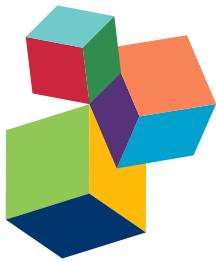


EVOLUTION OF ORGANISMAL FORM: FROM REGULATORY INTERACTIONS TO DEVELOPMENTAL PROCESSES AND BIOLOGICAL PATTERNS

EDITED BY: Sylvain Marcellini and Hector Escriva

PUBLISHED IN: Frontiers in Genetics and Frontiers in Ecology and Evolution
and Frontiers in Plant Science





frontiers

Frontiers Copyright Statement

© Copyright 2007-2017 Frontiers Media SA. All rights reserved.

All content included on this site, such as text, graphics, logos, button icons, images, video/audio clips, downloads, data compilations and software, is the property of or is licensed to Frontiers Media SA ("Frontiers") or its licensees and/or subcontractors. The copyright in the text of individual articles is the property of their respective authors, subject to a license granted to Frontiers.

The compilation of articles constituting this e-book, wherever published, as well as the compilation of all other content on this site, is the exclusive property of Frontiers. For the conditions for downloading and copying of e-books from Frontiers' website, please see the Terms for Website Use. If purchasing Frontiers e-books from other websites or sources, the conditions of the website concerned apply.

Images and graphics not forming part of user-contributed materials may not be downloaded or copied without permission.

Individual articles may be downloaded and reproduced in accordance with the principles of the CC-BY licence subject to any copyright or other notices. They may not be re-sold as an e-book.

As author or other contributor you grant a CC-BY licence to others to reproduce your articles, including any graphics and third-party materials supplied by you, in accordance with the Conditions for Website Use and subject to any copyright notices which you include in connection with your articles and materials.

All copyright, and all rights therein, are protected by national and international copyright laws.

The above represents a summary only. For the full conditions see the Conditions for Authors and the Conditions for Website Use.

ISSN 1664-8714

ISBN 978-2-88945-097-8

DOI 10.3389/978-2-88945-097-8

About Frontiers

Frontiers is more than just an open-access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

Frontiers Journal Series

The Frontiers Journal Series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the Frontiers Journal Series operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

Dedication to Quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews.

Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view.

By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the Frontiers Journals Series: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area! Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers Editorial Office: researchtopics@frontiersin.org

EVOLUTION OF ORGANISMAL FORM: FROM REGULATORY INTERACTIONS TO DEVELOPMENTAL PROCESSES AND BIOLOGICAL PATTERNS

Topic Editors:

Sylvain Marcellini, University of Concepcion, Chile

Hector Escriva, CNRS, UPMC université Paris 06, UMR 7232, BIOM, Observatoire Océanologique de Banyuls sur Mer, France



A live *Xenopus tropicalis* tadpole was exposed to a pulse of Alizarin red at stage NF56, and subsequently subjected to a Calcein pulse at stage NF60. This approach reveals growth patterns associated to the process of osteogenesis because it allows the simultaneous visualization of “old” and “newly deposited” mineralized bone matrix in red (bottom picture) and green (middle picture), respectively. The images show a dissected frontoparietal bone covered by some scattered pigment cells (black dots). Orientation of the merge picture: anterior (down) and posterior (up).

Authors: Carlos Henríquez and Sylvain Marcellini

Laboratory of Development and Evolution (LADE), Department of Cell Biology, Faculty of Biological Sciences, University of Concepcion, Concepcion, Chile.

Today’s biodiversity is the spectacular product of hundreds of millions of years of evolution. Understanding how this diversity of living organisms appeared is one of the most intriguing and challenging question in biology. Because organismal morphology is established during embryonic development, and because morphological traits diversified from ancestral forms during evolution, it can be inferred that changes in the mechanisms controlling embryonic development are instrumental for morphological evolution. This syllogism lies at the very heart

of a new discipline called Evo-Devo which is centered in the identification of the cellular and genetic mechanisms that, through modifications in developmental programmes, were at the base of morphological innovations during evolution.

After the discovery of the broad conservation of gene content and regulatory networks in the animal kingdom, as well as in plants, Evo-Devo is orienting towards the study of differences through experimental and functional approaches. Given the wide range of species, gene families, and developmental processes considered, a concerted effort is still required to shed light on the genetic, cellular and molecular mechanisms involved in phenotypic evolution. It is a particularly exciting time for this field of evolutionary developmental biology, as the advent of novel imaging, genome editing and sequencing technologies allows the study of almost any organism in ways that were unthinkable only a few years ago. Therefore, the aim of this Frontiers Research Topic is to gather an original collection of experimental approaches, concepts and hypotheses reflecting the current diversity of the Evo-Devo field. We have organized the articles according to the mechanistic depth with which they tackle specific evolutionary issues. Hence, comparisons of expression patterns have been grouped in Chapter 1, changes in regulatory interactions and gene networks are presented in Chapter 2, while Chapter 3 focuses on the evolution of developmental processes and biological patterns.

Citation: Marcellini, S., Escriva, H., eds. (2017). Evolution of Organismal Form: From Regulatory Interactions to Developmental Processes and Biological Patterns. Lausanne: Frontiers Media.
doi: 10.3389/978-2-88945-097-8

Table of Contents

- 05 Editorial: Evolution of Organismal Form: From Regulatory Interactions to Developmental Processes and Biological Patterns**

Sylvain Marcellini and Hector Escriva

CHAPTER 1. Expression patterns

- 07 Expression of Fox genes in the cephalochordate Branchiostoma lanceolatum**

Daniel Aldea, Anthony Leon, Stephanie Bertrand and Hector Escriva

- 19 Asymmetric Distribution of p110 and bruno2, New Members of a Conserved Core of Early Germline Determinants in Cephalochordates**

Simon C. Dailey, Roser Febrero Planas, Ariadna Rossell Espier, Jordi Garcia-Fernández and Ildikó M. L. Somorjai

- 31 Pax2/5/8 and Pax6 alternative splicing events in basal chordates and vertebrates: a focus on paired box domain**

Peter Fabian, Iryna Kozmikova, Zbynek Kozmik and Chrysoula N. Pantzartzi

- 43 Molecular footprinting of skeletal tissues in the catshark Scyliorhinus canicula and the clawed frog Xenopus tropicalis identifies conserved and derived features of vertebrate calcification**

Sébastien Enault, David N. Muñoz, Willian T. A. F. Silva, Véronique Borday-Birraux, Morgane Bonade, Silvan Oulion, Stéphanie Ventéo, Sylvain Marcellini and Mélanie Debiais-Thibaud

CHAPTER 2. Regulatory interactions and gene networks

- 57 On the evolutionary relationship between chondrocytes and osteoblasts**

Patsy Gómez-Picos and B. Frank Eames

- 72 The TALE face of Hox proteins in animal evolution**

Samir Merabet and Brigitte Galliot

- 85 A stranger in a strange land: the utility and interpretation of heterologous expression**

Elena M. Kramer

CHAPTER 3. Developmental processes and biological patterns

- 90 Evolution of epithelial morphogenesis: phenotypic integration across multiple levels of biological organization**

Thorsten Horn, Maarten Hilbrant and Kristen A. Panfilio

- 97 From limbs to leaves: common themes in evolutionary diversification of organ form**

Remco A. Mentink and Miltos Tsiantis

- 103 The differential view of genotype–phenotype relationships**

Virginie Orgogozo, Baptiste Morizot and Arnaud Martin



Editorial: Evolution of Organismal Form: From Regulatory Interactions to Developmental Processes and Biological Patterns

Sylvain Marcellini^{1*} and Hector Escriva²

¹ Laboratory of Development and Evolution, Department of Cell Biology, Faculty of Biological Sciences, Universidad de Concepción, Concepción, Chile, ² Observatoire Océanologique de Banyuls sur Mer, Centre National De La Recherche Scientifique, UPMC Université Paris 06, UMR 7232, BIOM, Banyuls sur Mer, France

Keywords: Evo-Devo, regulatory interactions, developmental processes, biological patterns, genomes

The Editorial on the Research Topic

Evolution of Organismal Form: From Regulatory Interactions to Developmental Processes and Biological Patterns

Living organisms display an astonishing morphological and behavioral diversity shaped by extrinsic environmental conditions and by intrinsic changes in developmental processes. In turn, such developmental trajectories are contingent on a myriad of regulatory interactions occurring at all possible steps of gene expression and cellular function. We are pleased to present a Frontiers Research Topic composed of 10 original research articles and reviews whose focus, ideas, and hypotheses reflect the current diversity and future directions of the field of Evo-Devo.

The evolution of gene families, gene expression patterns, and alternative splicing are addressed by examining germline determinants in cephalochordates and their implications for our understanding of multipotency and regeneration (Dailey et al.), by an extensive analysis of Fox members expressed during amphioxus early embryogenesis (Aldea et al.), and by deciphering the origin and expansion of alternative splicing of Pax genes in chordates (Fabian et al.). The evolution of cell types and tissue morphogenesis are addressed by illustrating, using the insect extraembryonic epithelia, how tissue architecture and physical context are crucial to understand gene function and evolution (Horn et al.), by proposing a model to explain how the transition between immature and mature cartilage might have facilitated the emergence of the osteoblastic regulatory network (Gomez-Picos and Eames), and by examining major fibrillary collagen genes expressed in the catshark and the clawed frog skeletons, thereby providing new insights on the origin of cartilage calcification (Enault et al.). Finally, some authors discuss important concepts in the field, such as the interpretations of heterologous assays and their possible pitfalls (Kramer), the convergent evolution of a two-steps morphogenetic mechanism controlling organ shape in plants and animals (Mentink and Tsiantis), the evolution of short peptide motives driving the generation of specific protein complexes involved in key bilaterian innovations (Merabet and Galliot), and the intricate relationships linking the genotypic and phenotypic dimensions (Orgogozo et al.).

So, what is the broad contribution to Evo-Devo of the 10 aforementioned manuscripts, and how do they relate to the future research directions that this discipline must prioritize in order to remain both successful and attractive to the community? Hints to our first answer are included within the topic title itself, inspired from a François Jacob influential review emphasizing the complex and crucial relationships between different levels of biological organization (Jacob, 1977). Undoubtedly, understanding how specific mutations and environmental factors affect molecular

OPEN ACCESS

Edited and reviewed by:

Samuel A. Cushman,
United States Forest Service Rocky
Mountain Research Station, USA

*Correspondence:

Sylvain Marcellini
smarcellini@udec.cl

Specialty section:

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

Received: 31 May 2016

Accepted: 03 August 2016

Published: 17 August 2016

Citation:

Marcellini S and Escriva H (2016)
Editorial: Evolution of Organismal
Form: From Regulatory Interactions to
Developmental Processes and
Biological Patterns.
Front. Genet. 7:148.
doi: 10.3389/fgene.2016.00148

networks, cells, organs, species, and ecosystems, represents one of the most stimulating Evo-Devo conceptual frameworks. While it is a long way to connect “regulatory interactions to developmental processes and biological patterns,” the initial and obligatory step is to report expression patterns of key molecular actors (Aldea et al.; Fabian et al.; Dailey et al.). One can then move to higher hierarchical levels, for instance by understanding the mechanisms facilitating the emergence of regulatory interactions involved in body plan patterning (Merabet and Galliot), or by integrating gene activity, cellular behavior and mechanical forces to reach a comprehensive view of embryonic evolution (Horn et al.). Our second answer relates to our ability to cope with a modern era continuously flooded by ever-improving imaging, genome editing and sequencing technologies. As a consequence, data analysis, and not data generation, should be our present priority (Moore, 2012). As recently argued, facing the next grand challenge in evolutionary biology will require a strong synergy between three major branches of the field: Experimental data, genomics, and modeling (Cushman, 2014). According to this strategy, evolutionary models must be validated by performing carefully designed and controlled experiments, which, in the case of heterotopic functional assays, should follow the guidelines

proposed by Kramer. This multidisciplinary approach will be particularly well-suited to extract universal principles underlying the development of multicellular organisms (Mentink and Tsiantis), to decipher the evolution of the bone and cartilage gene regulatory networks (Enault et al.; Gomez-Picos and Eames), or to understand, for any species of interest, how the interaction between genotype and environment generates complex phenotypic spaces (Cushman, 2014, Orgogozo et al.). These are exciting times for the Evo-Devo community, we hope that you will enjoy this collection of articles and look forward in the near future to reading any follow-up work that it will have inspired.

