understand the evolution of the *lectican* gene family, and its role in the evolution of vertebrate morphological novelties, we investigated the phylogeny, genomic arrangement, and expression patterns of all *lecticans* in the sea lamprey (*Petromyzon marinus*), a jawless vertebrate. Though both *P. marinus* and gnathostomes each have four *lecticans*, our phylogenetic and syntenic analyses are most consistent with the independent duplication of one of more *lecticans* in the lamprey lineage. Despite the likely independent expansion of the lamprey and gnathostome *lectican* families, we find highly conserved expression of *lecticans* in vertebrate-specific and mesenchyme-derived tissues. We also find that, unlike gnathostomes, lamprey expresses its *lectican* paralogs in distinct subpopulations of head skeleton precursors, potentially reflecting an ancestral diversity of skeletal tissue types. Together, these observations suggest that the ancestral pre-duplication *lectican* had a complex expression pattern, functioned to support mesenchymal histology, and likely played a role in the evolution of vertebrate-specific cell and tissue types.

1. Introduction

The emergence of vertebrates involved the elaboration of the ancestral chordate body plan with an array of new cell types, tissues, and organs. Among these are the expanded central and peripheral nervous systems, and the complex skeletomuscular systems of the head and trunk, which includes an array of new structural and connective tissues (Gans and Northcutt, 1983; Northcutt and Gans, 1983). Interestingly, large portions of these novelties are derived from the same embryonic source, neural crest cells, which also give rise to parts of the heart, teeth, endocrine system, and vascular smooth muscle (Gans and Northcutt, 1983; Northcutt and Gans, 1983; Eames et al., 2020). The evolution of these morphological and developmental novelties coincided with major

genome-wide changes including the origin of several new gene families, at least one whole genome duplication, and the evolution of new gene regulatory networks (Simakov et al., 2020; Martinez-Morales et al., 2007; Braasch and Schartl, 2014; Ohno, 1970; Van de Peer et al., 2009; Square et al., 2020; Sauka-Spengler et al., 2007; Hockman et al., 2019; Marlétaz et al., 2018; Meulemans and Bronner-Fraser, 2004; Meulemans and Bronner-Fraser, 2005; Martik et al., 2019). The timing of these genomic events has led to speculation that they facilitated the origin and morphological diversification of vertebrates by altering early development.

While alterations in embryogenesis can lead to major changes in the body plan, the evolution of truly novel tissues and cell types also requires the evolution of new cellular functions and histological properties.

* Corresponding author.

E-mail address: daniel.medeiros@colorado.edu (D.M. Medeiros).

<https://doi.org/10.1016/j.ydbio.2021.03.020>

Received 9 December 2020; Received in revised form 19 March 2021; Accepted 22 March 2021

Available online 19 April 2021

0012-1606/© 2021 Elsevier Inc. All rights reserved.

Extracellular matrix (ECM) proteins not only provide support and structure to cells and tissues, but also mediate signal transduction and mechanotransduction (Kawashima et al., 2009a). A key component of the ECM of many vertebrate tissues is a vertebrate-specific family of proteoglycans called Lecticans. Structurally, Lecticans are complex, consisting of hyaluronan-binding X-link domains, c-type lectin domains, a chondroitin/keratan sulfate binding domain, and an immunoglobulin domain. Because of this modular structure, Lecticans are able to interface with many different types of molecules and perform a range of functions in the ECM of diverse cells and tissues (Zelensky and Gready, 2005).

Genomically, all gnathostome *lectican* paralogs are closely linked to a *hapln* gene, which also encodes an X-link domain containing protein (Spicer et al., 2003a). The proximity of *lecticans* and *haplins*, together with their high sequence similarity indicate that they may have evolved via tandem duplication of an ancestral X-link protein-encoding gene, with subsequent exon shuffling resulting in the hybrid structure of Lecticans (Wada, 2013). After assembly of the primordial *lectican* gene, two genome-wide duplications are thought to have generated the four paralogs seen in modern jawed vertebrates: *aggrecan* (*acan*), *brevican* (*bcan*), *neurocan* (*ncan*), and *versican* (*vcan*). Since these duplications, the structures of the four gnathostome Lecticans have diverged, with *acan* acquiring an additional X-link domain, *bcan* and *ncan* losing an interglobular fold sequence adjacent to the immunoglobulin domain, and all Lecticans evolving chondroitin/keratan sulfate binding domains of different sizes (Zelensky and Gready, 2005).

Subfunctionalization, specialization, and/or neofunctionalization of gnathostome *lectican* paralogs resulted in each possessing distinct expression patterns and functions, in neural, skeletal, cardiac, and connective tissues (Milev et al., 1998; Miura et al., 1999). *acan* is known primarily for its role in the cartilage ECM (Mundlos et al., 1991; Watanabe et al., 1998), but it is also involved in neural crest cell migration and synaptic complexes in the brain (Perris and Johansson, 1990; Morawski et al., 2012; Seidenbecher et al., 1998; Krueger et al., 1992; Perissinotto et al., 2000). *acan* expression has also been found in the developing notochord as well as the epicardium and mesenchyme of the heart (Zanin et al., 1999; Casini et al., 2004). *vcan* is the most widely expressed *lectican* and is transcribed in mesoderm-derived tissues and organs including the kidneys, heart, muscles, and skeleton (Zanin et al., 1999; Casini et al., 2004; Landolt et al., 1995; Kang et al., 2004; Snow et al., 2005; Henderson and Copp, 1998), and various neuroectodermal derivatives like the otic vesicle, lens primordium (Kang et al., 2004), oligodendrocytes, Schwann Cells, the perineuronal net, ectodermal placodes, and migrating neural crest cells (Milev et al., 1998; Perissinotto et al., 2000; Ohashi et al., 2002; Courel et al., 1998; Bignami et al., 1993; Szabó et al., 2019). *bcan* and *ncan* are primarily expressed in the nervous system (Milev et al., 1998; Seidenbecher et al., 1998; Frischknecht and Seidenbecher, 2012; Ogawa et al., 2001; Gary et al., 1998; Jaworski et al., 1995; Yamada et al., 1994; Rauch et al., 1992; Oohira et al., 1994; Georgadaki and Zagris, 2014; Bell et al., 2004; Staudt et al., 2015; Owens et al., 2016), though notochord and heart expression has also been reported (Georgadaki and Zagris, 2014; Sander et al., 2001; Mishima and Hoffman, 2003; Barnette et al., 2013). Of the four *lecticans*, mutation of *acan* leads to the most significant defects, including severe chondrodysplasia (Watanabe et al., 1998), while *vcan* loss-of-function causes abnormal eye and heart development (Mukhopadhyay et al., 2006; Yamamura et al., 1997; Mjaatvedt et al., 1998). The functions of *bcan* and *ncan* are less clear, however, as mice deficient in these genes show only minor defects in neuronal potentiation (Brakebusch et al., 2002; Zhou et al., 2001).

It has been proposed that the evolution of novel interactions between Lecticans, hyaluronan, and other glycoproteins played an important role in the evolution of vertebrate tissues (Wada, 2013). However, our understanding of *lectican* expression, function, and evolution is based entirely on the information from model gnathostomes. Thus, it is unclear when in the vertebrate lineage *lecticans* originated, were duplicated, and acquired their diverse functions. The only two living jawless vertebrates, the lampreys and hagfish (the cyclostomes) have been indispensable for

understanding vertebrate evolution (Green and Bronner, 2014; Nikitina et al., 2009; Oisi et al., 2013; Ota et al., 2007; Shimeld and Donoghue, 2012). These modern agnathans diverged from the lineage leading to gnathostomes around 500 million years ago (Forey and Janvier, 1993; Delarbre et al., 2002). Due to accessibility, lampreys are the best studied modern agnathans. Despite their divergent morphologies and life histories, historical and modern comparisons have shown that lamprey and gnathostomes share most key aspects of their development, pointing to evolutionarily ancient features of vertebrate development (Green and Bronner, 2014; Jandzik, 2014; York and McCauley, 2020; Square et al., 2017; Murakami et al., 2005; Yang et al., 2016; York et al., 2019).

In this study, we used genomic and transcriptomic data from the sea lamprey, *Petromyzon marinus* to gain insight into the evolutionary history of *lectican* genes. In the context of recent work (Simakov et al., 2020), our data suggest that the four lamprey *lectican* genes were generated by one or more independent gene duplications in the lamprey lineage, though other scenarios cannot be excluded (Kuraku, 2013). We also characterized the expression patterns of *lecticans* in sea lamprey embryos and larvae, and show that *lectican* expression in neural, cardiac, and skeletal tissue is highly conserved across living vertebrates. In contrast, we find that expression of *lectican* paralogs in the head skeleton is markedly different between lamprey and gnathostomes. We posit that the ancestral pre-duplication *lectican* had a complex expression pattern which was independently partitioned between paralogs in the lamprey and gnathostome lineages. We further speculate that the origin of the primordial *lectican* gene facilitated the evolution of mesenchymal morphology and migratory ability in the cells of the first vertebrate embryos.

2. Results

2.1. The sea lamprey has four *lectican* genes encoding proteins with similar domain structures

We searched the *P. marinus* germline genome (Smith et al., 2018) and identified four different genomic scaffolds containing exons with sequence similarity to gnathostome Lecticans. We also searched all publicly available lamprey transcriptome data, as well as our own database of transcriptome sequences (see Methods) for gnathostome *lectican*-like sequences, and assembled these into 4 mRNAs corresponding to proteins of 1871aa, 1757aa, 1825aa, and 1343aa respectively (see Tab. S3 for accession numbers). All identified *lectican* exons aligned to parts of the reconstructed mRNAs, indicating there are only four sea lamprey *lectican* genes. We named these genes *lecticanA* (*lecA*), *lecticanB* (*lecB*), *lecticanC* (*lecC*), and *lecticanD* (*lecD*). We then searched for conserved domains in lamprey *lectican* conceptual translation products using NCBI's (Sayers et al., 2021) Conserved Domain search tool, and by alignment with gnathostome Lecticans. We found that although all lamprey Lectican protein sequences had largely archetypal domain structures, at least one domain appeared to be missing in each (Fig. 1A). LECA and LECC did not possess an identifiable complement control protein domain, while LECB did not have an immunoglobulin-like domain, and LECD did not have EGF-like domains. We also found that no lamprey Lectican possessed the extra X-link domain seen in ACAN (Fig. 1A).