AUTHOR CONTRIBUTIONS

SM and HE wrote, edited, and revised the manuscript.

ACKNOWLEDGMENTS

This collaborative effort has been supported a FONDECYT research grant 1151196 to SM.

REFERENCES

- Cushman, S. A. (2014). Grand challenges in evolutionary and population genetics: the importance of integrating epigenetics, genomics, modeling, and experimentation. *Front. Genet.* 5:197. doi: 10.3389/fgene.2014.00197
- Jacob, F. (1977). Evolution and tinkering. *Science* 196, 1161–1166. doi: 10.1126/science.860134
- Moore, A. (2012). Have we produced enough results yet, sir? *Bioessays* 34:163. doi: 10.1002/bies.201290005

Conflict of Interest Statement: 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 © 2016 Marcellini and Escriva. 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) or licensor 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.

Expression of Fox genes in the cephalochordate *Branchiostoma lanceolatum*

Daniel Aldea, Anthony Leon, Stephanie Bertrand* and Hector Escriva *

Centre National De La Recherche Scientifique, UPMC université Paris 06, UMR 7232, BIOM, Observatoire Océanologique de Banyuls sur Mer, Banyuls sur Mer, France

OPEN ACCESS

Edited by:

Naoki Osada,
Hokkaido University, Japan

Reviewed by:

Jr-Kai Yu,
Academia Sinica, Taiwan
Haruki Ochi,
Yamagata University, Japan

***Correspondence:**

Stephanie Bertrand and

Hector Escriva,

Laboratoire Arago, Avenue du
Fontaulé, 66650 Banyuls-sur-Mer,

France

stephanie.bertrand@obs-banyuls.fr;
hecriva@obs-banyuls.fr

Specialty section:

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

Received: 29 May 2015

Accepted: 07 July 2015

Published: 28 July 2015

Citation:

Aldea D, Leon A, Bertrand S and
Escriva H (2015) Expression of Fox
genes in the cephalochordate
Branchiostoma lanceolatum.
Front. Ecol. Evol. 3:80.
doi: 10.3389/fevo.2015.00080

Forkhead box (Fox) genes code for transcription factors that play important roles in different biological processes. They are found in a wide variety of organisms and appeared in unicellular eukaryotes. In metazoans, the gene family includes many members that can be subdivided into 24 classes. Cephalochordates are key organisms to understand the functional evolution of gene families in the chordate lineage due to their phylogenetic position as an early divergent chordate, their simple anatomy and genome structure. In the genome of the cephalochordate amphioxus *Branchiostoma floridae*, 32 Fox genes were identified, with at least one member for each of the classes that were present in the ancestor of bilaterians. In this work we describe the expression pattern of 13 of these genes during the embryonic development of the Mediterranean amphioxus, *Branchiostoma lanceolatum*. We found that *FoxK* and *FoxM* genes present an ubiquitous expression while all the others show specific expression patterns restricted to diverse embryonic territories. Many of these expression patterns are conserved with vertebrates, suggesting that the main functions of Fox genes in chordates were present in their common ancestor.

Keywords: Fox genes, amphioxus, Evo-Devo, chordates, embryonic development

Introduction

Forkhead box (Fox) transcription factors originated early during evolution and are specific to opisthokonts. They are present in fungi as well as in metazoans (Mazet et al., 2006; Larroux et al., 2008; Shimeld et al., 2010a) in which they play essential roles during embryonic development (Carlsson and Mahlapuu, 2002; Tuteja and Kaestner, 2007a,b; Benayoun et al., 2011). Fox proteins possess a helix-turn-helix DNA-binding domain called the forkhead domain which corresponds to a conserved region of approximately 110 amino acids (Weigel and Jackle, 1990; Clark et al., 1993). A molecular phylogeny-based classification of the Fox gene family allowed to propose its subdivision into 24 classes (ranged from *FoxA* to *FoxS* and including subfamilies that were recently subdivided: *FoxJ* (*FoxJ1* and *FoxJ2*), *FoxL* (*FoxL1* and *FoxL2*), and *FoxN* (*FoxN1/4* and *FoxN2/3*) (Mazet et al., 2003). Many Fox gene losses or duplications occurred in different bilaterian clades, affecting different Fox classes. For example, *FoxAB* is found in cephalochordates and in the sea urchin but not in tunicates or vertebrates (Tu et al., 2006; Yu et al., 2008a), and families R and S are vertebrate-specific (Wotton and Shimeld, 2006; Shimeld et al., 2010b). Using phylogenetic analyses, it has been proposed that 22 Fox gene families were already present in the bilaterian ancestor (Shimeld et al., 2010b).

Cephalochordates (i.e., amphioxus) belong to the chordate phylum together with tunicates and their sister group, the vertebrates. They present morphological, developmental, and genomic characteristics that are proposed to be very similar to the ancestral state in the chordate clade, making amphioxus a key model system to understand chordate evolution (Bertrand and Escriva, 2011, 2014). Interestingly, it has been shown that amphioxus is the only living bilaterian possessing at least one member of each of the 22 Fox gene families proposed to have been present in Urbilateria (Yu et al., 2008a). Thus, the study of Fox genes in this cephalochordate may shed light on the functional evolutionary history of this transcription factor gene family. Past studies using genomic data from the Caribbean cephalochordate *Branchiostoma floridae* described the presence of 32 Fox genes in this species (Yu et al., 2008a) and the expression pattern of 11 of these genes was previously described: *FoxAa* and *FoxAb* (formerly named AmHNF3-1 and AmHNF3-2, respectively) (Shimeld, 1997), *FoxB* (Mazet and Shimeld, 2002), *FoxC* (Mazet et al., 2006), *FoxD* (Yu et al., 2002b), *FoxE4* (Yu et al., 2002a), *FoxF* (Mazet et al., 2006; Onimaru et al., 2011), *FoxG* (Toresson et al., 1998), *FoxL1* (Mazet et al., 2006), *FoxN1/4a* (Bajoghli et al., 2009), *FoxQ1* and *FoxQ2* (Yu et al., 2003; Mazet et al., 2006). In this work we searched for Fox sequences in the transcriptome of the Mediterranean amphioxus *Branchiostoma lanceolatum*. We found 28 Fox sequences and we describe here the spatiotemporal expression pattern of 13 Fox genes during embryonic development, including seven previously described in *B. floridae* and six for which expression was not known. We show that in *B. lanceolatum* some Fox genes exhibit ubiquitous expression as *FoxK* and *FoxM*, while the others show specific and dynamic expression patterns restricted to diverse embryonic territories. These expression patterns suggest that Fox genes are performing both general and specific functions during amphioxus embryonic development, most of them being probably ancestral in the chordate clade.

Materials and Methods

Phylogenetic Analysis

All reference sequences, except for *B. lanceolatum*, were obtained from Genbank or from Fritzenwanker et al. (2014). The multiple alignment was performed only for the conserved Forkhead amino acid domain sequences using the MUSCLE module implemented in MEGA 6 and manually refined in its interface (Tamura et al., 2013). The best fit substitution model for phylogenetic reconstruction was estimated using MEGA 6 (Tamura et al., 2011). Bayesian inference (BI) tree was inferred using MrBayes 3.2 (Ronquist et al., 2012), with the model recommended by MEGA 6 under the Akaike information criterion (RtRev+Γ), at the CIPRES Science Gateway V. 3.1 (Miller et al., 2015). Two independent runs were performed, each with four chains and 1 million generations. A burn-in of 25% was used and a 50 majority-rule consensus tree was calculated for the remaining trees.

Cloning and Expression Study

B. lanceolatum Fox sequences were recovered from its reference transcriptome (Oulion et al., 2012) by TBLASTN using sequences

from *B. floridae* as queries. Specific primers were then designed for RT-PCR amplification from total RNA. Primer sequences are as follow:

```

FoxA_a_5' AAGTCGCCGGTGTACGAGATG
FoxA_a_3' GTATTATAGAGACGAAGGTTG
FoxA_b_5' CATTTCCTCAGAACAGACATG
FoxA_b_3' TCCTAAAGACTCCCAACAACA
FoxAB_5' CAGTGTGAGGTGAACATCATG
FoxAB_3' CGATTGACAGGTTGATAGAAC
FoxB_5' ACAACAGGACCCCTGACTCGT
FoxB_3' GCATTCCCTGACGTCTTG
FoxC_5' AACCGTCCCCTTTCCTCATG
FoxC_3' CAGTTTGATTGTAAGGACT
FoxD_5' ACAGCTGTGGAGTGGACACTT
FoxD_3' CACGAGACATGTAAGTCTCCG
FoxEa_5' AACCAACCCGTACCAGCATG
FoxEa_3' ATATGACACGGACACTGAAC
FoxG_5' ACGCACATTAGCACAGTTG
FoxG_3' ACTTGACCCCTGGCTTGACAC
FoxJ1_5' TACAGACAACGTAAACCATG
FoxJ1_3' TTGTAATGCAGGGTGGGGCCT
FoxK_5' GGAAGGCGGAGTTGGACAATG
FoxK_3' CCGGACACGTCCCTGCACCTGT
FoxM_5' AGGAGAGTGTGACAAACCATG
FoxM_3' TTCTCAGCTATTCAAGTAATAC
FoxN1/4a_5' GCGCACCGAGTATCGTTCTGA
FoxN1/4a_3' ACATAGGTAGGACTATGTACT
FoxN2/3_5' CAGTAAACACGAGCAGACATG
FoxN2/3_3' AGCTGAAGACAATGATGATCC

```

A mix of total mRNA of *B. lanceolatum* extracted from embryos at different developmental stages was used as a template for retro-transcription. Amplification was performed using Advantage 2 Polymerase kit (Clontech) and a touch-down PCR program with annealing temperature ranging from 65 to 40°C. Amplified fragments were cloned using the pGEM-T Easy system (Promega) and sub-cloned in pBluescript II KS+ for probe synthesis.

Whole Mount *In situ* Hybridization

Probes were synthesized using the DIG labeling system (Roche) after plasmid linearization with the appropriate enzymes. Ripe animals of *B. lanceolatum* were collected in Argelès-sur-Mer (France), and gametes were obtained by heat stimulation (Fuentes et al., 2004, 2007). *In vitro* fertilization was undertaken in Petri dishes filled with filtered sea water. Fixation and whole mount *in situ* hybridization were performed as described in Somorjai et al. (2008).

Results

Molecular Phylogenetic Analysis of *B. lanceolatum* Fox Gene Sequences

We looked for Fox gene sequences in the reference transcriptome of *B. lanceolatum* (Oulion et al., 2012). The sequences that were recovered were used to conduct a phylogenetic

tree reconstruction presented in **Figure 1**. We showed that *B. lanceolatum* possesses at least 28 Fox genes, each of them being orthologous to one of the 32 genes described in *B. floridae* and corresponding to at least one member of each of the 22 families present in the bilaterian ancestor (Yu et al., 2008a). Specific duplications, that occurred in the cephalochordate clade at least in the ancestor of *B. floridae* and *B. lanceolatum*, gave rise to three members in the FoxQ2 group (*FoxQ2a*, *FoxQ2b*, *FoxQ2c*), two members in the FoxN1/4 group (*FoxN1/4a* and *FoxN1/4b*), and two genes in the FoxE group (*FoxEa* and *FoxEc*). We then analyzed the expression pattern during *B. lanceolatum* embryonic development of 13 of these 28 Fox genes corresponding to those showing a higher expression level in the transcriptome (Oulion et al., 2012).

FoxAa and FoxAb

FoxAa (formerly named *AmHNF3-1*) (Shimeld, 1997) was first expressed at the gastrula stage in the anterior ventral endoderm and in the mesendodermal layer of the dorsal blastoporal lip (**Figures 2A,B**). At the late gastrula stage, we detected transcripts in the axial dorsal mesendoderm corresponding to the presumptive notochord territory, as well as in mesendoderm cells of the archenteron floor (**Figures 2C,D**). Expression in the axial mesoderm and endoderm persisted through mid-late neurula stage (**Figures 2E,F**). Later on, at late neurula stage before the mouth opens, the expression in the notochord was restricted to the most anterior and posterior tips of the embryo, while the endodermal expression was restricted to the middle region of the gut (**Figure 2G**). At the larva stage, the expression at the anterior tip of the notochord and in the tailbud was still observed and we detected a diffuse expression in the gut (**Figure 2H**).

FoxAb (formerly named *AmHNF3-2*) (Shimeld, 1997) expression was first detected at the gastrula stage as a weak signal in the mesendodermal part of the dorsal blastoporal lip (**Figures 2I,J**). At the late gastrula stage, we detected expression in the central paraxial mesoderm on both sides of the notochord anlagen (**Figures 2K,L**). At the mid-late neurula stage transcripts were detected in the neural tube, including the cerebral vesicle, and in the dorsal part of the endoderm (**Figures 2M,N**). At the late neurula stage, before the mouth opens, *FoxAb* was expressed in the neural tube and in the most anterior part of the pharynx. In the posterior region, expression was detected in the tailbud and in the dorsal midline of the gut (**Figure 2O**). At the larva stage, we observed expression in the pharynx, in the preoral pit, in the club-shaped gland and in the tailbud. At this stage, the expression in the neural tube gets restricted to some neurons and to the posterior part of the cerebral vesicle (**Figure 2P** and Figure S1A).

FoxAB

FoxAB transcripts were detected as a weak and ubiquitous signal from the eight-cell stage to the blastula stage (**Figures 2Q,R**). This ubiquitous expression was confirmed by the presence of reads in transcriptome analyses (data not shown). At the gastrula stage we observed a strong specific expression in the dorsal blastoporal lip, the amphioxus putative organizer (**Figures 2S,T**).

At the late gastrula stage, expression gets restricted to the presumptive notochord territory (**Figures 2U,V**). No expression could be detected by *in situ* hybridization in later stages.

FoxB

FoxB expression was first detected dorsally, both in the ectoderm and in the mesendoderm, as a weak signal in mid gastrula stage embryos (**Figures 2W,X**). Later on, in early neurula stage embryos, a signal could be observed in the neural plate on either side of the midline, as well as in two patches in the posterior paraxial mesendoderm (**Figures 2Y,Z**). During the late neurula stage, expression was detected in the most posterior paraxial mesoderm that give rise to the newly formed somites and in the neural tube posterior to the cerebral vesicle (**Figures 2A',B'**). Then, *FoxB* expression in the mesoderm faded away in late neurulae (**Figure 2C'**) and get later restricted to the cerebral vesicle and to some neurons along the neural tube in larvae (**Figure 2D'** and Figure S1B).

FoxC

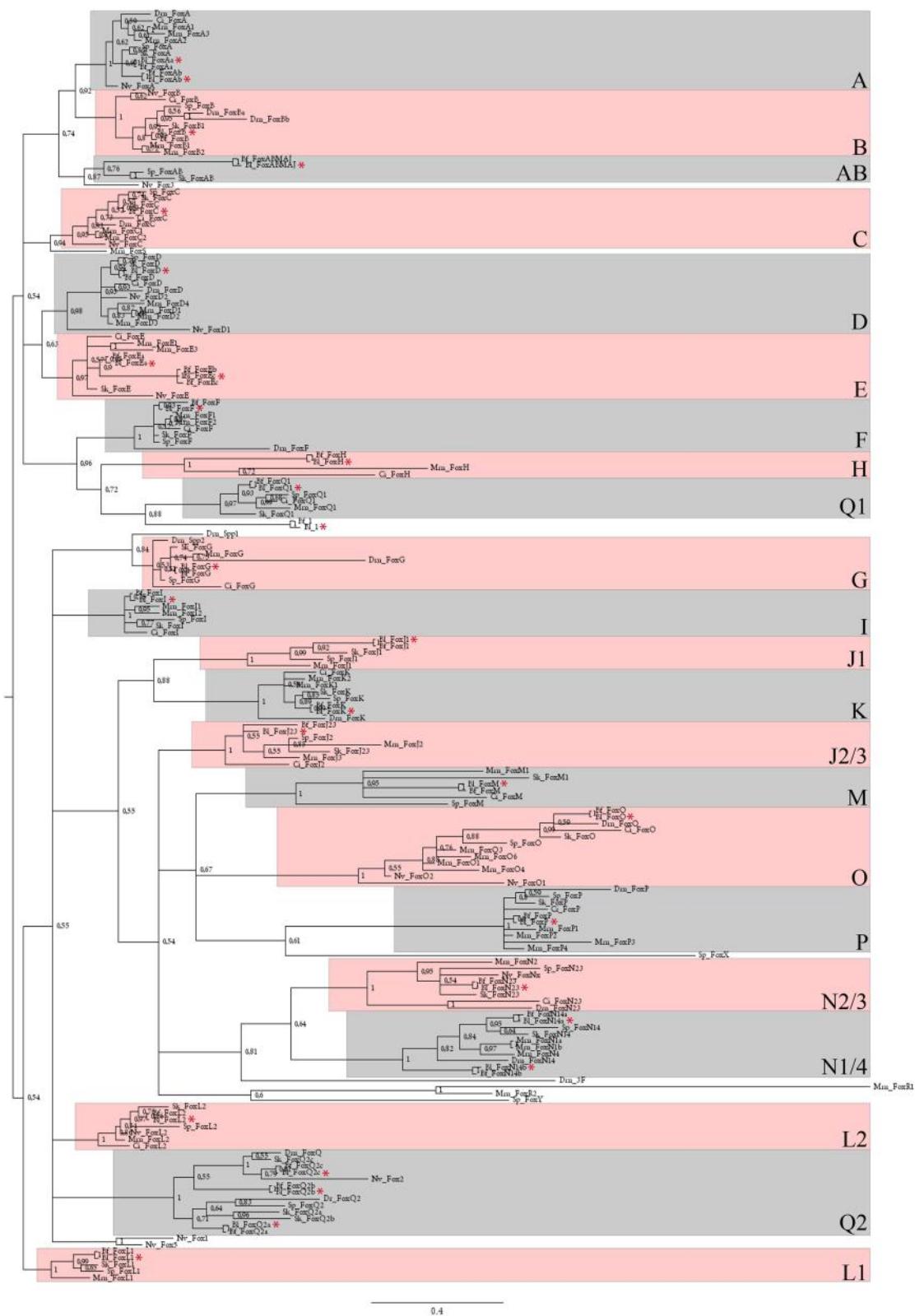
FoxC was expressed at the gastrula stage in the dorsal paraxial mesendoderm (**Figures 3A,B**). Later on, at the late gastrula stage, expression was detected in the region that gives rise to the three most anterior somites (**Figures 3C,D**). In mid-late neurulae, the transcripts remained all along the body in the somites and a new expression domain appeared in the anterior endoderm at the level where the first gill slit opens (**Figure 3E**). At the late neurula stage, the expression persisted in the pharynx and somites and was also detected in the club-shaped gland anlagen (**Figures 3F,G**). At the larva stage a diffuse expression was observed in the somites as well as in the preoral pit, in the club-shaped gland and in the first gill slit (**Figure 3H** and Figure S1C).

FoxD

FoxD transcripts were first detected at the gastrula stage in the dorsal blastoporal lip (**Figures 3I,J**). Then, at the late gastrula stage, *FoxD* was expressed in the dorsal axial mesendoderm, in part of the dorsal paraxial mesendoderm as two patches on both sides of the midline and in the anterior region of the neural plate (**Figures 3K,L**). At the mid-late neurula stage, the notochord and the somites, as well as the cerebral vesicle, were labeled (**Figure 3M**). At the late neurula stage, before the mouth opens, transcripts were detected in the paraxial somitic mesoderm, in the notochord, in the cerebral vesicle and in the posterior endoderm (**Figures 3N,O**). A faint labeling was also detected at this stage in the first gill slit and in the club-shaped gland anlagens. At the larva stage, we observed a low expression level in the cerebral vesicle, in the preoral pit, in the club-shaped gland, in the first gill slit, in the notochord and in the posterior part of the gut. We also observed an anterior to posterior gradient of expression in the somites (**Figure 3P** and Figure S1D).

FoxEa

FoxEa (formerly named *FoxE4* in *B. floridae*) expression was first detected at early neurula stage in the antero-ventral mesendoderm (**Figures 3Q,R**). Later on, at the mid-late neurula

**FIGURE 1 | Phylogenetic analysis of *B. lanceolatum* Fox genes.**

Unrooted 50 majority-rule consensus Bayesian inference tree based on the amino acid sequences of the forkhead domain. Posterior probabilities are shown at each node. The different paralogy groups are colored in pink or light

blue boxes. Divergent sequences appeared outside these boxes. Only one amphioxus Fox gene, named Fox1 (Yu et al., 2008a), that probably originated by a specific duplication and fast evolutionary rate in cephalochordates,

(Continued)

FIGURE 1 | Continued

localizes outside these paralogy groups. Abbreviations: Dm, *Drosophila melanogaster*; Mm, *Mus musculus*; Dr, *Danio rerio*; Ci, *Ciona intestinalis*; Sp, *Strongylocentrotus purpuratus*; Sk, *Saccoglossus kowalevskii*; Nv, *Nematostella vectensis*; Bf, *Branchiostoma floridae*; Bl, *Branchiostoma lanceolatum*. Red stars indicate Bl sequences. Scale bar represents 0.4 amino acid substitution per site.