2.2. Phylogenetic analyses do not support one-to-one orthology of lamprey and gnathostome *lecticans*

To deduce relationships between lamprey and gnathostome *lecticans*, we used *lectican* protein sequences to perform maximum likelihood phylogenetic analyses, with different taxa, substitution models, and individual parameters for tests (Roshan and Livesay, 2006; Abascal et al., 2005; Miller et al., 2010; Rambaut, 2018) (Figs. S1–S4). Among gnathostome *lecticans*, we recovered all four known paralog groups and found good support for *acan*, *bcan* and *vcan* + *ncan* subfamilies. In contrast, we found that none of the lamprey Lecticans consistently group within any of the four gnathostome Lectican paralogy groups, nor the

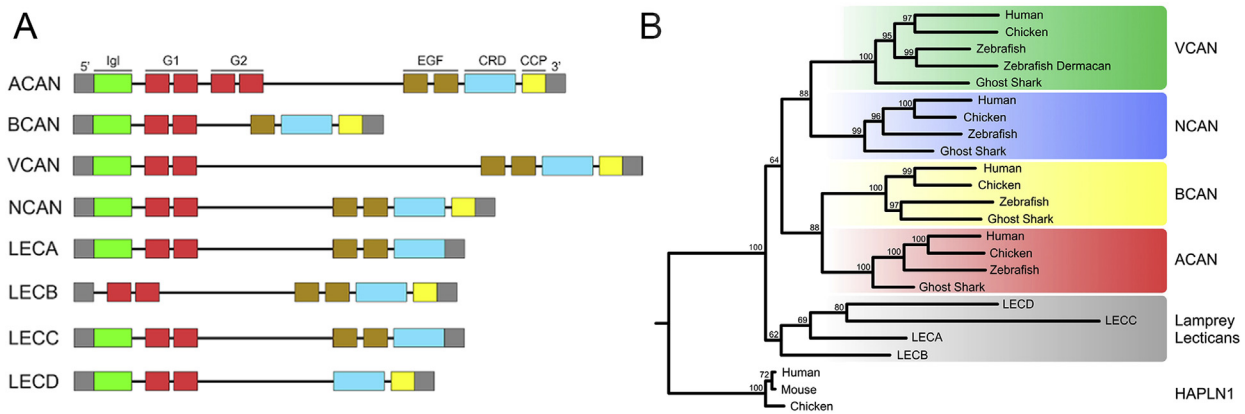


Fig. 1. Structural comparison and molecular phylogeny of vertebrate lecticans. (A) Domain structure of vertebrate Lecticans with the N-terminus to the left. **Key-words:** Igl: Immunoglobulin-like domain; G1/G2: link domains; EGF: EGF-like domains; CRD: carbohydrate recognition domains; CCP: complement control protein domain. (B) Phylogenetic relationships of vertebrate Lecticans based on amino acid sequence alignments for the entire genes. Lamprey sequences are in gray boxes while individual gnathostome paralogy groups are in colored boxes. Bootstrap value scores are shown at the respective node. HAPLN1 sequences were designated as outgroup. Original tree and accession numbers for all sequences can be found in Fig. S1 and Table. S1.

acan + *bcan* and *vcan* + *ncan* subfamilies regardless of the parameters used to build the phylogenies (Fig. 1B, Fig. S5). Considering that *lecticans* and *haplns* likely originated from a tandem duplication event early in the vertebrate lineage, we reasoned that building a phylogenetic tree using HAPLN1s and the HAPLN-aligning portion of Lectican protein sequences might help resolve the relationships between lamprey and gnathostome Lecticans (Fig. S4). As with the full-length Lectican phylogeny, none of the lamprey Lecticans grouped with any gnathostome paralogy group with high confidence.

2.3. Analyses of syntenic genes also fail to support one-to-one orthology between lamprey and gnathostome lecticans

All gnathostome *lectican* paralogs are adjacent to a corresponding HAPLN paralog (Spicer et al., 2003a). We thus searched for HAPLN-like reading frames in the lamprey genome (Smith et al., 2018), and used these to build a phylogeny of chordate HAPLN-related genes in hopes of resolving the relationships between vertebrate *lectican*/HAPLN loci. We identified one lamprey HAPLN gene linked to *lecticanD*. However, as with lamprey Lecticans, lamprey HAPLN fails to group convincingly with any single gnathostome paralogy group (Fig. S6). We also found that gnathostome HAPLN1s and HAPLN4s form a weakly supported clade, consistent with the relationships of their adjacent *lecticans*, *vcan* and *ncan*. We expanded our search to include other possible conserved linked genes. We found that all gnathostome and lamprey *lecticans* are linked to paralogs of the myocyte enhancer factor *mef2* gene family. We thus

assembled a phylogenetic tree of MEF2 amino acid sequences to see if it could provide insights into the evolution of the vertebrate *lectican* locus. As with HAPLN genes, none of the lamprey MEF2 sequences clustered convincingly with any gnathostome MEF2 paralogy group using any parameters (Fig. S7).

As a final test of orthology between lamprey and gnathostome *lecticans*, we compared the gene complement around the gnathostome and lamprey *lectican* loci. For each lamprey *lectican*, we asked if any of the surrounding 40 genes (when available) had homologs that were syntenic with any chick, spotted gar, or elephant shark *lecticans* (Fig. 2A, Fig. S8). We found that *lecA* had the most conserved linked genes, with 21/40 of adjacent genes having gnathostome homologs closely linked to one or more *lecticans*. Of those, 15 were exclusively linked to a *acan* or a *bcan*, while only 4 were exclusively linked to a *vcan* or *ncan*. Around the *lecB*, *lecC*, and *lecD* loci, 30–40% of genes were homologs of genes linked to gnathostome *lecticans*, with similar proportions linked to the *acan* + *bcan* and *vcan* + *ncan* subfamilies (Fig. S8). Thus, comparisons of linked genes provide weak support for placing *lecA* within the *acan* + *bcan* subfamily (Fig. 2B).

2.4. Expression of lecticans in sea lamprey embryos and larvae

We first detected *lecA* expression at Tahara (1988) stage 21 (st. T21) in the presumptive neural tube and axial/paraxial mesoderm (Fig. 3A). This expression continued through st. T22 and st. T23 (Fig. 3B and C), with additional expression in the heart becoming apparent at st. T23.

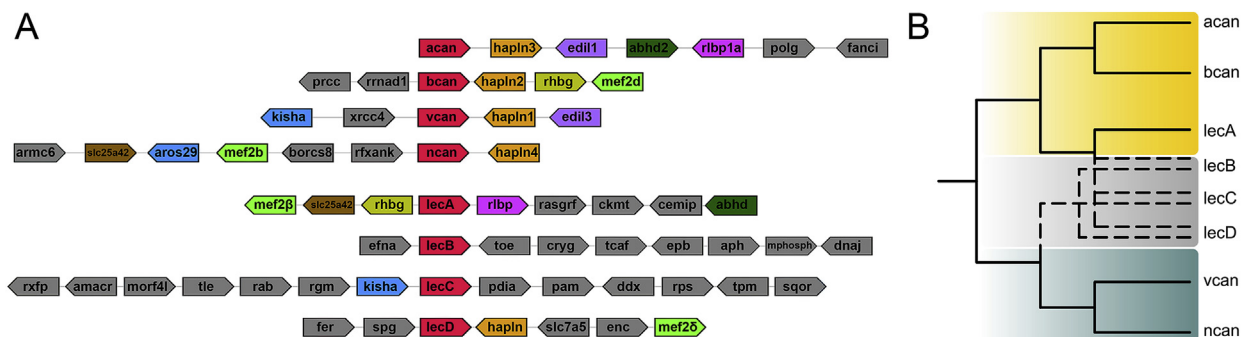


Fig. 2. Summary of *lectican* microsynteny and implications for their evolutionary relationships. (A) Conserved genes adjacent to the gnathostome (top four) and lamprey (bottom four) lectican loci. Linked genes are shown in orientation with respect to their linked *lectican* gene. *Lecticans* are in red. Homologous genes are colored the same. Gnathostome genes were surveyed within a 300 kb radius while lamprey genes were surveyed within a 400 kb radius. Macrosyntentic analyses can be found in Table S1. (B) Hypothetical maximum parsimony scenarios for the evolution of vertebrate lecticans when synteny data is incorporated. While syntentic analyses suggest that *lecticanA* is orthologous to the *aggreccan/brevican* gene family, the relationship of *lecticanB*, *lecticanC*, or *lecticanD* is unclear.

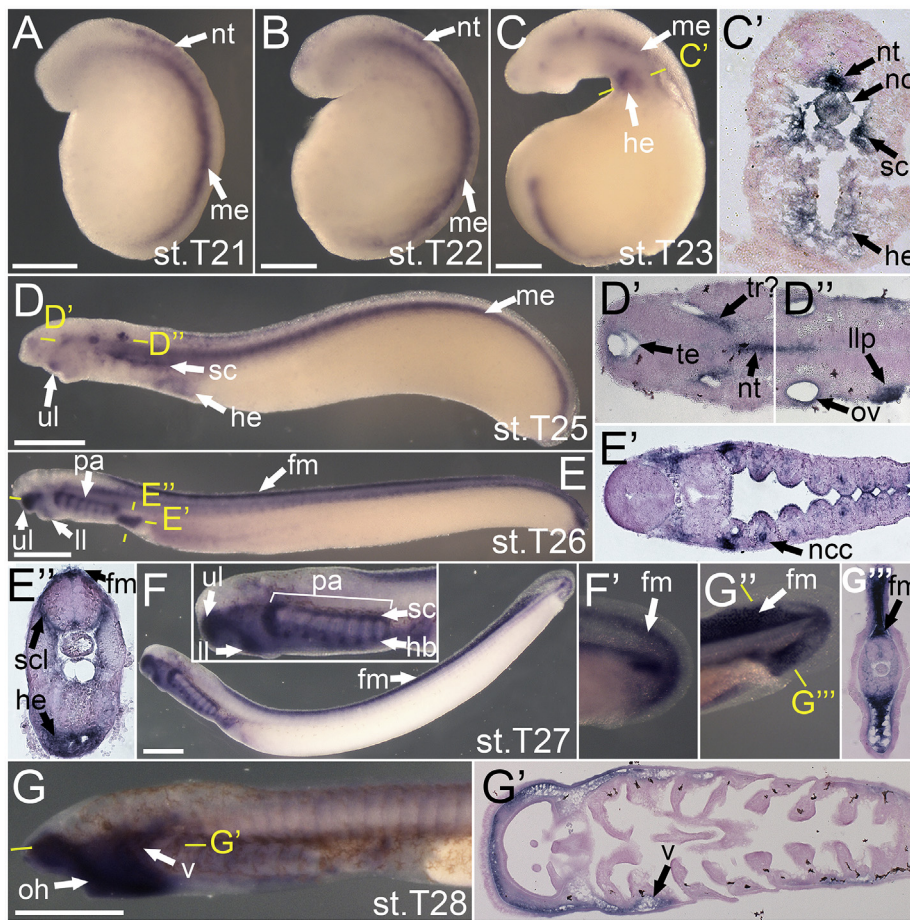


Fig. 3. Expression of *lecA* in embryos and larvae. Left lateral view in all wholemount images. Scale bars represent 500 μ m. Staging as in Tahara, 1988. Panels with the same letter depict specimens of the same stage. Yellow lines and lettering indicate approximate plane of section for the indicated panel. (A–B) At st. T21–T22 *lecA* transcripts mark cells in the dorsal neural tube and axial mesoderm. (C) At st. T23, *lecA* transcription is activated in the heart and sectioning reveals expression in the notochord, neural tube floor plate, and sclerotome. (D) At st. T25, *lecA* expression begins in the upper lip and dorsal pharyngeal mesenchyme, presaging formation of the subchordal cartilages. Expression is also seen in the heart, otic vesicle, presumptive lateral line placode, nascent telencephalon, neural tube floor plate, axial mesoderm and mesenchyme at the approximate position of the future trabecular cartilages. (E) In st. T26 larvae, *lecA* is expressed in the upper and lower lip mesenchyme, sclerotome, neural crest-derived mesenchyme in the pharyngeal arches, the heart, and fin mesenchyme. (F) In the trunk of st. T27 larvae, *lecA* transcripts mark fin mesenchyme. In the head (inset), expression is seen throughout skeletogenic mesenchyme, including the upper and lower lip mesenchyme, and pharyngeal mesenchyme as the position of the future pharyngeal arch bars, subchordal cartilages and hypobranchial bar. (G) At st. T28, *lecA* expression is seen in the mesenchyme of the upper lip, lower lip, and first pharyngeal arch as these cells differentiate into the oral hood and velum. In the trunk, *lecA* transcripts mark the differentiating mesenchyme of the dorsal and ventral fin folds. **Keywords:** fm: fin mesenchyme; hb: hypobranchial bar; he: heart; ht: heart tube; ll: lower lip; me: axial mesoderm; no: notochord; ncc: neural crest cells; nt: neural tube; oh: oral hood; ov: otic vesicle; pa: pharyngeal arches; llp: lateral line placode; sc: subchordal cartilage; scl: sclerotome; te: telencephalon; ul: upper lip; v: velum.