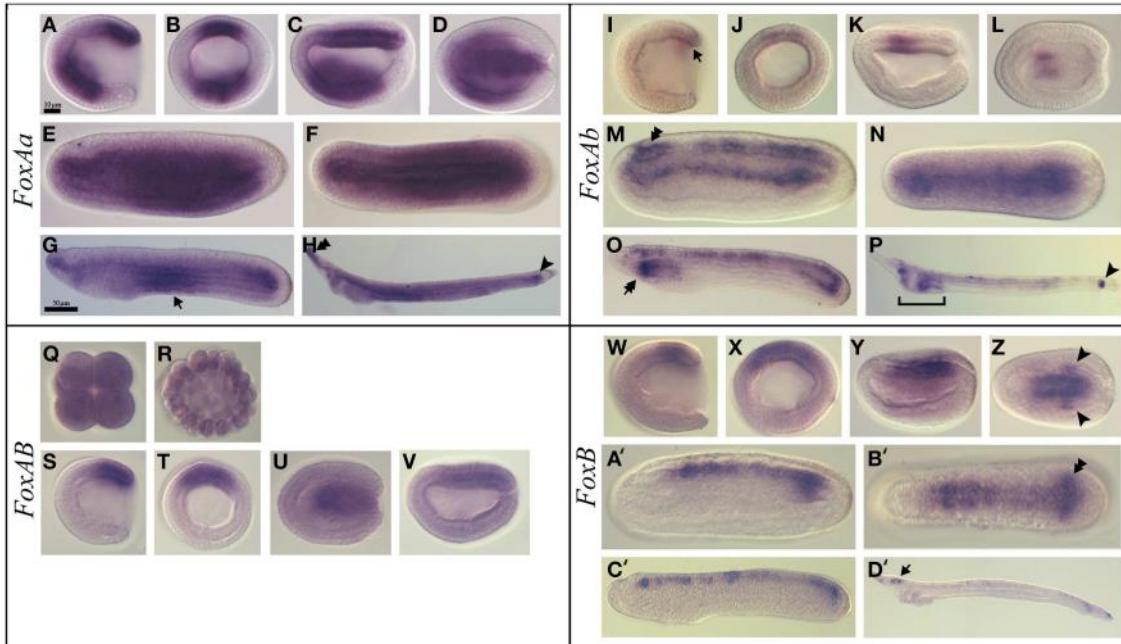


FIGURE 2 | Expression of *B. lanceolatum* *FoxAa*, *FoxAb*, *FoxAB*, and *FoxB*. In all the panels except (B, J, Q, R, T, X) anterior is to the left. In lateral and blastoporal views dorsal is to the top. *FoxAa* expression pattern (A–H). Gastrula lateral (A) and blasporal (B) views. Late gastrula lateral (C) and dorsal (D) views. Mid-late neurula lateral (E) and dorsal (F) views. In the late neurula lateral view (G) arrow marks the endodermal expression in the middle region. In the larva stage lateral view (H), the double arrowhead indicates the expression in the anterior tip of the notochord and the arrowhead marks the expression in the tailbud. *FoxAb* expression pattern (I–P). In the gastrula lateral (I) and blastoporal (J) views the arrow indicates the expression in the mesendodermal part of the dorsal blastoporal lip. Late gastrula lateral (K) and dorsal (L) views. In the mid-late neurula lateral (M) and dorsal (N) views the double arrowhead marks the expression in the

cerebral vesicle. In the late neurula lateral view (O), the double arrow marks the expression in the most anterior part of the pharynx. In larva lateral view (P) the arrowhead indicates the expression in the tailbud. *FoxAB* expression pattern (Q–V). Eight-cell stage (Q). Blastula stage (R). Gastrula lateral (S) and blasporal (T) views. Late gastrula lateral (U) and dorsal (V) views. *FoxB* expression pattern (W–D'). Gastrula lateral (W) and blastoporal (X) views. Early neurula dorsal (Y). In the early neurula dorsal (Z) view the arrowhead indicates the two expression patches in the posterior paraxial mesendoderm. Mid-late neurula lateral (A') and dorsal (B') views. The double arrowhead marks the expression in the newly formed somites. Late neurula lateral view (C'). In larva lateral view (D') the arrow indicates the expression in the cerebral vesicle. Scale bar: 10 µm (A–F), (I–N), (Q–V), (W–B'), and 50 µm (G,H), (O,P), (C',D').

stage, *FoxEa* transcripts were detected ventrally in the endoderm with a higher expression level on the right side of the pharynx (**Figures 3S,T**), and a slight expression domain in the posterior gut was also visible. At the late neurula stage, *FoxEa* transcripts remained ventrally in the pharyngeal endoderm on the right side (**Figure 3U**). Finally, at the larva stage, transcripts were detected in the club-shaped gland (**Figure 3V** and Figure S1E).

FoxG

FoxG expression was first observed at the neurula stage in the anterior region of the first somites (**Figures 3W,Y**). At the late neurula stage, *FoxG* was expressed in the anterior ventral region of the three most anterior somites (**Figures 3X,Z**). Later on, in late neurula before the mouth opens, a neural expression appeared in some individual neurons within the neural tube,

while the expression observed in the first somites disappeared (**Figure 3A'**). This expression persisted in the larva stage embryos in which *FoxG* was also detected in some neurons of the cerebral vesicle (**Figure 3B'** and Figure S1F).

FoxJ1

FoxJ1 showed a dynamic expression pattern. Expression began during gastrulation and was detected in the ectoderm except the ectoderm around the blastopore (**Figures 4A,B**). Later on, at the late gastrula stage, this expression pattern persisted in the ectoderm that give rise to the epidermis (**Figures 4C,D**). At the mid-late neurula stage, we detected transcripts in the neural tube while the expression in the epidermis was completely lost (**Figures 4E,F**). This neural tube expression was no more observed in late neurula stage embryos before the mouth

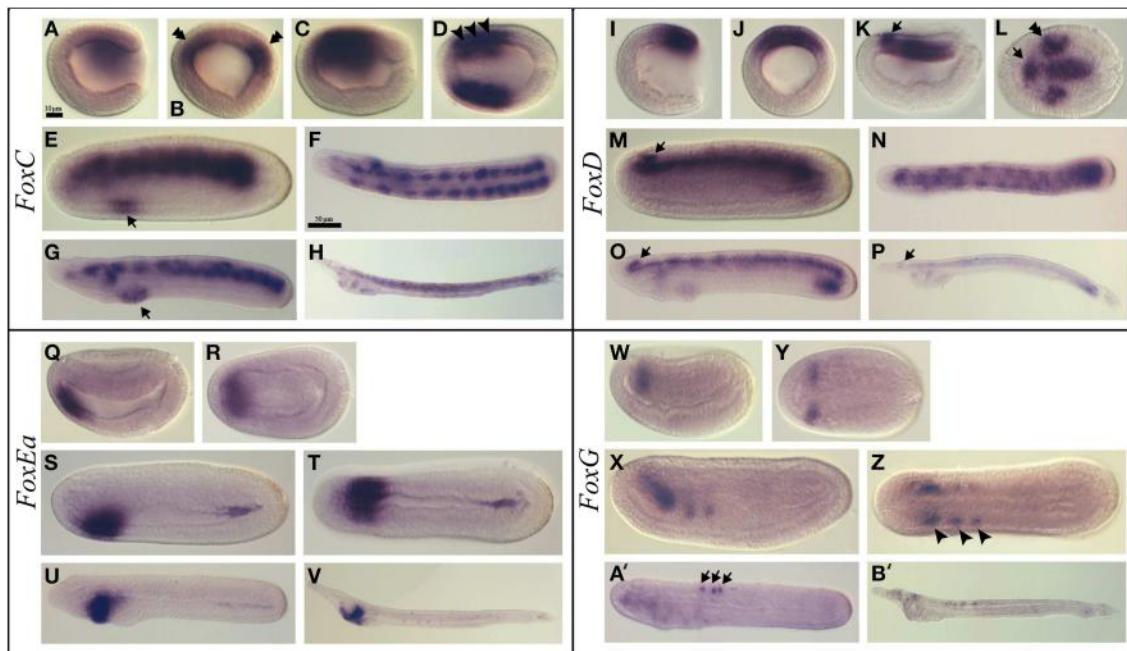


FIGURE 3 | Expression of *B. lanceolatum* FoxC, FoxD, FoxEa, and FoxG. In all the panels except (B,J), anterior is to the left. In lateral and blastoporal views dorsal is to the top. FoxC expression pattern (A–H). Gastrula lateral (A) and blastoporal (B) views. The double arrowhead indicates the expression in the paraxial mesoderm. Late gastrula lateral (C) and dorsal (D) views. The arrowheads marks the region that will give rise to the three most anterior somites. In mid-late neurula lateral view (E) the arrow indicates a new expression domain in the anterior endoderm. Late neurula dorsal (F) and lateral (G) views. The arrow marks the expression domain in the pharynx. Larva lateral view (H). FoxD expression pattern (I–P). Gastrula lateral (I) and blasporal (J) views. Late gastrula lateral (K) and dorsal (L) views. The arrow indicates the expression in the anterior region of the neural

plate and the double arrowhead marks the expression in the paraxial dorsal mesendoderm. Mid-late neurula lateral view (M). Late neurula dorsal (N) and lateral (O) views. Larva lateral view (P). In (M, O, P) the arrows indicate the expression domain in the cerebral vesicle. FoxE expression pattern (Q–V). Early neurula lateral (Q) and dorsal (R) views. Mid-late neurula lateral (S) and dorsal (T) views. Late neurula lateral view (U). Larva lateral view (V). FoxG expression pattern (W–B'). Early neurula lateral (W) and dorsal (Y) views. Mid-late neurula lateral (X) and dorsal (Z) views. The arrowhead indicates the expression in the three most anterior somites. In the late neurula stage lateral view (A') the arrows mark the neurons within the neural tube. Larva stage lateral view (B'). Scale bar: 10 μm (A–E), (I–L), (Q–T), (W–Z), and 50 μm (F–H), (N–P), (U,V), (A',B').

opens (data not show), however at the larva stage we observed expression at the anterior tip of the embryo and in the pharynx at the level of the preoral pit and of the first gill slit (Figure 4G and Figure S1G).

FoxK

FoxK was ubiquitously expressed from the eight-cell stage to the blastula stage (Figures S2A,B). At the gastrula stage, the expression became restricted to the mesendoderm (Figures S2C,D), and by the late gastrula stage transcripts were detected mostly in the dorsal mesoderm (Figures S2E,F). At the mid-late neurula stage, we detected a stronger expression in the most anterior region of the embryo (Figures S2G,H). Transcripts were then detected in the whole embryo at the late neurula stage with a stronger expression in the anterior tip (Figures S2I,J). Finally, at the larva stage, we observed a ubiquitous expression with a higher level at the anterior tip and in the pharynx (Figure S2K).

FoxM

FoxM transcripts were detected ubiquitously during the whole embryonic development, from the eight-cell stage until the mid-late neurula stage except in the epidermis (Figures S2L–S). Later

on, at late neurula stage, *FoxM* expression could not be detected anymore by *in situ* hybridization (Figure S2T).

FoxN1/4a

Ubiquitous *FoxN1/4a* expression was detected from the eight-cell stage until the blastula stage (Figures 4H,I). At the gastrula stage, a signal was detected in the anterior ectoderm (Figures 4J,K). Later on, at the early neurula stage, we observed transcripts in the anterior endoderm as well as in the axial central mesoderm (Figures 4L,M). At the mid-late neurula stage, we detected three major expression domains: one anterior, at the level of the cerebral vesicle, a second one in the anterior ventral endoderm and a third one in the posterior mesoderm (Figure 4N). At the late neurula stage before the mouth opens, we observed expression in the anterior and posterior endoderm (Figure 4O). Finally, at the larva stage, we detected expression in the posterior region of the gut and in the anus (Figure 4P).

FoxN2/3

Ubiquitous expression of *FoxN2/3* was observed from the eight-cell stage (Figure 4Q) to the blastula stage (Figure 4R). Then, at the gastrula stage, the expression was restricted to the

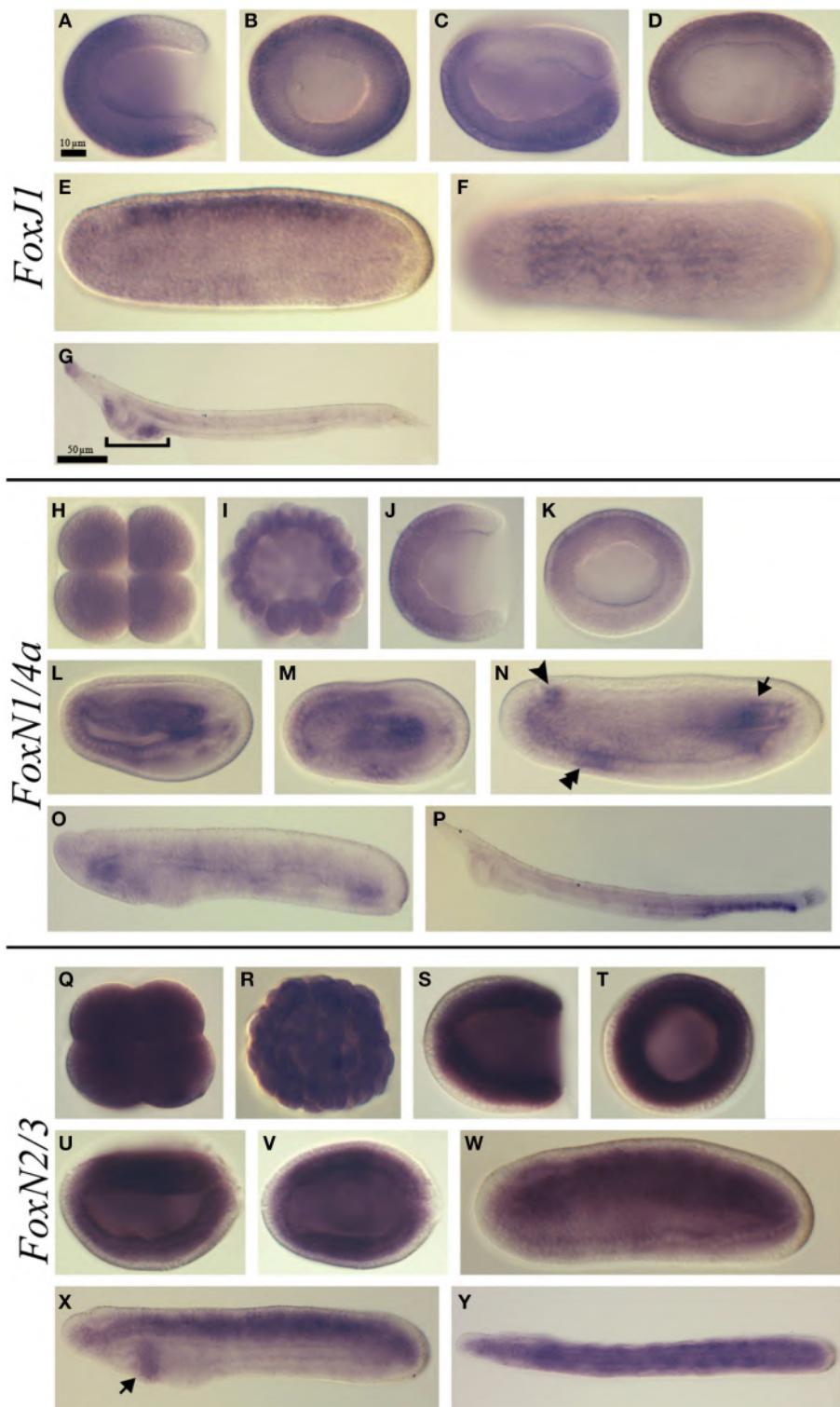


FIGURE 4 | Expression of *B. lanceolatum* *FoxJ1*, *FoxN1/4a*, and *FoxN2/3*. In all the panels except (B, H, I, K, Q, R, T) anterior is to the left. In lateral and blastoporal views dorsal is to the top. *FoxJ1* expression pattern (A–G). Gastrula lateral (A) and blasporal (B) views. Late gastrula lateral (C) and dorsal (D) views. Mid-late neurula lateral (E) and dorsal (F) views. In the larva lateral view (G) the bracket indicates the pharyngeal region. *FoxN1/4a*

expression pattern (H–P). Eight-cell stage (H). Blastula stage (I). Gastrula lateral (J) and blastoporal views (K). Early neurula lateral (L) and dorsal (M) views. In the mid-late neurula lateral view (N), the arrowhead, double arrowhead and arrow mark the three main expression domains: at the level of the cerebral vesicle, in the anterior ventral endoderm and in the posterior

(Continued)

FIGURE 4 | Continued

mesoderm, respectively. Late neurula stage lateral view (**O**). Larva stage lateral view (**P**). *FoxN2/3* expression pattern (**Q–Y**). Eight-cell stage (**Q**). Blastula stage (**R**). Gastrula lateral (**S**) and blasporal (**T**) views. Late gastrula

lateral (**U**) and dorsal (**V**) views. Mid-late neurula lateral view (**W**). Late neurula lateral (**X**) and dorsal (**Y**) views. The arrow in (**X**) indicates the expression domain in the pharyngeal endoderm. Scale bar: 10 μm (**A–F**), (**H–N**), (**Q–W**), and 50 μm (**G**), (**O–P**), (**X,Y**).

mesendoderm (**Figures 4S,T**). At the late gastrula stage, the expression remained strong in the mesendoderm but started to become lower in the ventral part (**Figures 4U,V**). By the mid-late neurula stage, *FoxN2/3* transcripts were detected in the mesoderm and in the neural tube (**Figure 4W**). At the late neurula stage, before the mouth opens, the expression was mainly detected in the paraxial mesoderm (somites) and in the notochord. A new expression domain also appeared at this stage in the pharyngeal endoderm (**Figures 4X,Y**). At the larva stage, we did not detect any specific signal using *in situ* hybridization.

Discussion

Fox Genes Expression in Cephalochordate Species

The complete or partial embryonic expression patterns of *FoxAa*, *FoxAb*, *FoxB*, *FoxC*, *FoxD*, *FoxEa*, *FoxG*, and *FoxN1/4a* were previously described in *B. floridae* and/or *B. belcheri* (Shimeld, 1997; Terazawa and Satoh, 1997; Toresson et al., 1998; Mazet and Shimeld, 2002; Yu et al., 2002a,b; Mazet et al., 2006; Bajoghli et al., 2009). These genes overwhelmingly show a similar embryonic expression to what we observed in *B. lanceolatum*, as we have previously noticed for other important developmental genes (Somorjai et al., 2008). However, our work brings some new information.

First, in contrast to what has been described in *B. floridae*, we showed that *FoxAa* and *FoxAb* have different expression patterns. Indeed, in *B. floridae*, *FoxAb* *in situ* hybridization data showed that it has a similar expression to *FoxAa* at early stages whereas expression was no more detected after the eight somites stage (Shimeld, 1997). Here we showed that although both genes were expressed in the mesendodermal part of the dorsal blastoporal lip at the gastrula stage, the overall expression patterns are consistently different between the two genes and we observed a restricted expression of *FoxAb* from the gastrula to the larva stage. These discrepancies might be explained by the fact that the level of expression of *FoxAb* is very low. Indeed, staining of embryos hybridized to *FoxAb* took very long suggesting a low expression level. Thus, the staining time used in *B. floridae* might have been too short to detect expression in late stage embryos. Moreover, the expression we observed for *FoxAa* in *B. lanceolatum* is different from what was observed in *B. floridae* but similar to what has been described in *B. belcheri* (Terazawa and Satoh, 1997). Indeed, as in *B. belcheri*, *FoxAa* was not expressed in the central nervous system of *B. lanceolatum*. On the other hand, *FoxAb* showed a very specific expression in the ventral part of the neural tube in neurula stage embryos, which has been proposed to be homologous to the vertebrate floor plate. Vertebrates have three FoxA group paralogous genes that are expressed in the organizer, the notochord, the floor plate and the endoderm (Friedman and Kaestner, 2006). In *Ciona* (Di Gregorio

et al., 2001), *Ci-fkh* is also expressed in the notochord, the floor plate and the endoderm. The data we obtained in *B. lanceolatum* suggest that the expression of *FoxA* in the chordate ancestor was similar to what is observed in tunicates and that independent sub-functionalizations occurred in cephalochordates after specific gene duplication and in vertebrates after the two rounds of whole genome duplications.

Concerning *FoxB*, expression in *B. floridae* was first detected in neurulae with five somites (Mazet and Shimeld, 2002). Here we showed that in *B. lanceolatum* *FoxB* expression could be observed in gastrula embryos in the dorsal posterior mesendoderm and ectoderm. Then, in neurulae, we detected expression in the neural plate similar to *B. floridae*, as well as an expression in the most posterior somites that was not previously described. This expression in the neural plate/neural tube and in the lastly formed somites persisted until the late neurula stage. Interestingly, in amphioxus three different somitic populations have been described (Bertrand et al., 2011). The first, most anterior, population forms under the control of the FGF signal and the two posterior populations forms independently of the FGF signal. Several genes are expressed specifically in these three somitic populations but only one gene, *Mox*, (Minguillon and Garcia-Fernandez, 2002) is expressed in the second and third populations. The present data suggest that *FoxB* also plays a role in the formation of these somitic population since it is also expressed in the two most-posterior somitic populations.

In *B. floridae*, *FoxC* has been described as being firstly expressed in the mesoderm of neurulae but its expression was described only in one developmental stage (Mazet et al., 2006). Here we showed that expression starts much earlier, at the gastrula stage, in the dorsal paraxial mesendoderm, the presumptive somitic mesoderm territory. Expression persisted in the paraxial mesoderm/somites until the larva stage, and at the late neurula stage we started to observe expression in the club-shaped gland anlagen and at the place where the first gill slit opens. These data suggest a major ancestral role of *FoxC* during somitogenesis which would have been conserved in vertebrates (Kume et al., 2001; Wilm et al., 2004; Wotton et al., 2008) and lost in tunicates in which *FoxC* is expressed in neural and palp cells (Imai et al., 2006).

FoxD and *FoxEa* expression in *B. lanceolatum* was very similar to previous descriptions in *B. floridae* (Yu et al., 2002a,b). However we noticed expression in some specific regions of the pharynx in late neurulae and larvae for *FoxD*, and a transient expression in mid-late and late neurula stage embryos in the posterior endoderm for *FoxEa* that were not described in the Caribbean species.

FoxG, previously known as *Brain Factor 1 (BF-1)*, was described in *B. floridae* as a gene that is ventrally expressed in the cerebral vesicle and in the anterior-most portion of the first somite pair (Toresson et al., 1998). Our results showed

a conserved expression pattern in the cerebral vesicle area in *B. lanceolatum*. However, mesoderm expression is not only limited to the first somite pair but the first three somite pairs exhibit the same pattern at the neurula stage suggesting that this gene might play a role during anterior somitogenesis. This result highlights the functional differences between the formation of the anterior somites which is under the control of the FGF signaling pathway and the formation of the most posterior somites which is not FGF-dependent (Bertrand et al., 2011). Moreover, expression is localized in the ventral part of these three most anterior somites which will give rise to the perivisceral coelom, suggesting a function of *FoxG* in the establishment of the somitic compartments.