Sectioning at T23 revealed transcripts in the notochord, neural tube floor plate, heart tube, and sclerotome (Fig. 3C'). At st. T25, *lecA* expression was activated the upper lip, portions of the pharyngeal mesenchyme, and some cranial placodes/ganglia (Fig. 3D). At stage T26, *lecA* transcripts were found throughout the skeletogenic mesenchyme of the oropharyngeal region and fin fold, as well as in the sclerotome and heart (Fig. 3E). At st. T27, when skeletogenesis begins, we observed continued expression in the oral region, fin fold, and dorsal and ventral aspects of the branchial basket (the presumptive subchordal and hypobranchial cartilages) (Fig. 3F). However, expression appeared reduced in the anteriormost branchial arch bars (pharyngeal arch 3–5 cartilages), which chondrify first, suggesting *lecA* is downregulated in differentiating hyaline cartilage. At stage T28, expression of *lecA* in the head was almost entirely restricted to the mucocartilage of the oral hood and velum, which is derived from pharyngeal arches 1 and 2 (Fig. 3G). Fin mesenchyme expression also persisted through st. T28, expanding into the caudal fin fold (Fig. 3G).

Expression of *lecB* was first observed at stage T23 in the oral ectoderm (Fig. 4A). At st. T25, *lecB* expression expanded into the neural tube, somites, and pharynx (Fig. 4B). By pharyngeal stage (st. T26), new expression was observed in the pronephros and additional neural structures, including the pineal organ and cranial placodes/ganglia (Fig. 4C). At st. T27, we found *lecB* transcription was downregulated in neural tissues but persisted in the differentiating mucocartilage of the lips, ventral pharynx, and pharyngeal arches 1 and 2 (the forming velum) (Fig. 4D). Sectioning at this stage revealed that iterated *lecB* expression in the posterior pharyngeal arches (arches 3–8) was largely endodermal. By stage T28, *lecB* was only detectable in the mucocartilage of the lips,

velum, and ventral pharynx, and the ectoderm of the oral papillae (Fig. 4E).

lecC expression was first activated in the forming otic vesicle and notochord at st. T24 (Fig. 5A). Expression in these structures persisted through st. T25 (Fig. 5B), but was lost by st. T26.5, when transcription was activated in the intermediate domain of the posterior pharyngeal arches (Fig. 5C). Interestingly, this expression closely tracked alcian blue reactivity in the differentiating hyaline cartilage of the pharyngeal cartilage bars, as previously described (Martin et al., 2009). Consistent with this, expression expanded to the seventh and eighth arch cartilage and encompassed the full dorsoventral extent of the forming pharyngeal bars of pharyngeal arches 3–6 at st. T27 (Fig. 5D). By stage T28, *lecC* expression was seen in all hyaline cartilage bars in the posterior pharynx (Fig. 5E).

We first observed *lecD* expression at stage T21 in the developing somites (Fig. 6A). By stage T23, *lecD* was additionally detected in the heart and first pharyngeal arch, and sectioning revealed somitic expression was restricted to the sclerotome (Fig. 6B). At st. T24 and T25, we found that *lecD* expression appeared to track somite age, with expression appearing strongest in newly formed posterior somites, and reduced in the oldest, anterior somites (Fig. 6C and D). At stage 26, we detected *lecD* transcripts in the heart, somites, tail bud, and mesenchyme of the lower lip, ventral pharynx, and first pharyngeal arch (Fig. 6E). By st. T27, *lecD* labelled the mucocartilage of the first pharyngeal arch, lower lip, and ventral pharynx, while sectioning revealed additional expression in pharyngeal endoderm (Fig. 6F). By stage T28, expression was restricted to the heart, and mucocartilage of the lower lip ventral pharynx (Fig. 6G).

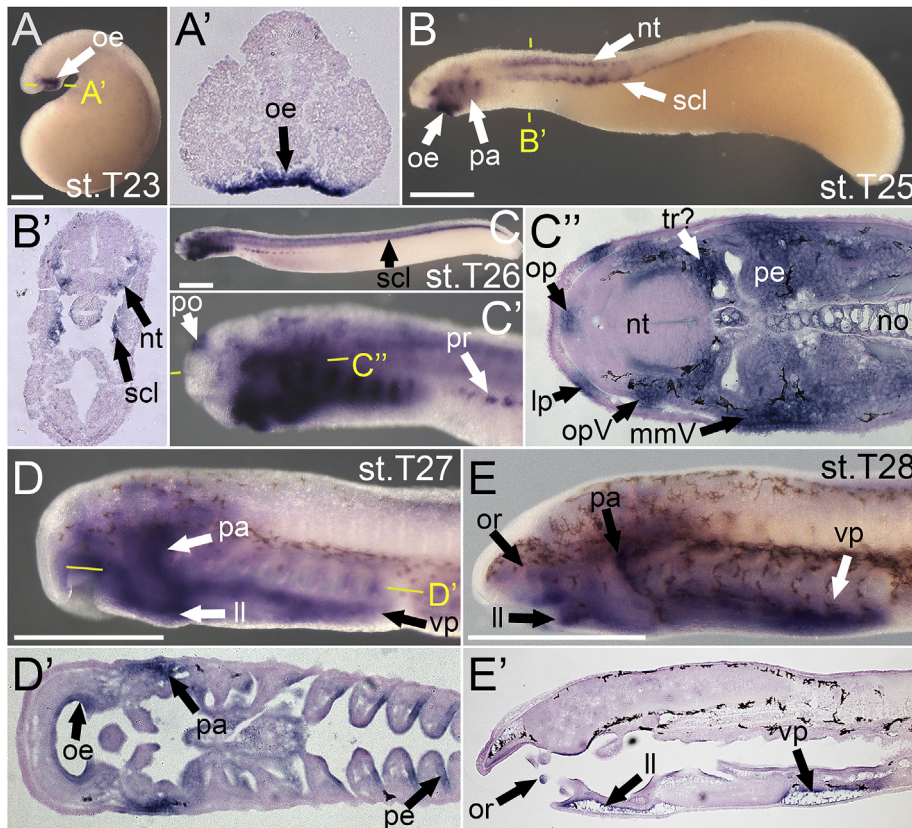


Fig. 4. Expression of *lecB* in *P. marinus* embryos and larvae. Left lateral view in all wholemount images. Scale bars represent 500 μ m. Staging as in Tahara, 1988. Panels with the same letter depict specimens of the same stage. Yellow lines and lettering indicate approximate plane of section for the indicated panel. (A) *lecB* transcription is first activated in the oral ectoderm at st. T23. (B) At st. T25, new expression is seen in the lateral neural tube at the level of the hindbrain, the pharynx, and the sclerotome. (C) At st. T26, *lecB* expression continues in the oral ectoderm, neural tube, pharynx, and sclerotome, and expands to include the presumptive pineal organ, pronephros, and the cranial placodes/ganglia. Sectioning reveals expression in the presumptive lens placode, olfactory placode, ophthalmic trigeminal placode, and maxillomandibular trigeminal placode. Additional expression is detected in the pharyngeal endoderm and head mesenchyme at the approximate position of the future trabecular cartilage. (D) *lecB* transcripts continue to mark the oral ectoderm, and are detected in the differentiating mucocartilage of pharyngeal arches 1 and 2 and the lower lip, which together form the velum. Expression is also seen in mucocartilage of the ventral pharynx. Sectioning at this stage reveals staining in stripes of pharyngeal endodermal. (E) As the oropharyngeal skeleton continues to differentiate during st. T28, *lecB* expression persists in the mucocartilage of the velum and ventral pharynx, and also marks the oral papillae. **Keywords:** ll: lower lip; lp: lens placode; mmV: maxillomandibular trigeminal ganglion; no: notochord; nt: neural tube; oe: oral ectoderm; op: olfactory placode; opV: ophthalmic trigeminal ganglion; or: oral papillae; pa: pharyngeal arch; pe: pharyngeal endoderm; po: pineal organ; pr: pronephros; scl: sclerotome; tr: trabecular; vp: ventral pharynx.

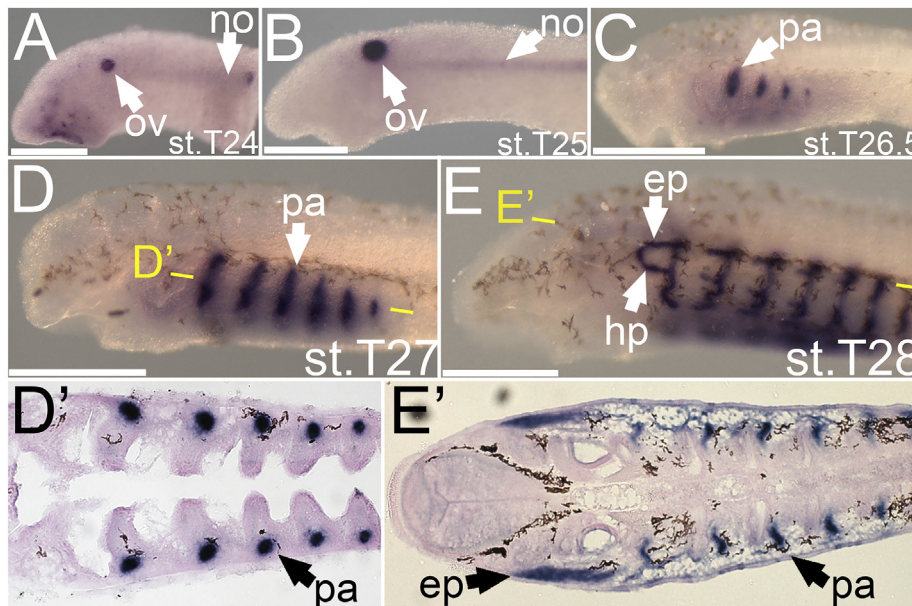


Fig. 5. Expression of *lecC* in *P. marinus* embryos and larvae. Left lateral view in all wholemount images. Scale bars represent 500 μ m. Staging as in Tahara, 1988. Panels with the same letter depict specimens of the same stage. Yellow lines and lettering indicate approximate plane of section for the indicated panel. (A) *lecC* transcription is first seen in the forming otic vesicle and notochord at st. T24. (B–C) From st. T25.5 to st. T26.5, expression in the otic vesicle and notochord ceases, while new expression is activated in middle of the posterior pharyngeal arches, tracking differentiation of the vertical pharyngeal arch cartilages. (D) At st. T27, expression is seen in all posterior pharyngeal arches (arches 3–7), and has expanded through the dorsoventral extent the most mature pharyngeal arch cartilages (arches 3–5). (E) At st. T28, *lecC* transcripts mark all pharyngeal arch cartilages, and become detectable in the differentiating epitrematic and hypotrematic processes of each. **Keywords:** ep: epitrematic process; hp: hypotrematic process; no: notochord; ov: otic vesicle; pa: pharyngeal arch cartilage.

3. Discussion

The evolution of vertebrate developmental and morphological novelties has been linked to a variety of genetic and genomic events, including the evolution of new gene regulatory interactions between ancient developmental regulators, the origin of new gene families, and

genome-wide duplication events (Martinez-Morales et al., 2007; Ohno, 1970; Meulemans and Bronner-Fraser, 2004; Martik et al., 2019; McLy-saght et al., 2002). To better understand the role of new gene families and gene duplications in vertebrate morphological evolution, we investigated the phylogeny and expression of *lectican* genes in the sea lamprey. To our knowledge, this work constitutes the first comprehensive expression

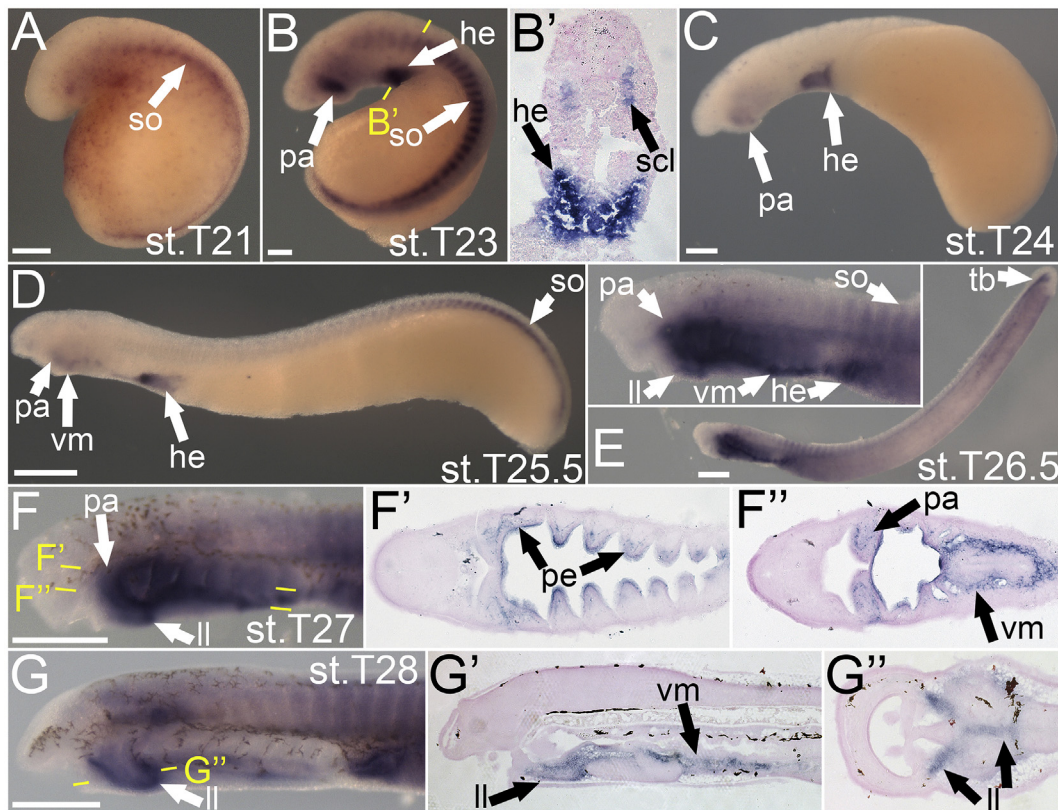


Fig. 6. Expression of *lecD* in *P. marinus* embryos and larvae. Left lateral view in all wholemount images. For panels A–C, scale bar represents 250 μ m. For panels D–G, scale bar represents 500 μ m. Staging as in Tahara, 1988. Panels with the same letter depict specimens of the same stage. Yellow lines and lettering indicate approximate plane of section for the indicated panel. (A) *lecD* transcripts are first detected at st. T21 in the somites. (B) At st. T23, additional *lecD* expression is seen in the first pharyngeal arch and nascent heart, while sectioning reveals somite expression is restricted to the presumptive sclerotome. (C) At st. T24, *lecD* expression continues in the first pharyngeal arch and heart, but is reduced in the anterior somites. (D) At st. T25.5 *lecD* transcripts are detected in the heart and mesenchyme of the ventral pharynx and first pharyngeal arch. Transcripts are also detectable in the somites, with the newly formed somites in the caudal third of larvae showing the strongest expression. (E) At st. T26.5, *lecD* expression becomes apparent in the tail bud. In the head, strong expression is seen in the heart, mesenchyme of the ventral pharynx, first pharyngeal arch, and lower lip (inset). Weak expression is detectable in the maturing somites, presumably the sclerotome (inset). (F–G) Expression in ventral mesenchyme, first pharyngeal arch and the lower lip continues as these cells differentiate into mucocartilage from st. T27–T28. Sectioning reveals additional expression in the pharyngeal endoderm. **Keywords:** he: heart; ll: lower lip; pa: first pharyngeal arch; pe: pharyngeal endoderm; scl: sclerotome; so: somites; tb: tail bud; vm: ventral mucocartilage.

analysis of all *lecticans* in a single vertebrate, and the first description of these genes in a jawless vertebrate.

3.1. Evolutionary history of the lectican family

Our own queries of invertebrate deuterostome genomes confirm previous work suggesting *lecticans* are *bona fide* vertebrate-specific genes (Kawashima et al., 2009b; Yoneda et al., 2010). However, these searches also suggest that exons encoding all of their main functional domains are present in invertebrate chordates, as previously shown (Kawashima et al., 2009b; Yoneda et al., 2010). In addition, gnathostome *lecticans* are linked to the closely related *haplins*, another vertebrate specific gene family (Spicer et al., 2003b). Together, these observations suggest *lecticans* arose by tandem duplication of an ancestral *hapln*-like gene in the vertebrate stem, followed by exon shuffling events that juxtaposed the *hapln*-related exons with other functional domains (Kawashima et al., 2009b; Yoneda et al., 2010). Consistent with this scenario, we found one lamprey *hapln* closely linked to the *lecD* locus (Fig. 2A). The timing of the duplication events that created the gnathostome and lamprey *lectican* families is less clear. Like previous reports, our phylogenetic analyses place all gnathostome *lecticans* into four paralogy groups, with *acan* + *bcn* and *vcn* + *ncn* forming two subfamilies. This topology is typical of gnathostome gene families and strongly suggests the four gnathostome *lecticans* were generated during the two vertebrate genome-scale

duplication events (1R and 2R) (Simakov et al., 2020; Dehal and Boore, 2005; Smith and Keinath, 2015; Guigo et al., 1996). In contrast, the relationships among lamprey *lecticans*, and between lamprey and gnathostome *lecticans*, are inconclusive. Regardless of tree building parameters, lamprey *lecticans* fail to consistently group with gnathostome paralogy groups, often clustering weakly with each other (Fig. 1B). Phylogenetic analyses of the neighboring genes *hapln* and *mef2*, and comparisons of syntenic genes yielded similarly inconclusive and weakly-supported phylogenies. There are several scenarios that could account for lack of clear one-to-one orthology between lamprey and gnathostome *lecticans*. One explanation is that the lamprey and gnathostome *lecticans* are the result of independent duplications of a single ancestral *lectican* in each lineage. At the other extreme, lamprey *lecticans* could be cryptic one-to-one orthologs of gnathostome *lecticans* that group together due to confounding factors such as the GC-bias of lamprey protein coding exons (Smith et al., 2013) or long branch attraction (Kuraku, 2008). Finally, the distinct lamprey and gnathostome *lectican* families could represent differential retention of ancient paralogs (Kuraku, 2013). For example, the vertebrate ancestor may have had eight *lecticans*, with a different set of four retained in each lineage. Various intermediate scenarios involving shared duplication, asymmetrical loss of paralogs, and independent duplication are also possible. However, a prerequisite for cryptic one-to-one orthology is that lamprey diverged from gnathostomes after the 2R genome duplications. Recent

comprehensive comparisons of chordate genome structure refute this, showing lamprey most likely diverged from gnathostomes before the second, “2R”, genome duplication (Simakov et al., 2020; Kuraku et al., 2009). If this is the case, the most parsimonious scenario is that the last common ancestor of lamprey and gnathostomes had two *lecticans*, an ancestral *acan* + *bcan* and an ancestral *vcan* + *ncan*, with independent duplications generating four *lecticans* in each lineage (Fig. 2B). In some support of this, we find 1) lamprey *Lecticans* cluster together in our phylogenies, 2) the genomic region surrounding *lecA* is *acan* + *bcan*-like but shows no particular similarity to either the *acan* or *bcan* regions (Fig. 2A, Fig. S8). In contrast, we find that none of the other lamprey *lectican* genomic regions are *ncan* and/or *vcan*-like. Based on our synteny analysis, we further speculate that *lecA* is likely derived from the *acan* + *bcan*-related 1R duplicate. The relationship between *lecB*, *lecC* and *lecD* and gnathostome *lecticans* is unresolved. While they could be cryptic *vcan* + *ncan* family members, it is also possible the *vcan* + *ncan* subfamily was lost in the lamprey lineage, and all lamprey *lecticans* are *acan* + *bcan* co-orthologs (Fig. 2B).

3.2. *Lectican* expression in the nervous system is ancestral within vertebrates

Regardless of their phylogenetic relationships, we found that almost every gnathostome *lectican* expression domain was conserved in lamprey, with only a few minor differences. To what degree these differences reflect divergence in *lectican* regulation between lamprey and gnathostomes, or the incomplete documentation of *lectican* expression in gnathostomes, is unclear. In the central nervous system, *lecA* and frog *vcan* both display expression in the neural tube floor plate (Fig. 3C') (Casini et al., 2004). Lamprey *lecB* expression is also observed in the lateral neural tube (Fig. 4B'), though there are no reports of gnathostome *lectican* transcription in this region. *lecA* and *lecB* are expressed in the developing brain like *bcan* and *ncan* (Fig. 3D', D'', Fig. 4D'), though not as broadly. Like *ncan*, *bcan*, and *vcan*, *lecA* and *lecB* are expressed in the cranial placodes and sensory ganglia (Fig. 3D', D'', Fig. 4D'), though in different neural populations (Szabó et al., 2019; Georgadaki and Zagris, 2014; Bell et al., 2004). Although *lectican* expression in the forming nervous system appears to be conserved among living vertebrates, the role of *lecticans* in neural development is unclear, as *bcan* and *ncan*-deficient mice show only minor differences in neuron function (Brakebusch et al., 2002; Zhou et al., 2001). Regardless of their precise functions, our data suggest that the LCA of cyclostomes and gnathostomes expressed *lecticans* in both the peripheral and central nervous systems.

3.3. *Lectican* expression in mesoderm-derived tissues is conserved across vertebrates

As in the forming nervous system *lectican* expression in mesodermal derivatives is largely conserved between lamprey and gnathostomes. In the gnathostome heart, *aggrekan* marks migratory cardiac mesoderm (Zanin et al., 1999), while *ncan* marks the forming myocardium and splanchnic mesoderm, and *vcan* marks the endocardium and the heart tube (Zanin et al., 1999; Georgadaki and Zagris, 2014; Mishima and Hoffman, 2003; Rambeau et al., 2017). In lamprey *lecA* and *lecD* are both expressed in the heart. As in neural tissue, the precise role of *lecticans* in the gnathostome heart is unclear, though mouse *vcan* mutants have major defects in the developing heart tube and endocardial cushion (Mukhopadhyay et al., 2006; Yamamura et al., 1997; Mjaatvedt et al., 1998).

Aside from cardiac mesoderm, we also noted expression of one or more lamprey *lecticans* in the notochord, pronephros, fin mesenchyme, and sclerotome. All of these mesodermal tissues express one or more *lecticans* in gnathostomes in temporal and spatial patterns virtually identical to their lamprey counterparts. The only notable difference in mesodermal *lectican* expression we observed was an absence of lamprey *lecticans* in somatic lateral plate mesoderm (LPM), which gives rise to *acan* and *vcan*-expressing skeletal tissue in gnathostome paired fins and

limbs.