FoxJ1 and the Formation of Motile Cilia

FoxJ1 orthologs were identified in many eumetazoans as well as in sponges (Larroux et al., 2006) and choanoflagellates (King et al., 2008). In vertebrates, *FoxJ1* plays an essential role in the generation of motile cilia and in mediating Left/Right asymmetry (Chen et al., 1998; Brody et al., 2000; Yu et al., 2008b). It has also recently been shown that misexpression of *FoxJ1* from placozoans, echinoderms and platyhelminthes in zebrafish embryos induces the expression of ciliary genes, whereas the inactivation of *FoxJ1* in the flatworm *Schmidtea mediterranea* impairs the normal differentiation of motile cilia, suggesting a conserved function in metazoans (Vij et al., 2012). This conserved function is also supported by the embryonic expression of *FoxJ1* in different phyla (Choi et al., 2006; Tu et al., 2006; Fritzenwanker et al., 2014). In *B. lanceolatum*, we showed that *FoxJ1* is first expressed in the ectoderm of the gastrulae, excluding the blastoporal region and the presumptive neural plate, at the time at which motile cilia start to grow. Then, in neurulae, expression was lost in the epidermis and appeared in the closed neural tube. At the larva stage, expression was restricted to the anterior tip of the animal and to the ciliated preoral pit and first gill slit. This expression pattern suggests that in amphioxus *FoxJ1* might also play a role in the formation of motile cilia. However, other cells, like the epithelial gut cells, also harbor motile cilia and do not express *FoxJ1*, suggesting that other genes might also be implicated in ciliogenesis in these embryonic structures.

FoxAB

In *B. lanceolatum*, *FoxAB* was transiently expressed in the organizer at the gastrula stage and in the presumptive notochord later on. No expression could be detected in mid-neurulae or larvae. *FoxAB* family genes were described in hemichordates (Fritzenwanker et al., 2014), sea urchin (Tu et al., 2006) and cnidarians and are absent in vertebrates and tunicates, the two other chordate clades (Yu et al., 2008a). In the hemichordate *Saccoglossus kowalevskii*, *FoxAB* is expressed in the ectoderm and the mouth perforates through the ring expressing this gene in the ventral side (Fritzenwanker et al., 2014). In bryozoans, *FoxAB* also shows an ectodermal expression (Fuchs et al., 2011). Therefore, it is still difficult to propose any scenario for the evolution of the function of *FoxAB* family genes in bilaterians. *FoxAB* could have been recruited for the patterning of the

notochord field in the ancestor of chordates, but the absence of genes of this family in tunicates and vertebrates make this hypothesis unlikely.

FoxK and FoxM Ubiquitous Expression

We detected a ubiquitous expression of *FoxK* starting at the eight-cell stage until the larva stage. In other bilaterians data are scarce. In vertebrates, there are two paralogs in the *FoxK* family, *FoxK1* and *FoxK2*. In mouse, the study of the function of *FoxK1* during embryonic development was undertaken showing that the gene is involved in myogenic differentiation (Bassel-Duby et al., 1994). In *Ciona intestinalis* (Imai et al., 2004) as in the hemichordate *S. kowalevskii* (Fritzenwanker et al., 2014), the expression of *FoxK* is quite ubiquitous as observed for *B. lanceolatum*. Finally, studies in *Drosophila* have shown that *FoxK* is involved in the differentiation of midgut in the fly embryo (Casas-Tinto et al., 2008). Altogether these data do not allow us to infer any putative ancestral function for *FoxK* family genes and further studies are required in different animal phyla.

FoxM expression is also ubiquitous in *B. lanceolatum* and was first detected as early as the eight-cell stage. Then the expression level continuously decreased while development proceeds and became undetectable by *in situ* hybridization at the late neurula stage. In *Xenopus*, *FoxM1* is maternally expressed and transcripts are thereafter detected in the neuroectoderm (Pohl et al., 2005). Moreover this gene has been shown to be important for early neuronal differentiation (Ueno et al., 2008). In mouse, *FoxM1* is expressed in dividing cells and knock-out animals exhibit embryonic lethal phenotype due to many malformations affecting different organs such as the liver, the heart, the lung, or the vasculature (Kalin et al., 2011). As for *FoxK*, the data available up to now do not give us any indication on the putative ancestral function of genes belonging to the *FoxM* family.

FoxN1/4a and FoxN2/3 Expression

In all vertebrates studied so far, *FoxN1* plays an essential role in thymus development (Ma et al., 2012; Neves et al., 2012; Lee et al., 2013; Romano et al., 2013). Moreover, in mammals, *FoxN1* is essential for hair formation whereas it is also expressed in chick during feather development (Darnell et al., 2014). Although mammal and fish *FoxN1s* are able to activate the expression of hair keratin genes, *FoxN1/4* from amphioxus is not because its N-terminal region of the forkhead domain is different compared with vertebrates (Schlake et al., 2000). On the other hand, *FoxN4* is expressed in the nervous system, including retina, during vertebrate development (Danilova et al., 2004; Kelly et al., 2007; Boije et al., 2013). Outside vertebrates, embryonic expression has been described in *S. kowalevskii* (Fritzenwanker et al., 2014) and in a single developmental stage of *B. floridae* (Bajoghli et al., 2009). In the hemichordate, expression of *FoxN1/4* is ubiquitous during early development and is thereafter observed in the ectoderm. In *B. lanceolatum*, the expression of *FoxN1/4a* was very dynamic with a maternal ubiquitous expression followed by restricted expression in the ectoderm at the gastrula stage, in the endoderm and axial mesoderm in neurulae, in the cerebral vesicle, the pharynx and the posterior somites later on, and, finally, in

the posterior gut of the larvae. These data suggest that *FoxN1* and *FoxN4* probably acquired new functions in vertebrates, and analysis of the expression of *FoxN1/4* family genes in tunicates will be needed to better understand this point. Interestingly, the gut of amphioxus larva and adult is considered as a major organ for immunity and *FoxN1/4a* might, as vertebrates *FoxN1*, play a role in the control of immune system function in amphioxus. However, further functional studies are required to test this hypothesis.

In vertebrates, *FoxN3* is important for craniofacial and eye development (Schuff et al., 2007; Samaan et al., 2010; Schmidt et al., 2011). In *Xenopus*, *FoxN3* is expressed in neural crest and eye field whereas *FoxN2* is expressed early in the eye field and then in branchial arches, retina and vagal ganglion (Schuff et al., 2006). In mouse, *FoxN2* is expressed in craniofacial, limb, nervous system and somitic tissues (Triboli et al., 2002). In *Ciona intestinalis*, expression of *FoxN2/3* is quite ubiquitous during early development and becomes more intense in the sensory vesicle, the mesenchyme, the notochord and the palps after gastrulation (Imai et al., 2004). In sea urchin *FoxN2/3* is expressed in the non-skeletogenic mesoderm and, later on, in the endoderm and it has been shown that *FoxN2/3* function is important for ingression and for the expression of genes coding for proteins of the skeletal matrix (Rho and Mcclay, 2011). Here, we show that *FoxN2/3* in amphioxus was ubiquitously expressed at early stages. Then, at the gastrula stage, its expression was restricted to the endomesoderm and later on we observed a specific expression in the somites. Altogether, this suggests a conserved role of *FoxN2/3* in the development of mesoderm in deuterostomes, although genes of this family seem to have acquired specific functions in each chordate lineage.

References

- Bajoghli, B., Aghaallaei, N., Hess, I., Rode, I., Netuschil, N., Tay, B. H., et al. (2009). Evolution of genetic networks underlying the emergence of thymopoiesis in vertebrates. *Cell* 138, 186–197. doi: 10.1016/j.cell.2009.04.017
- Bassel-Duby, R., Hernandez, M. D., Yang, Q., Rochelle, J. M., Seldin, M. F., and Williams, R. S. (1994). Myocyte nuclear factor, a novel winged-helix transcription factor under both developmental and neural regulation in striated myocytes. *Mol. Cell Biol.* 14, 4596–4605.
- Benayoun, B. A., Caburet, S., and Veitia, R. A. (2011). Forkhead transcription factors: key players in health and disease. *Trends Genet.* 27, 224–232. doi: 10.1016/j.tig.2011.03.003
- Bertrand, S., Camasses, A., Somorjai, I., Belgacem, M. R., Chabrol, O., Escande, M. L., et al. (2011). Amphioxus FGF signaling predicts the acquisition of vertebrate morphological traits. *Proc. Natl. Acad. Sci. U.S.A.* 108, 9160–9165. doi: 10.1073/pnas.1014235108
- Bertrand, S., and Escrivá, H. (2011). Evolutionary crossroads in developmental biology: amphioxus. *Development* 138, 4819–4830. doi: 10.1242/dev.066720
- Bertrand, S., and Escrivá, H. (2014). “Chordates: The acquisition of an axial backbone,” *The Tree of Life*, eds P. Vargas and R. Zardoya (Sunderland, MA: Sinauer Associates, Inc), 460–468.
- Boije, H., Shirazi Fard, S., Ring, H., and Hallbook, F. (2013). Forkheadbox N4 (*FoxN4*) triggers context-dependent differentiation in the developing chick retina and neural tube. *Differentiation* 85, 11–19. doi: 10.1016/j.diff.2012.12.002
- Brody, S. L., Yan, X. H., Wuerffel, M. K., Song, S. K., and Shapiro, S. D. (2000). Ciliogenesis and left-right axis defects in forkhead factor HFH-4-null mice. *Am. J. Respir. Cell Mol. Biol.* 23, 45–51. doi: 10.1165/ajrcmb.23.1.4070
- Carlsson, P., and Mahlapuu, M. (2002). Forkhead transcription factors: key players in development and metabolism. *Dev. Biol.* 250, 1–23. doi: 10.1006/dbio.2002.0780
- Casas-Tinto, S., Gomez-Velazquez, M., Granadino, B., and Fernandez-Funez, P. (2008). FoxK mediates TGF-beta signalling during midgut differentiation in flies. *J. Cell Biol.* 183, 1049–1060. doi: 10.1083/jcb.200808149
- Chen, J., Knowles, H. J., Hebert, J. L., and Hackett, B. P. (1998). Mutation of the mouse hepatocyte nuclear factor/forkhead homologue 4 gene results in an absence of cilia and random left-right asymmetry. *J. Clin. Invest.* 102, 1077–1082. doi: 10.1172/JCI4786
- Choi, V. M., Harland, R. M., and Khokha, M. K. (2006). Developmental expression of FoxJ1.2, FoxJ2, and FoxQ1 in *Xenopus tropicalis*. *Gene Expr. Patterns* 6, 443–447. doi: 10.1016/j.modgep.2005.11.007
- Clark, K. L., Halay, E. D., Lai, E., and Burley, S. K. (1993). Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* 364, 412–420. doi: 10.1038/364412a0
- Danilova, N., Visel, A., Willett, C. E., and Steiner, L. A. (2004). Expression of the winged helix/forkhead gene, foxn4, during zebrafish development. *Brain Res. Dev. Brain Res.* 153, 115–119. doi: 10.1016/j.devbrainres.2004.05.014
- Darnell, D. K., Zhang, L. S., Hannenhalli, S., and Yaklichkin, S. Y. (2014). Developmental expression of chicken FOXN1 and putative target genes during feather development. *Int. J. Dev. Biol.* 58, 57–64. doi: 10.1387/ijdb.130023sy
- Di Gregorio, A., Corbo, J. C., and Levine, M. (2001). The regulation of forkhead/HNF-3beta expression in the *Ciona* embryo. *Dev. Biol.* 229, 31–43. doi: 10.1006/dbio.2000.9964

Conclusions

Analyzing the expression of Fox genes in the Mediterranean amphioxus, *B. lanceolatum* showed us several points. First, as previously described for other gene families (Somorjai et al., 2008), the expression of orthologous genes in different amphioxus species shows a high degree of stasis. However, differences may be found that can easily be explained by variation in experimental sensitivity. And, second, the comparative analyzes of the expression of amphioxus Fox genes with other metazoans and particularly chordates have shown a high degree of conservation for some genes (e.g., *FoxC*, *FoxD*), but also divergent patterns in others (e.g., *FoxM*, *FoxN1/4a*). This indicates that Fox genes were necessary for essential functions in metazoans but they were also instrumental for the evolution of new functions. Further studies in amphioxus and other metazoans, and particularly functional studies, will be extremely important in the future to establish the complete picture of Fox genes expression and function and their role in the evolution of animals.