3.4. Combinatorial *lectican* expression suggests lamprey possesses a diverse array of neural crest-derived skeletal tissues

We find that expression of multiple *lecticans* in forming and differentiated skeletal tissue is a conserved feature of vertebrate development. However, we also noted that gnathostomes typically transcribe only two *lecticans* in skeletogenic neural crest cells, *acan* and *vcan*, whereas lamprey expresses all four. Furthermore, lamprey *lecticans* are expressed in spatiotemporally distinct patterns throughout development, creating a combinatorial code of *lectican* expression in different parts of the nascent lamprey head skeleton. The histological heterogeneity of the lamprey head skeleton, which includes a mesenchymal chondroid tissue called mucocartilage, has been noted before (Schneider, 1879; Gaskell, 2019; Mangia and Palladini, 1970; Wright and Youson, 1982; Armstrong et al., 1987; Yao et al., 2011; Cattell et al., 2011). Anatomical work on adult hagfishes has also revealed diverse histology in the head skeleton (Ota and Kuratani, 2010; Ohtani et al., 2008; Parker, 1883; Cole, 1906), suggesting that the LCA of cyclostomes likely had multiple chondroid tissue types. It is possible the combinatorial co-expression of *lecticans* in the lamprey head skeleton elements reflects histological differences between different subtypes of mucocartilage. If this is the case, it would suggest that either 1) the LCA of cyclostomes and gnathostomes had a diversity of neural crest-derived chondroid tissues and the gnathostome lineage has retained only a few; or 2) the LCA of cyclostomes and gnathostomes had only a few neural crest-derived cartilage subtypes and the diversity seen in the sea lamprey head skeleton is a derived feature of lampreys, or cyclostomes. It has been previously shown that the pharyngeal skeleton of cyclostomes is patterned using the same basic mechanisms as seen in gnathostomes (Square et al., 2020; Square et al., 2017; Yao et al., 2011; Fujimoto et al., 2013; Cerny et al., 2010; Medeiros and Crump, 2012; Takio et al., 2007). In gnathostomes, this patterning acts a scaffold for proper deployment of the morphogenetic programs that control skeletal element shape and the tissue differentiation. In lamprey, which has a largely symmetrical oropharyngeal skeleton, this patterning may function mainly to control the activation of distinct differentiation programs in different parts of the head skeleton as previously proposed (Square et al., 2017; Cerny et al., 2010; Medeiros and Crump, 2012).

3.5. Different patterns of specialization and subfunctionalization after *lectican* duplication in lamprey and gnathostomes

Gene duplication is thought to facilitate evolutionary novelty by creating additional copies of genes that can then diverge to gain new expression domains and functions (neofunctionalization). More commonly, however, duplication leads to partitioning of ancestral expression domains (subfunctionalization) as described by the duplication-degeneration-complementation model (Force et al., 1999). Recent functional genomic comparisons have also highlighted the importance of specialization after duplication in the vertebrate lineage. During specialization, one paralog loses most aspects of its ancestral expression pattern and becomes specialized for a particular domain, while other paralogs maintain the complete ancestral pattern (Marlétaz et al., 2018). Our data suggest the ancestral *lectican* had a complex expression pattern that was then partitioned among its 1R duplicates. These paralogs were then independently duplicated in the lamprey and gnathostome lineages, with little apparent neofunctionalization in either (Fig. 7). We also find that the relative roles of specialization and subfunctionalization differ between the gnathostome and lamprey *lectican* families. Striking specialization is apparent in the lamprey *lectican* family, where *lecA* is transcribed in virtually all major *lectican* expression domains, while *lecC* has highly restricted expression in cell-rich hyaline cartilage (Fig. 7). Similarly, *lecD* transcripts are only seen in the sclerotome, heart, and a subpopulation of skeletogenic NCCs. In contrast, no

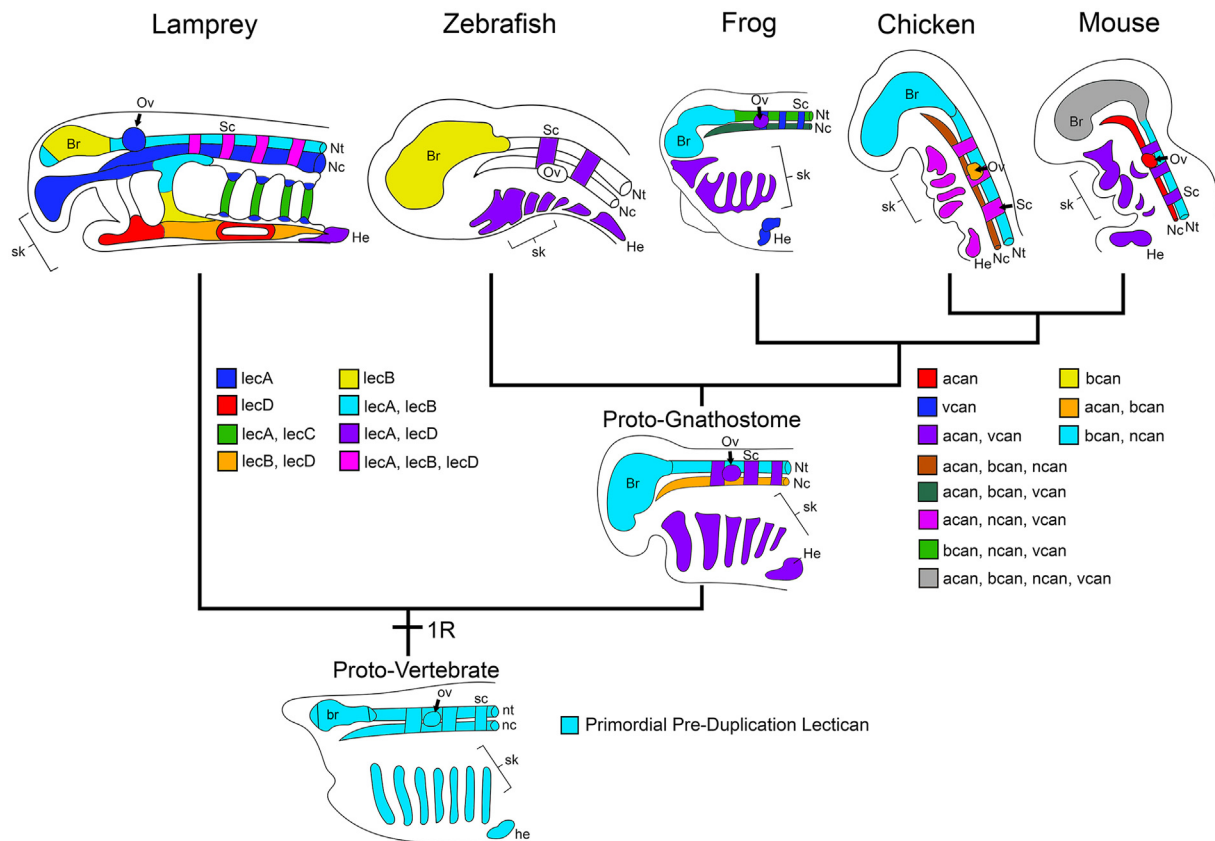


Fig. 7. The evolution of *lectican* expression patterns in the head. Modern gnathostome lectican expression is depicted based on current data for zebrafish (*Danio rerio*), frog (*Xenopus laevis*), chicken (*Gallus gallus*), and mouse (*Mus musculus*), and their consensus expression is depicted as an idealized proto-gnathostome. Expression data for chondrichthyans is currently not available. **Keywords:** br: brain; he: heart; nc: notochord; nt: neural tube; ov: otic vesicle; sc: sclerotome; sk: skeletal mesenchyme.

gnathostome *lectican* is so strictly specialized, and all paralogs are expressed in partially overlapping subsets of the ancestral expression pattern that could be described as “overlapping subfunctionalization” (Mundlos et al., 1991; Watanabe et al., 1998; Morawski et al., 2012; Zanin et al., 1999; Snow et al., 2005; Henderson and Copp, 1998; Frischknecht and Seidenbecher, 2012; Rauch et al., 1992; Oohira et al., 1994; Sander et al., 2001) (Fig. 7). Whether the different modes of expression pattern evolution have any significant consequences for Lectican protein function is unclear. Specialization of ohnologs is usually associated with rapid divergence in protein coding sequence (Marlétaz et al., 2018). Contrary to this prediction, lamprey *lecC*, the most specialized lamprey *lectican*, and *lecA*, the most broadly expressed lamprey *lectican* have similar, archetypical *lectican* structures. Meanwhile, all gnathostome *lecticans* vary significantly in length, and have lost and gained different functional domains. Nevertheless, it is provocative that both lamprey and gnathostomes typically express multiple *lecticans* in each expression domain. This suggests Lectican proteins are not entirely redundant, and supports the idea that combinations of functionally distinct Lectican proteins may confer subtle histological differences in related tissues.

3.6. The primordial lectican likely contributed to the evolution of vertebrate traits and functioned to support mesenchymal histology

Vertebrate evolution involved the acquisition of new organs, tissues, and cell types, as well as the elaboration of many pre-existing cell and tissue types. The *lectican* family is also a vertebrate novelty that arose at around the same time as these histological and morphological innovations. To what degree the evolution of new gene families drove the evolution of vertebrate traits is an open question. We used our expression data to ask if *lectican*-expressing cells and tissues were usually vertebrate

novelties, or had clear homologs in invertebrate chordates (Table 1). If a recognized homolog was present, we next asked if the histology of the vertebrate cell/tissue differed fundamentally from its invertebrate counterpart. These comparisons reveal that *lectican* transcription is largely restricted to cells and tissues that are either *bona fide* vertebrate novelties, or have unique histology in vertebrates.

If *lecticans* are indeed expressed mainly in vertebrate histological innovations, what specifically did the first *lectican* contribute to the vertebrate phenotype? Perhaps the best studied Lectican is Aggrecan, which is a major component of the hyaline cartilage ECM, and confers many of its defining histological and structural properties. It was previously thought that hyaline cartilage was unique to vertebrates, though a clear homolog with virtually all of its defining features has recently been described in the invertebrate chordate amphioxus (Jandzik et al., 2015). Thus, the origin of the first *lectican* was likely not prerequisite for the evolution of vertebrate-type cellular cartilage. Nevertheless, it is possible that *lecticans* contributed to the evolution of a more rigid type of cell rich hyaline cartilage, or the evolution of ECM-rich hyaline cartilage (Table 1).

Aside from hyaline cartilage, *lecticans* are expressed in a variety of other tissues during development. This suggests the evolution of the first *lectican* conferred a more general property upon vertebrate cells and tissues. Provocatively, a common theme among *lectican*-expressing cell types is that they differentiate from migratory and/or mesenchymal precursors (Table 1). Furthermore, gnathostome *lecticans* have been shown to regulate the migration of neural crest cells, the major mesenchymal cell type in the nascent vertebrate head (Perissinotto et al., 2000). We thus speculate that the primordial Lectican may have functioned to promote mesenchymal histology and/or migratory behavior during development. In support of this scenario, development in invertebrate deuterostomes is largely, or completely epithelial. Indeed, the

Table 1

Lecticans are mainly expressed in cells and tissues that are unique to vertebrates, or have distinct histology in vertebrates.

	Anterior Mesoderm and/or Cranial Neural Crest			Axial Mesoderm	Paraxial Mesoderm		Intermed. Meso.	Lateral plate Mesoderm		Oropharyngeal Epithelia		Neural	
	Cell-rich hyaline cartilage	ECM-rich hyaline cartilage	Muco-cartilage	Notochord	Fin Mes.	Scleroto-	Pro-nephros	Somatic	Cardiac	Phar. endoderm	Oral ectoderm	CNS neurons	PNS neurons
Gnathostomes	<i>acan</i> <i>vcan</i>	<i>acan</i> <i>vcan</i>	N/A	<i>acan</i> <i>bcan</i>	<i>acan</i>	<i>acan</i> <i>vcan</i>	<i>vcan</i>	<i>acan</i> <i>vcan</i>	<i>vcan</i>	<i>vcan</i>	<i>bcan</i>	<i>bcan</i> <i>ncan</i> <i>vcan</i>	<i>bcan</i> <i>ncan</i> <i>vcan</i>
Lamprey	<i>lecC</i> <i>lecA</i>	?	<i>lecA</i> <i>lecB</i> <i>lecD</i>	<i>lecA</i>	<i>lecA</i>	<i>lecA</i> <i>lecB</i>	<i>lecB</i>	No <i>lec</i>	<i>lecA</i> <i>lecD</i>	<i>lecB</i>	<i>lecB</i>	<i>lecA</i> <i>lecB</i>	<i>lecA</i> <i>lecB</i>
Urochordate	N/A	N/A	N/A	Present	N/A	N/A	N/A	N/A	Present	Present	Present	Present	Present
Cephalochordate	Present	N/A	N/A	Present	Present	Present	?	Present	N/A	Present	Present	Present	Present
Cell/tissue is unique to vertebrates?	No	Yes	Yes	No	No	No	Yes ⁴	No	No	No	No	No	No
Cell/tissue has distinct histology in vertebrates?	Yes ¹	-	-	No	Yes ²	Yes ³	-	Yes ⁵	No	No	No	No	Yes ⁶

1. Unlike amphioxus cell-rich hyaline cartilage, vertebrate cell-rich hyaline cartilage develops from mesenchymal cells and can differentiate into ECM-rich hyaline cartilage.
2. Amphioxus fin box mesoderm is epithelial and forms connective tissue, while vertebrate fin mesoderm delaminates and migrates as mesenchyme and can form cartilage and bone.
3. Amphioxus sclerotome is epithelial and forms connective tissue, while vertebrate sclerotome cells delaminate and migrate as mesenchyme and form cartilage and bone.
4. Amphioxus nephridia and the vertebrate kidney consist of different cell types with no clear evolutionary relationship. The vertebrate kidney forms from migratory mesenchymal cells.
5. The amphioxus homolog of somatic LPM is epithelial and forms connective tissue, while gnathostome LPM cells delaminate and migrate as mesenchyme and form cartilage and bone in the limbs.
6. The PNS neurons of invertebrate chordates migrate short distances as neuroblasts, vertebrate PNS neurons migrate as multipotent mesenchymal cells that aggregate to form ganglia.

invertebrate with the most vertebrate-like body plan, amphioxus, develops without any discernible mesenchyme (Conklin, 1932; Carvalho et al., 2020). Comprehensive analyses of lectican function in a wider range of vertebrates, including lamprey, using new methods for loss-of-function perturbation (Square et al., 2015) should help test this hypothesis.

4. Methods

4.1. Isolation of lamprey lectican homologs

Lamprey lectican sequences were assembled from transcriptomic reads of Tahara t. 26.5 embryos and late larval oral disc tissue that were previously gathered and submitted to GenBank (Sayers et al., 2021) (see Tab. S2 for accession numbers). Sequences from these files were used for our phylogenetic and syntenic analyses. For in situ hybridizations for *lecA*, primers were designed from lamprey genomic sequence to amplify conserved exon sequences, which were cloned into the pJet1.2 vector from ThermoFisher®. For *lecB*, *lecC*, and *lecD*, 500bp regions from transcriptomic sequences were selected and ordered as fragments in pUC57-amp vector from Synbio Tech®.

4.2. Phylogenetic analyses

Full-length human Lectican protein sequences were obtained from NCBI (Sayers et al., 2021) and used to identify lecticans and lectican-related genes in vertebrate and invertebrate deuterostome genomes using tBLASTN (Altschul et al., 1990). GenBank accession numbers for the sequences used in our analyses are given in Tab. S1–S6. In the case of sturgeon (*Acipenser oxyrinchus*), zebrafish lectican sequences were used to find transcript reads using tBLASTn which were subsequently assembled, translated, and searched against zebrafish genome to confirm their identity. To identify lamprey and hagfish sequences, we used zebrafish Lecticans to query lamprey (*P. marinus*) and hagfish (*Eptatretus burgeri*) transcriptomes using tBLASTn. These transcript reads were tiled,

assembled, and translated before being searched against gnathostome genomes confirm their identity. We then aligned all sequences using the PROBALIGN (Roshan and Livesay, 2006) program on CIPRES (Miller et al., 2010) servers. For all alignments, we used a gap open penalty of 20 and a gap extension penalty of 1. To determine the optimal substitution model for our phylogenetic analysis, we used ProtTest v3.4.2 (Abascal et al., 2005). For all tests, we allowed the possibility of invariant sites, empirical frequencies, and we used a fixed BIONJ tree topology to determine our ideal model. For our phylogenetic analysis, we used maximum likelihood analyses using RAXML-HP2 Workflow on CIPRES servers. Using the parameters recommended by ProtTest, our likelihood scores were bootstrapped with 1000 trees for each test to derive a consensus tree. Our consensus trees were lastly visualized using FigTree v1. 4.4 (Rambaut, 2018). Because of the ancient divergence of cyclostomes and gnathostomes, and the strong GC bias of lamprey protein coding sequences, we focused on full-length Lectican sequences for our primary phylogenetic analyses. Due to the fragmented condition of the currently sequenced cyclostome genomes, the only cyclostome Lecticans included (aside from the *P. marinus* Lecticans) were two unnamed hagfish (*E. burgeri*) Lecticans. Full-length gnathostome Lecticans were chosen to best represent the breadth of gnathostome diversity. For ray-finned fish, we chose zebrafish, gar, and sturgeon to include Lecticans that diverged before and after the teleost genome duplication. From the sarcopterygii, we included both amniote (human, mouse, chick) and anamniote (*X. laevis*) Lecticans. To represent the chondrichthyyi, two holocephalans were used (Elephant shark and Ghost shark). Because the N-terminal portions of Lecticans are evolutionarily related to HAPLNs, we also performed phylogenetic analyses using Lectican N-termini, HAPLNs, and invertebrate X-Link-containing genes. Finally, because Lecticans have overall more sequence conservation at the N- and C-termini, we tested substitution models that were specific to each termini, removing the intermediate domain for these phylogenetic analyses. For our original lectican test, a DCMut + I + G + F model was calculated with a log likelihood score of −80,905.555 under Akaike Information Criterion (AIC). For our HAPLN test, a JTT + G model was calculated with a log

likelihood score of −29,806.183 under AIC. Our *lectican* + hagfish test yielded a VT + I + G + F model with a log likelihood score of −152,828.685 under AIC. Our *lectican* + hagfish test yielded a VT + I + G + F model with a log likelihood score of −152,828.685 under AIC. Our HAPLN + N terminus test yielded a WAG + G + F model with a log likelihood score of −37,630.993 under AIC. When testing specific domain models, a WAG + I + G + F model was calculated for the N terminus with a log likelihood score of −37,763.00 under AIC. Conversely, our C terminus yielded a JTT + I + G model with a log likelihood score of −21,947.234 under AIC. Lastly, for our *mef2* test, a JTT + G + F model was calculated with a log likelihood score of −18,094.435 under AIC.

4.3. Synteny analyses

For our microsynteny analysis, we gathered peptide sequences of gnathostome *lecticans* on NCBI and found their respective genomic location using UCSC's Genome Browser and BLAT tool (Kent et al., 2002; Kent, 2002) as well as ENSEMBL (Yates et al., 2020). We then used sequences of elephant shark (*Callorhynchus milii*), dog (*Canis lupus familiaris*), chicken (*Gallus gallus*), mouse (*Mus musculus*), African clawed frog (*Xenopus tropicalis*), Spotted gar (*Lepisosteus oculatus*), and human (*Homo sapiens*) to reconstruct the ancestral arrangement of genes around the *acan*, *ncan*, and *vcn* loci. To do this, we compared genes in the ± 250–300 kb around each gnathostome locus. Linked genes conserved in six of the seven gnathostomes (or five of seven gnathostomes, if one of the five organisms was elephant shark, the most basally diverging gnathostome analyzed), were included in the reconstructed loci. The orientation of each linked gene was determined by the majority orientation. We then used UCSC's Genome Browser to identify all genes within ± 400 kb of each lamprey *lectican*. Because comparisons of the genes immediately adjacent to the gnathostome and lamprey *lectican* loci revealed very few conserved linked genes (Fig. 2A), we expanded our analysis to include a larger selection of syntenic genes (Fig. S8). To do this we identified the 20 genes (when available) immediately 5' and 3' of the *acan*, *ncan*, and *vcn* loci of three distantly related gnathostomes; chicken, elephant shark, and spotted gar. For *bcn*, which is not present in the current elephant shark genome assembly, we identified the 20 genes immediately 5' and 3' of the chicken, spotted gar, and zebrafish *bcn* loci. We then identified the 20 genes immediately 5' and 3' of lamprey *lecB* and *lecD*. For *lecC*, which sits near the 5' end of a scaffold, we identified the seven 5' genes, and the 33 3' genes. For *lecD*, which sits on a scaffold with less than 40 genes, all 29 flanking genes on the scaffold were used. For each gene identified as flanking a lamprey *lectican*, we then asked if there was a linked gene near one or more gnathostome *lecticans*. We then color-coded the lamprey genes based on the gnathostome *lectican*(s) its gene is linked to (Fig. S8).

4.4. Embryo collection and staging

Embryos for in situ hybridization were obtained from adult spawning-phase sea lampreys (*Petromyzon marinus*) collected from Lake Huron, MI, and kept in chilled holding tanks as previously described (Square et al., 2020). Embryos were staged according to the method of Tahara (1988), fixed in MEMFA (Mops buffer, EGTA, MgSO₄, and formaldehyde), rinsed in Mops buffer, dehydrated into methanol, and stored at −20 °C. All methods involving *P. marinus* were reviewed for ethical and humane care and use by the University of Colorado, Boulder Institutional Animal Care and Use Committee and approved as protocol #2392.

4.5. In situ hybridization

Riboprobes were made for anti-sense fragments using SP6 RNA Polymerase. Sequences for probes and genes are available upon request. In our experience, full-length *P. marinus* riboprobes, or riboprobes generated against untranslated regions of *P. marinus* transcripts, give higher

background than short riboprobes against coding sequences. We believe that this is because lamprey noncoding sequences, especially 3' UTRs, often have an excessive GC-repeat content, causing corresponding riboprobes to hybridize nonspecifically to off-targets. To mitigate this, we made short 500-bp riboprobes against coding regions and used a high-stringency hybridization protocol (Cerny et al., 2010; Square et al., 2016). Key parameters of this protocol include post-hybridization washes at 70 °C and the use of a low-salt, low-pH hybridization buffer (50% formamide; 1.3 × SSC, pH 5.0; 5 mM EDTA, pH 8.0; 50 µg/mL tRNA; 0.2% Tween-20; 0.5% CHAPS; and 100 µg/mL heparin).

4.6. Histology and sectioning

After in situ hybridization, embryos were postfixed in 4% paraformaldehyde/PBS (4 °C, overnight), rinsed in PBS, cryo-protected with 15% sucrose in water, embedded in 15% sucrose, 7.5% gelatin/15% sucrose (37 °C, several hours to overnight), and 20% gelatin/15% sucrose (37 °C overnight), frozen in −70 °C, and mounted with Tissue-Tek OCT compound (Sakura Finetek). Cryo-sections of 10 µm were collected on Super Frost Plus slides (Fisher Scientific), degelatinized in 3% gelatin in 38% ethanol, counterstained using Nuclear Fast Red (Vector Laboratories), dried, and cover-slipped with DPX (Fluka) (Jandzik et al., 2014).