Acknowledgments

DA holds a fellowship from CONICYT “Becas Chile.” Part of this study was supported by the EXOMOD grant from the CNRS.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fevo.2015.00080>

- Friedman, J. R., and Kaestner, K. H. (2006). The Foxa family of transcription factors in development and metabolism. *Cell Mol. Life Sci.* 63, 2317–2328. doi: 10.1007/s00018-006-6095-6
- Fritzenwanker, J. H., Gerhart, J., Freeman, R. M. Jr., and Lowe, C. J. (2014). The Fox/Forkhead transcription factor family of the hemichordate *Saccoglossus kowalevskii*. *EvoDevo* 5:17. doi: 10.1186/2041-9139-5-17
- Fuchs, J., Martindale, M. Q., and Hejnol, A. (2011). Gene expression in bryozoan larvae suggest a fundamental importance of pre-patterned blastemic cells in the bryozoan life-cycle. *EvoDevo* 2:13. doi: 10.1186/2041-9139-2-13
- Fuentes, M., Benito, E., Bertrand, S., Paris, M., Mignardot, A., Godoy, L., et al. (2007). Insights into spawning behavior and development of the European amphioxus (*Branchiostoma lanceolatum*). *J. Exp. Zool. B Mol. Dev. Evol.* 308, 484–493. doi: 10.1002/jez.b.21179
- Fuentes, M., Schubert, M., Dalfo, D., Candiani, S., Benito, E., Gardenyes, J., et al. (2004). Preliminary observations on the spawning conditions of the European amphioxus (*Branchiostoma lanceolatum*) in captivity. *J. Exp. Zool. B Mol. Dev. Evol.* 302, 384–391. doi: 10.1002/jez.b.20025
- Imai, K. S., Hino, K., Yagi, K., Satoh, N., and Satou, Y. (2004). Gene expression profiles of transcription factors and signaling molecules in the ascidian embryo: towards a comprehensive understanding of gene networks. *Development* 131, 4047–4058. doi: 10.1242/dev.01270
- Imai, K. S., Levine, M., Satoh, N., and Satou, Y. (2006). Regulatory blueprint for a chordate embryo. *Science* 312, 1183–1187. doi: 10.1126/science.1123404
- Kalin, T. V., Ustiyan, V., and Kalinichenko, V. V. (2011). Multiple faces of FoxM1 transcription factor: lessons from transgenic mouse models. *Cell Cycle* 10, 396–405. doi: 10.4161/cc.10.3.14709
- Kelly, L. E., Nekkalapudi, S., and El-Hodiri, H. M. (2007). Expression of the forkhead transcription factor FoxN4 in progenitor cells in the developing *Xenopus laevis* retina and brain. *Gene Expr. Patterns* 7, 233–238. doi: 10.1016/j.modgep.2006.10.003
- King, N., Westbrook, M. J., Young, S. L., Kuo, A., Abedin, M., Chapman, J., et al. (2008). The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451, 783–788. doi: 10.1038/nature06617
- Kume, T., Jiang, H., Topczewska, J. M., and Hogan, B. L. (2001). The murine winged helix transcription factors, Foxc1 and Foxc2, are both required for cardiovascular development and somitogenesis. *Genes Dev.* 15, 2470–2482. doi: 10.1101/gad.907301
- Larroux, C., Fahey, B., Liubicich, D., Hinman, V. F., Gauthier, M., Gongora, M., et al. (2006). Developmental expression of transcription factor genes in a demosponge: insights into the origin of metazoan multicellularity. *Evol. Dev.* 8, 150–173. doi: 10.1111/j.1525-142X.2006.00086.x
- Larroux, C., Luke, G. N., Koopman, P., Rokhsar, D. S., Shimeld, S. M., and Degnan, B. M. (2008). Genesis and expansion of metazoan transcription factor gene classes. *Mol. Biol. Evol.* 25, 980–996. doi: 10.1093/molbev/msn047
- Lee, Y. H., Williams, A., Hong, C. S., You, Y., Senoo, M., and Saint-Jeannet, J. P. (2013). Early development of the thymus in *Xenopus laevis*. *Dev. Dyn.* 242, 164–178. doi: 10.1002/dvdy.23905
- Ma, D., Wang, L., Wang, S., Gao, Y., Wei, Y., and Liu, F. (2012). Foxn1 maintains thymic epithelial cells to support T-cell development via mcm2 in zebrafish. *Proc. Natl. Acad. Sci. U.S.A.* 109, 21040–21045. doi: 10.1073/pnas.1217021110
- Mazet, F., Amemiya, C. T., and Shimeld, S. M. (2006). An ancient Fox gene cluster in bilaterian animals. *Curr. Biol.* 16, R314–R316. doi: 10.1016/j.cub.2006.03.088
- Mazet, F., and Shimeld, S. M. (2002). The evolution of chordate neural segmentation. *Dev. Biol.* 251, 258–270. doi: 10.1006/dbio.2002.0831
- Mazet, F., Yu, J. K., Liberles, D. A., Holland, L. Z., and Shimeld, S. M. (2003). Phylogenetic relationships of the Fox (Forkhead) gene family in the Bilateria. *Gene* 316, 79–89. doi: 10.1016/S0378-1119(03)00741-8
- Miller, M. A., Schwartz, T., Pickett, B. E., He, S., Klem, E. B., Scheuermann, R. H., et al. (2015). A RESTful API for access to phylogenetic tools via the CIPRES science gateway. *Evol. Bioinform. Online* 11, 43–48. doi: 10.4137/EBO.S21501
- Minguión, C., and García-Fernandez, J. (2002). The single amphioxus Mox gene: insights into the functional evolution of Mox genes, somites, and the asymmetry of amphioxus somitogenesis. *Dev. Biol.* 246, 455–465. doi: 10.1006/dbio.2002.0660
- Neves, H., Dupin, E., Parreira, L., and Le Douarin, N. M. (2012). Modulation of Bmp4 signalling in the epithelial-mesenchymal interactions that take place in early thymus and parathyroid development in avian embryos. *Dev. Biol.* 361, 208–219. doi: 10.1016/j.ydbio.2011.10.022
- Onimaru, K., Shoguchi, E., Kuratani, S., and Tanaka, M. (2011). Development and evolution of the lateral plate mesoderm: comparative analysis of amphioxus and lamprey with implications for the acquisition of paired fins. *Dev. Biol.* 359, 124–136. doi: 10.1016/j.ydbio.2011.08.003
- Oulion, S., Bertrand, S., Belgacem, M. R., Le Petillon, Y., and Escrivá, H. (2012). Sequencing and analysis of the Mediterranean amphioxus (*Branchiostoma lanceolatum*) transcriptome. *PLoS ONE* 7:e36554. doi: 10.1371/journal.pone.0036554
- Pohl, B. S., Rossner, A., and Knochel, W. (2005). The Fox gene family in *Xenopus laevis*: FoxI2, FoxM1 and FoxP1 in early development. *Int. J. Dev. Biol.* 49, 53–58. doi: 10.1387/ijdb.051977bp
- Rho, H. K., and McClay, D. R. (2011). The control of foxN2/3 expression in sea urchin embryos and its function in the skeletogenic gene regulatory network. *Development* 138, 937–945. doi: 10.1242/dev.058396
- Romano, R., Palamaro, L., Fusco, A., Giardino, G., Gallo, V., Del Vecchio, L., et al. (2013). FOXN1: a master regulator gene of thymic epithelial development program. *Front. Immunol.* 4:187. doi: 10.3389/fimmu.2013.00187
- Ronquist, F., Teslenko, M., Van Der Mark, P., Ayres, D. L., Darling, A., Hohna, S., et al. (2012). MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst. Biol.* 61, 539–542. doi: 10.1093/sysbio/sys029
- Samaan, G., Yugo, D., Rajagopalan, S., Wall, J., Donnell, R., Goldowitz, D., et al. (2010). Foxn3 is essential for craniofacial development in mice and a putative candidate involved in human congenital craniofacial defects. *Biochem. Biophys. Res. Commun.* 400, 60–65. doi: 10.1016/j.bbrc.2010.07.142
- Schlake, T., Schorpp, M., Maul-Pavicic, A., Malashenko, A. M., and Boehm, T. (2000). Forkhead/winged-helix transcription factor Whn regulates hair keratin gene expression: molecular analysis of the nude skin phenotype. *Dev. Dyn.* 217, 368–376. doi: 10.1002/(SICI)1097-0177(200004)217:4<368::AID-DVDDY>3.0.CO;2-Z
- Schmidt, J., Schuff, M., and Olsson, L. (2011). A role for FoxN3 in the development of cranial cartilages and muscles in *Xenopus laevis* (Amphibia: Anura: Pipidae) with special emphasis on the novel rostral cartilages. *J. Anat.* 218, 226–242. doi: 10.1111/j.1469-7580.2010.01315.x
- Schuff, M., Rossner, A., Donow, C., and Knochel, W. (2006). Temporal and spatial expression patterns of FoxN genes in *Xenopus laevis* embryos. *Int. J. Dev. Biol.* 50, 429–434. doi: 10.1387/ijdb.052126ms
- Schuff, M., Rossner, A., Wacker, S. A., Donow, C., Gessert, S., and Knochel, W. (2007). FoxN3 is required for craniofacial and eye development of *Xenopus laevis*. *Dev. Dyn.* 236, 226–239. doi: 10.1002/dvdy.21007
- Shimeld, S. M., Boyle, M. J., Brunet, T., Luke, G. N., and Seaver, E. C. (2010a). Clustered Fox genes in lophotrochozoans and the evolution of the bilaterian Fox gene cluster. *Dev. Biol.* 340, 234–248. doi: 10.1016/j.ydbio.2010.01.015
- Shimeld, S. M., Degnan, B., and Luke, G. N. (2010b). Evolutionary genomics of the Fox genes: origin of gene families and the ancestry of gene clusters. *Genomics* 95, 256–260. doi: 10.1016/j.ygeno.2009.08.002
- Shimeld, S. M. (1997). Characterisation of amphioxus HNF-3 genes: conserved expression in the notochord and floor plate. *Dev. Biol.* 183, 74–85. doi: 10.1006/dbio.1996.8481
- Somorjai, I., Bertrand, S., Camasses, A., Haguenuer, A., and Escrivá, H. (2008). Evidence for stasis and not genetic piracy in developmental expression patterns of *Branchiostoma lanceolatum* and *Branchiostoma floridae*, two amphioxus species that have evolved independently over the course of 200 Myr. *Dev. Genes Evol.* 218, 703–713. doi: 10.1007/s00427-008-0256-6
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: molecular evolutionary genetics analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Terazawa, K., and Satoh, N. (1997). Formation of the chordamesoderm in the amphioxus embryo: analysis with Brachyury and fork head/HNF-3 genes. *Dev. Genes Evol.* 207, 1–11. doi: 10.1007/s004270050086
- Toresson, H., Martinez-Barbera, J. P., Bardsley, A., Caubit, X., and Krauss, S. (1998). Conservation of BF-1 expression in amphioxus and zebrafish suggests evolutionary ancestry of anterior cell types that contribute to the vertebrate telencephalon. *Dev. Genes Evol.* 208, 431–439. doi: 10.1007/s004270050200