4.7. Imaging

Whole-mount in situ hybridized *P. marinus* embryos and larvae were photographed using a Carl Zeiss Axiocam MRc5, Carl Zeiss Discovery V8 dissecting microscope, and Axiovision 4.9.1 software. Sections were photographed using a Carl Zeiss Imager A2 compound microscope.

Acknowledgements

The authors thank Scott Miehl at the USGS Hammond Bay Biological Station for providing adult sea lampreys. They also thank Jeremiah J. Smith at the University of Kentucky, Jr-Kai Yu at the Academia Sinica in Taipei, Taiwan, as well as Juan Pascual-Anaya and Shigeru Kuratani at the RIKEN Institute in Kobe, Japan for supplemental transcriptomic data. Zachary Root, Marek Romasek, Tyler Square, David Jandzik, and Daniel Medeiros were supported by National Science Foundation grants IOS 1656843 and IOS 1257040 to Daniel Medeiros. Zachary Root, Cara Allen, and Margaux Brewer were also supported by the Beverly Sears and EBIO grants through the University of Colorado Boulder. David Jandzik was additionally supported by the Scientific Grant Agency of the Slovak Republic VEGA grant No.1/0415/17.

Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ydbio.2021.03.020>.

References

- Abascal, F., Zardoya, R., Posada, D., 2005. ProtTest: selection of best-fit models of protein evolution. *Bioinformatics* 21 (9), 2104–2105.
- Altschul, S.F., et al., 1990. Basic local alignment search tool. *J. Mol. Biol.* 215 (3), 403–410.
- Armstrong, L., Wright, G.M., Youson, J., 1987. Transformation of mucocartilage to a definitive cartilage during metamorphosis in the sea lamprey, *Petromyzon marinus*. *J. Morphol.* 194 (1), 1–21.
- Barnette, D.N., et al., 2013. Tgfb-Smad and MAPK signaling mediate scleraxis and proteoglycan expression in heart valves. *J. Mol. Cell. Cardiol.* 65, 137–146.
- Bell, G.W., Yatskevich, T.A., Antin, P.B., 2004. GEISHA, a whole-mount in situ hybridization gene expression screen in chicken embryos. *Dev. Dynam.: Off. Publ. Am. Assoc. Anat.* 229 (3), 677–687.
- Bignami, A., Perides, G., Rahemtulla, F., 1993. Versican, a hyaluronate-binding proteoglycan of embryonal precartilaginous mesenchyma, is mainly expressed postnatally in rat brain. *J. Neurosci. Res.* 34 (1), 97–106.
- Braasch, L., Schartl, M., 2014. Evolution of endothelin receptors in vertebrates. *Gen. Comp. Endocrinol.* 209, 21–34.

- Brakebusch, C., et al., 2002. Brevican-deficient mice display impaired hippocampal CA1 long-term potentiation but show no obvious deficits in learning and memory. *Mol. Cell Biol.* 22 (21), 7417–7427.
- Carvalho, J.E., et al., 2020. An Updated Staging System for Cephalochordate Development: One Table Suits Them All. *bioRxiv*.
- Casini, P., et al., 2004. Identification and gene expression of versican during early development of *Xenopus*. *Int. J. Dev. Biol.* 52 (7), 993–918.
- Cattell, M., et al., 2011. A new mechanistic scenario for the origin and evolution of vertebrate cartilage. *PLoS One* 6 (7).
- Cerny, R., et al., 2010. Evidence for the prepattern/cooption model of vertebrate jaw evolution. *Proc. Natl. Acad. Sci. Unit. States Am.* 107 (40), 17262–17267.
- Cole, F.J., 1906. XXX—a monograph on the general morphology of the myxinoïd fishes, based on a study of *Myxine*. Part I. The anatomy of the skeleton. *Earth Environ. Sci. Trans. R. Soc. Edinburgh* 41 (3), 749–788.
- Conklin, E.G., 1932. The embryology of amphioxus. *J. Morphol.* 54 (1), 69–151.
- Courel, M.-N., et al., 1998. Hyaluronectin is produced by oligodendrocytes and Schwann cells in vitro. *J. Neurocytol.* 27 (1), 27–32.
- Dehal, P., Boore, J.L., 2005. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol.* 3 (10).
- Delarbre, C., et al., 2002. Complete mitochondrial DNA of the hagfish, *Eptatretus burgeri*: the comparative analysis of mitochondrial DNA sequences strongly supports the cyclostome monophyly. *Mol. Phylogenet. Evol.* 22 (2), 184–192.
- Eames, B.F., Gomez-Picos, P., Jandzik, D., 2020. On the Evolution of Skeletal Cells before and after Neural Crest. *Eames, B.F., Medeiros, D.M., Adameyko, I. (Eds.), In: Evolving Neural Crest Cells. CRC Press, pp. 185–218.*
- Force, A., et al., 1999. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151 (4), 1531–1545.
- Forey, P., Janvier, P., 1993. Agnathans and the origin of jawed vertebrates. *Nature* 361 (6408), 129–134.
- Frischknecht, R., Seidenbecher, C.I., 2012. Brevican: a key proteoglycan in the perisynaptic extracellular matrix of the brain. *Int. J. Biochem. Cell Biol.* 44 (7), 1051–1054.
- Fujimoto, S., et al., 2013. Non-parsimonious evolution of hagfish *Dlx* genes. *BMC Evol. Biol.* 13 (1), 15.
- Gans, C., Northcutt, R.G., 1983. Neural crest and the origin of vertebrates: a new head. *Science* 220 (4594), 268–273.
- Gary, S.C., Kelly, G.M., Hockfield, S., 1998. BEHAB/brevican: a brain-specific lectican implicated in gliomas and glial cell motility. *Curr. Opin. Neurobiol.* 8 (5), 576–581.
- Gaskell, W.H., 2019. The Origin of Vertebrates. *Good Press*.
- Georgiadaki, K., Zagris, N., 2014. Neurocan developmental expression and function during early avian embryogenesis. *Res. J. Dev. Biol.* 1 (1).
- Green, S.A., Bronner, M.E., 2014. The lamprey: a jawless vertebrate model system for examining origin of the neural crest and other vertebrate traits. *Differentiation* 87 (1–2), 44–51.
- Guigo, R., Muchnik, I., Smith, T.F., 1996. Reconstruction of ancient molecular phylogeny. *Mol. Phylogenet. Evol.* 6 (2), 189–213.
- Henderson, D.J., Copp, A.J., 1998. Versican expression is associated with chamber specification, septation, and valvulogenesis in the developing mouse heart. *Circ. Res.* 83 (5), 523–532.
- Hockman, D., et al., 2019. A genome-wide assessment of the ancestral neural crest gene regulatory network. *Nat. Commun.* 10 (1), 1–15.
- Jandzik, D., et al., 2014. Roles for FGF in lamprey pharyngeal pouch formation and skeletogenesis highlight ancestral functions in the vertebrate head. *Development* 141 (3), 629–638.
- Jandzik, D., et al., 2015. Evolution of the new vertebrate head by co-option of an ancient chordate skeletal tissue. *Nature* 518 (7540), 534–537.
- Jaworski, D.M., Kelly, G.M., Hockfield, S., 1995. The CNS-specific hyaluronan-binding protein BEHAB is expressed in ventricular zones coincident with gliogenesis. *J. Neurosci.* 15 (2), 1352–1362.
- Kang, J.S., et al., 2004. Characterization of dermacan, a novel zebrafish lectican gene, expressed in dermal bones. *Mech. Dev.* 121 (3), 301–312.
- Kawashima, T., et al., 2009a. Domain shuffling and the evolution of vertebrates. *Genome Res.* 19 (8), 1393–1403.
- Kawashima, T., et al., 2009b. Domain shuffling and the evolution of vertebrates. *Genome Res.* 19 (8), 1393–1403.
- Kent, W.J., 2002. BLAT—the BLAST-like alignment tool. *Genome Res.* 12 (4), 656–664.
- Kent, W.J., et al., 2002. The human genome browser at UCSC. *Genome Res.* 12 (6), 996–1006.
- Krueger, R., Hennig, A., Schwartz, N., 1992. Two immunologically and developmentally distinct chondroitin sulfate proteoglycans in embryonic chick brain. *J. Biol. Chem.* 267 (17), 12149–12161.
- Kuraku, S., 2008. Insights into cyclostome phylogenomics: pre-2R or post-2R. *Zool. Sci.* 25 (10), 960–968.
- Kuraku, S., 2013. Impact of asymmetric gene repertoire between cyclostomes and gnathostomes. *Semin. Cell Dev. Biol.* 24 (2), 119–127.
- Kuraku, S., Meyer, A., Kuratani, S., 2009. Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after? *Mol. Biol. Evol.* 26 (1), 47–59.
- Landolt, R.M., et al., 1995. Versican is selectively expressed in embryonic tissues that act as barriers to neural crest cell migration and axon outgrowth. *Development* 121 (8), 2303–2312.
- Mangia, F., Palladini, G., 1970. Histochemical studies on the mucocartilage of the lamprey during its larval ontogenesis. *Arch. Anat. Microsc. Morphol. Exp.* 59 (3), 283.
- Marlétaz, F., et al., 2018. Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature* 564 (7734), 64–70.
- Martik, M.L., et al., 2019. Evolution of the new head by gradual acquisition of neural crest regulatory circuits. *Nature* 574 (7780), 675–678.
- Martin, W.M., Bumm, L.A., McCauley, D.W., 2009. Development of the viscerocranial skeleton during embryogenesis of the sea lamprey, *Petromyzon marinus*. *Developmental Dynamics* 238 (12), 3126–3138.
- Martinez-Morales, J.-R., et al., 2007. New genes in the evolution of the neural crest differentiation program. *Genome Biol.* 8 (3), R36.
- McLysaght, A., Hokamp, K., Wolfe, K.H., 2002. Extensive genomic duplication during early chordate evolution. *Nat. Genet.* 31 (2), 200–204.
- Medeiros, D.M., Crump, J.G., 2012. New perspectives on pharyngeal dorsoventral patterning in development and evolution of the vertebrate jaw. *Dev. Biol.* 371 (2), 121–135.
- Meulemans, D., Bronner-Fraser, M., 2004. Gene-regulatory interactions in neural crest evolution and development. *Dev. Cell* 7 (3), 291–299.
- Meulemans, D., Bronner-Fraser, M., 2005. Central role of gene cooption in neural crest evolution. *J. Exp. Zool. B Mol. Dev. Evol.* 304 (4), 298–303.
- Milev, P., et al., 1998. Differential regulation of expression of hyaluronan-binding proteoglycans in developing brain: aggrecan, versican, neurocan, and brevican. *Biochem. Biophys. Res. Commun.* 247 (2), 207–212.
- Miller, M.A., Pfeiffer, W., Schwartz, T., 2010. *Creating the CIPRES Science Gateway for Inference of Large Phylogenetic Trees*. In *2010 Gateway Computing Environments Workshop (GCE)*. Ieee.
- Mishima, N., Hoffman, S., 2003. Neurocan in the embryonic avian heart and vasculature. *Anat. Rec. Part A: Discoveries in Molecular, Cellular, and Evolutionary Biology* 272 (2), 556–562.
- Miura, R., et al., 1999. The proteoglycan lectin domain binds sulfated cell surface glycolipids and promotes cell adhesion. *J. Biol. Chem.* 274 (16), 11431–11438.
- Mjaatvedt, C., et al., 1998. The *Cspg2* gene, disrupted in the *hdf* mutant, is required for right cardiac chamber and endocardial cushion formation. *Dev. Biol.* 202 (1), 56–66.
- Morawski, M., et al., 2012. Aggrecan: beyond cartilage and into the brain. *Int. J. Biochem. Cell Biol.* 44 (5), 690–693.
- Mukhopadhyay, A., et al., 2006. Erosive vitreoretinopathy and wagner disease are caused by intronic mutations in *CSPG2/Versican* that result in an imbalance of splice variants. *Invest. Ophthalmol. Vis. Sci.* 47 (8), 3565–3572.
- Mundlos, S., et al., 1991. Distribution of cartilage proteoglycan (aggrecan) core protein and link protein gene expression during human skeletal development. *Matrix* 11 (5), 339–346.
- Murakami, Y., et al., 2005. Evolution of the brain developmental plan: insights from agnathans. *Dev. Biol.* 280 (2), 249–259.
- Nikitina, N., Bronner-Fraser, M., Sauka-Spengler, T., 2009. Microinjection of RNA and morpholino oligos into lamprey embryos. *Cold Spring Harb. Protoc.* 2009 (1) pdb. prot5123.
- Northcutt, R.G., Gans, C., 1983. The genesis of neural crest and epidermal placodes: a reinterpretation of vertebrate origins. *Quarterly review of biology* 58 (1), 1–28.
- Ogawa, T., et al., 2001. Brevican in the developing hippocampal fimbria: differential expression in myelinating oligodendrocytes and adult astrocytes suggests a dual role for brevican in central nervous system fiber tract development. *J. Comp. Neurol.* 432 (3), 285–295.
- Ohno, S., 1970. *Evolution by Gene Duplication*. Springer-Verlag, New York.
- Ohtani, K., et al., 2008. Expression of Sox and fibrillar collagen genes in lamprey larval chondrogenesis with implications for the evolution of vertebrate cartilage. *J. Exp. Zool. B Mol. Dev. Evol.* 310 (7), 596–607.
- Oisi, Y., et al., 2013. Craniofacial development of hagfishes and the evolution of vertebrates. *Nature* 493 (7431), 175–180.
- Oohashi, T., et al., 2002. Bral1, a brain-specific link protein, colocalizing with the versican V2 isoform at the nodes of Ranvier in developing and adult mouse central nervous systems. *Mol. Cell. Neurosci.* 19 (1), 43–57.
- Oohira, A., et al., 1994. Developmentally regulated expression of a brain specific species of chondroitin sulfate proteoglycan, neurocan, identified with a monoclonal antibody 1G2 in the rat cerebellum. *Neuroscience* 60 (1), 145–157.
- Ota, K.G., Kuratani, S., 2010. Expression pattern of two collagen type 2 $\alpha 1$ genes in the Japanese inshore hagfish (*Eptatretus burgeri*) with special reference to the evolution of cartilaginous tissue. *J. Exp. Zool. B Mol. Dev. Evol.* 314 (2), 157–165.
- Ota, K.G., Kuraku, S., Kuratani, S., 2007. Hagfish embryology with reference to the evolution of the neural crest. *Nature* 446 (7136), 672–675.
- Owens, N.D., et al., 2016. Measuring absolute RNA copy numbers at high temporal resolution reveals transcriptome kinetics in development. *Cell Rep.* 14 (3), 632–647.
- Parker, W.K., 1883. On the skeleton of the marsipobranch fishes. Part I. The Myxinoïds (*Myxine*, and *Bdelostoma*). *Phil. Trans. Roy. Soc. Lond.* 174, 373–409.
- Perissinotto, D., et al., 2000. Avian neural crest cell migration is diversely regulated by the two major hyaluronan-binding proteoglycans PG-M/versican and aggrecan. *Development* 127 (13), 2823–2842.
- Perris, R., Johansson, S., 1990. Inhibition of neural crest cell migration by aggregating chondroitin sulfate proteoglycans is mediated by their hyaluronan-binding region. *Dev. Biol.* 137 (1), 1–12.
- Rambaut, A., 2018. FigTree V1. 4.4, a Graphical Viewer of Phylogenetic Trees. 2014. Java.
- Rambeau, P., et al., 2017. Reduced aggrecan expression affects cardiac outflow tract development in zebrafish and is associated with bicuspid aortic valve disease in humans. *Int. J. Cardiol.* 249, 340–343.
- Rauch, U., et al., 1992. Cloning and primary structure of neurocan, a developmentally regulated, aggregating chondroitin sulfate proteoglycan of brain. *J. Biol. Chem.* 267 (27), 19536–19547.
- Roshan, U., Livesay, D.R., 2006. Probalign: multiple sequence alignment using partition function posterior probabilities. *Bioinformatics* 22 (22), 2715–2721.

- Sander, V., Müllegger, J., Lepperdinger, G., 2001. *Xenopus* brevican is expressed in the notochord and the brain during early embryogenesis. *Mech. Dev.* 102 (1–2), 251–253.
- Sauka-Spengler, T., et al., 2007. Ancient evolutionary origin of the neural crest gene regulatory network. *Dev. Cell* 13 (3), 405–420.
- Sayers, E.W., et al., 2021. Database resources of the National center for biotechnology information. *Nucleic Acids Res.* 49 (D1), D10–D17.
- Schneider, A., 1879. *Anatomie und entwicklungsgeschichte von Petromyzon und ammocoetes*. Beiträge zur vergleichenden. Anatomie und Entwicklungsgeschichte der Wirbeltiere. Reimer., Berlin, pp. 85–92.
- Seidenbecher, C.I., et al., 1998. Transcripts for secreted and GPI-anchored brevican are differentially distributed in rat brain. *Eur. J. Neurosci.* 10 (5), 1621–1630.
- Shimeld, S.M., Donoghue, P.C., 2012. Evolutionary crossroads in developmental biology: cyclostomes (lamprey and hagfish). *Development* 139 (12), 2091–2099.
- Simakov, O., et al., 2020. Deeply conserved synteny resolves early events in vertebrate evolution. *Nat. Ecol. Evol.* 4, 820–830.
- Smith, J.J., Keinath, M.C., 2015. The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. *Genome Res.* 25 (8), 1081–1090.
- Smith, J.J., et al., 2013. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nat. Genet.* 45 (4), 415–421.
- Smith, J.J., et al., 2018. The sea lamprey germline genome provides insights into programmed genome rearrangement and vertebrate evolution. *Nat. Genet.* 50 (2), 270–277.
- Snow, H.E., et al., 2005. Versican expression during skeletal/joint morphogenesis and patterning of muscle and nerve in the embryonic mouse limb. *Anat. Rec. Part A Discov. Mol. Cell. Evol. Biol. Off. Publ. Am. Assoc. Anat.* 282 (2), 95–105.
- Spicer, A.P., Joo, A., Bowling, R.A., 2003a. A hyaluronan binding link protein gene family whose members are physically linked adjacent to chondroitin sulfate proteoglycan core protein genes the missing links. *J. Biol. Chem.* 278 (23), 21083–21091.
- Spicer, A.P., Joo, A., Bowling, R.A., 2003b. A hyaluronan binding link protein gene family whose members are physically linked adjacent to chondroitin sulfate proteoglycan core protein genes - the missing links. *J. Biol. Chem.* 278 (23), 21083–21091.
- Square, T., et al., 2015. CRISPR/Cas9-mediated mutagenesis in the sea lamprey *Petromyzon marinus*: a powerful tool for understanding ancestral gene functions in vertebrates. *Development* 142 (23), 4180–4187.
- Square, T., et al., 2016. Embryonic expression of endothelins and their receptors in lamprey and frog reveals stem vertebrate origins of complex Endothelin signaling. *Sci. Rep.* 6 (1), 1–13.
- Square, T., et al., 2017. The origin and diversification of the developmental mechanisms that pattern the vertebrate head skeleton. *Dev. Biol.* 427 (2), 219–229.
- Square, T.A., Jandzik, D., Massey, J.L., Romášek, M., Stein, H.P., Hansen, A.W., Purkayastha, A., Cattell, M.V., Medeiros, D.M., 2020. Evolution of the endothelin pathway drove neural crest cell diversification. *Nature* 585, 563–568. <https://doi.org/10.1038/s41586-020-2720-z>.
- Staudt, N., et al., 2015. A panel of recombinant monoclonal antibodies against zebrafish neural receptors and secreted proteins suitable for wholemount immunostaining. *Biochem. Biophys. Res. Commun.* 456 (1), 527–533.
- Szabó, A., et al., 2019. Neural crest streaming as an emergent property of tissue interactions during morphogenesis. *PLoS Comput. Biol.* 15 (4), e1007002.
- Tahara, Y., 1988. Normal stages of development in the lamprey, *Lampetra reissneri* (Dybowski). *Zool. Sci.* 5 (1), 109–118.
- Takio, Y., et al., 2007. Hox gene expression patterns in *Lethenteron japonicum* embryos—insights into the evolution of the vertebrate Hox code. *Dev. Biol.* 308 (2), 606–620.
- Van de Peer, Y., Maere, S., Meyer, A., 2009. The evolutionary significance of ancient genome duplications. *Nat. Rev. Genet.* 10 (10), 725–732.
- Wada, H., 2013. Domain shuffling and the evolution of vertebrate extracellular matrix. In: *Evolution of Extracellular Matrix*. Springer, pp. 27–37.
- Watanabe, H., Yamada, Y., Kimata, K., 1998. Roles of aggrecan, a large chondroitin sulfate proteoglycan, in cartilage structure and function. *J. Biochem.* 124 (4), 687–693.
- Wright, G.M., Youson, J.H., 1982. Ultrastructure of mucocartilage in the larval anadromous sea lamprey, *Petromyzon marinus* L. *Am. J. Anat.* 165 (1), 39–51.
- Yamada, H., et al., 1994. Molecular cloning of brevican, a novel brain proteoglycan of the aggrecan/versican family. *J. Biol. Chem.* 269 (13), 10119–10126.
- Yamamura, H., et al., 1997. A heart segmental defect in the anterior-posterior axis of a transgenic mutant mouse. *Dev. Biol.* 186 (1), 58–72.
- Yang, X., Si-Wei, Z., Qing-Wei, L., 2016. Lamprey: a model for vertebrate evolutionary research. *Zool. Res.* 37 (5), 263.
- Yao, T., et al., 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 (5), 339–346.
- Yates, A.D., et al., 2020. Ensemble 2020. *Nucleic Acids Res.* 48 (D1), D682–D688, 2020.
- Yoneda, M., et al., 2010. Evidence for the heparin-binding ability of the Ascidian Xlink domain and insight into the evolution of the Xlink domain in chordates. *J. Mol. Evol.* 71 (1), 51–59.
- York, J.R., McCauley, D.W., 2020. The origin and evolution of vertebrate neural crest cells. *Open Biol.* 10 (1), 190285.
- York, J.R., Lee, E.M.-J., McCauley, D.W., 2019. The lamprey as a model vertebrate in evolutionary developmental biology. In: *Lampreys: Biology, Conservation and Control*. Springer, pp. 481–526.
- Zanin, M.K., et al., 1999. Distinct spatial and temporal distributions of aggrecan and versican in the embryonic chick heart. *Anat. Rec. Off. Publ. Am. Assoc. Anat.* 256 (4), 366–380.
- Zelensky, A.N., Gready, J.E., 2005. The C-type lectin-like domain superfamily. *FEBS J.* 272 (24), 6179–6217.
- Zhou, X.-H., et al., 2001. Neurocan is dispensable for brain development. *Mol. Cell Biol.* 21 (17), 5970–5978.