- Tribioli, C., Robledo, R. F., and Lufkin, T. (2002). The murine fork head gene Foxn2 is expressed in craniofacial, limb, CNS and somitic tissues during embryogenesis. *Mech. Dev.* 118, 161–163. doi: 10.1016/S0925-4773(02)00220-4
- Tu, Q., Brown, C. T., Davidson, E. H., and Oliveri, P. (2006). Sea urchin Forkhead gene family: phylogeny and embryonic expression. *Dev. Biol.* 300, 49–62. doi: 10.1016/j.ydbio.2006.09.031
- Tuteja, G., and Kaestner, K. H. (2007a). Forkhead transcription factors II. *Cell* 131, 192. doi: 10.1016/j.cell.2007.09.016
- Tuteja, G., and Kaestner, K. H. (2007b). SnapShot: forkhead transcription factors I. *Cell* 130, 1160. doi: 10.1016/j.cell.2007.09.005
- Ueno, H., Nakajo, N., Watanabe, M., Isoda, M., and Sagata, N. (2008). FoxM1-driven cell division is required for neuronal differentiation in early Xenopus embryos. *Development* 135, 2023–2030. doi: 10.1242/dev.019893
- Vij, S., Rink, J. C., Ho, H. K., Babu, D., Eitel, M., Narasimhan, V., et al. (2012). Evolutionarily ancient association of the FoxJ1 transcription factor with the motile ciliogenic program. *PLoS Genet* 8:e1003019. doi: 10.1371/journal.pgen.1003019
- Weigel, D., and Jackle, H. (1990). The fork head domain: a novel DNA binding motif of eukaryotic transcription factors? *Cell* 63, 455–456. doi: 10.1016/0092-8674(90)90439-L
- Wilm, B., James, R. G., Schultheiss, T. M., and Hogan, B. L. (2004). The forkhead genes, Foxc1 and Foxc2, regulate paraxial versus intermediate mesoderm cell fate. *Dev. Biol.* 271, 176–189. doi: 10.1016/j.ydbio.2004.03.034
- Wotton, K. R., Mazet, F., and Shimeld, S. M. (2008). Expression of FoxC, FoxF, FoxL1, and FoxQ1 genes in the dogfish *Scyliorhinus canicula* defines ancient and derived roles for Fox genes in vertebrate development. *Dev. Dyn.* 237, 1590–1603. doi: 10.1002/dvdy.21553
- Wotton, K. R., and Shimeld, S. M. (2006). Comparative genomics of vertebrate Fox cluster loci. *BMC Genomics* 7:271. doi: 10.1186/1471-2164-7-271
- Yu, J. K., Holland, L. Z., Jamrich, M., Blitz, I. L., and Hollan, N. D. (2002a). AmphiFoxE4, an amphioxus winged helix/forkhead gene encoding a protein closely related to vertebrate thyroid transcription factor-2: expression during pharyngeal development. *Evol. Dev.* 4, 9–15. doi: 10.1046/j.1525-142x.2002.01057.x
- Yu, J. K., Holland, N. D., and Holland, L. Z. (2002b). An amphioxus winged helix/forkhead gene, AmphiFoxD: insights into vertebrate neural crest evolution. *Dev. Dyn.* 225, 289–297. doi: 10.1002/dvdy.10173
- Yu, J. K., Holland, N. D., and Holland, L. Z. (2003). AmphiFoxQ2, a novel winged helix/forkhead gene, exclusively marks the anterior end of the amphioxus embryo. *Dev. Genes Evol.* 213, 102–105. doi: 10.1007/s00427-003-0302-3
- Yu, J. K., Mazet, F., Chen, Y. T., Huang, S. W., Jung, K. C., and Shimeld, S. M. (2008a). The Fox genes of *Branchiostoma floridae*. *Dev. Genes Evol.* 218, 629–638. doi: 10.1007/s00427-008-0229-9
- Yu, X., Ng, C. P., Habacher, H., and Roy, S. (2008b). FoxJ1 transcription factors are master regulators of the motile ciliogenic program. *Nat. Genet.* 40, 1445–1453. doi: 10.1038/ng.263

Conflict of Interest Statement: 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 © 2015 Aldea, Leon, Bertrand and Escrivá. 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) or licensor 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.



Asymmetric Distribution of *pl10* and *bruno2*, New Members of a Conserved Core of Early Germline Determinants in Cephalochordates

Simon C. Dailey^{1†}, Roser Febrero Planas^{2†}, Ariadna Rossell Espier², Jordi Garcia-Fernández¹ and Ildikó M. L. Somorjai^{1,2*}

¹ Gatty Marine Laboratory, Scottish Oceans Institute, University of St Andrews, St Andrews, UK, ² Department of Genetics, University of Barcelona, Barcelona, Spain

OPEN ACCESS

Edited by:

Sylvain Marcellini,
University of Concepcion, Chile

Reviewed by:

Brian I. Crother,
Southeastern Louisiana University,
USA

Eduardo E. Zattara,
Indiana University, USA

*Correspondence:

Ildikó M. L. Somorjai
imls@st-andrews.ac.uk

[†]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 Ecology and Evolution

Received: 01 September 2015

Accepted: 28 December 2015

Published: 22 January 2016

Citation:

Dailey SC, Febrero Planas R, Rossell Espier A, Garcia-Fernández J and Somorjai IML (2016) Asymmetric Distribution of *pl10* and *bruno2*, New Members of a Conserved Core of Early Germline Determinants in Cephalochordates. *Front. Ecol. Evol.* 3:156.
doi: 10.3389/fevo.2015.00156

Molecular fingerprinting of conserved germline and somatic “stemness” markers in different taxa have been key in defining the mechanism of germline specification (“preformation” or “epigenesis”), as well as expression domains of somatic progenitors. The distribution of molecular markers for primordial germ cells (PGCs), including *vasa*, *nanos*, and *piwil1*, as well as Vasa antibody staining, support a determinative mechanism of germline specification in the cephalochordate *Branchiostoma lanceolatum*, similarly to other amphioxus species. *pl10* and *bruno2*, but not *bruno4/6*, are also expressed in a pattern consistent with these other germline genes, adding to our repertoire of PGC markers in lancelets. Expression of *nanos*, *vasa*, and the remaining markers (*musashi*, *pufA*, *pufB*, *pumilio*, and *piwil2*) may define populations of putative somatic progenitors in the tailbud, the amphioxus posterior growth zone, or zones of proliferative activity. Finally, we also identify a novel expression domain for *musashi*, a classic neural stem cell marker, during notochord development in amphioxus. These results are discussed in the context of germline determination in other taxa, stem cell regulation, and regenerative capacity in adult amphioxus.

Keywords: amphioxus, posterior stem cells, evolution, regeneration, PGCs, preformation

INTRODUCTION

One of the key innovations coupled to the evolution of multicellularity was the ability to segregate the germline and the soma, with transcriptional repression of a somatic programme being key to maintaining the germ cell fate (Hallmann, 2011). Historically, two main mechanisms of germline specification have been defined in animals: preformation and epigenesis (Extavour and Akam, 2003), or determinative and inductive modes, respectively. In the determinative mode, cytoplasmic determinants associated with the germ plasm in the egg are inherited by a limited number of daughter cells during cleavage, which are thus specified as presumptive germ cells (PGCs), and go on to form the mature adult gametes. In contrast, during induction, inductive cues cause somatic cells to become specified as germline. Studies in mouse, axolotl, and cricket suggest that BMP signaling may be an ancient mechanism for PGC induction from mesoderm in animals (Chatfield et al., 2014; Donoughe et al., 2014). The phylogenetic distribution of these two mechanisms of germ cell specification suggests that the inductive mode may represent the ancestral state, and that germ

plasm has evolved independently multiple times (Blackstone and Jasker, 2003; Extavour and Akam, 2003; Johnson et al., 2003a,b; Crother et al., 2007; Extavour, 2007; Ewen-Campen et al., 2010).

Comparative studies in multiple taxa have revealed that the molecular signature of germ cells may often be shared across species that use both determinative and inductive modes of PGC specification, leading to the proposition of a conserved germline multipotency programme (Extavour, 2007; Juliano et al., 2010). Interestingly, some basal metazoans appear to use a combination of mechanisms to specify germ cells, and many of the classic germline markers are in fact also expressed in adult somatic stem cells in these organisms (Alié et al., 2011; Leclère et al., 2012). Germline-associated genes are most often RNA-binding proteins, but there is considerable species-specific variation in the suite employed (see Gazave et al., 2013 for a compilation of much of the recent literature). However, a key core of proteins including Vasa/PL10, Tudor and a PIWI domain containing protein may represent an ancestral “pluripotency module” (Ewen-Campen et al., 2010). Tweaking upstream regulators or downstream targets, combined with the addition of new germ cell genes, such as *nanos* or *bruno*, would have generated the diversity in germline specification mechanisms in early metazoans (Ewen-Campen et al., 2010).

Until recently, little was known about germline specification in cephalochordates (lancelets or amphioxus), the sister group to the vertebrates and tunicates, and the best living proxy for the ancestral chordate (Bertrand and Escrivà, 2011). Classic studies suggested that lancelets might employ an inductive mode of PGC specification (reviewed in Extavour and Akam, 2003). Electron microscopy data however showed that in *Branchiostoma floridae*, the pole plasm localizes to the vegetal cortex soon after fertilization and segregates into a single blastomere during cleavage, putting into question this hypothesis (Holland and Holland, 1992). Although functional data are still lacking, blastomere separation experiments combined with expression data for molecular markers traditionally associated with the germline, including *piwi-like1*, *nanos*, *vasa*, and Vasa protein, strongly support a determinative mode of PGC specification in cephalochordates (Wu et al., 2011; Zhang et al., 2013 and **Figure 1**). Zygotic expression domains from gastrulation onwards of these genes, as well as *piwi-like2* and *tudor7*, also suggest a function in somatic progenitors/stem cells of the posterior growth zone. Together, these data provide a general framework for understanding how markers for PGCs and posterior progenitors may be expressed in cephalochordates during development (**Figure 1**).

Currently, the most convincing evidence for the existence of somatic stem cells in cephalochordates comes from studies of tail regeneration in the European amphioxus, *Branchiostoma lanceolatum*, whose adult regenerative ability is comparable to that seen in many ambulacrarians (echinoderms and hemichordates; Somorjai et al., 2012a,b). Unfortunately, no germ cell markers have so far been characterized during development in this species, and few putative somatic stem cell markers exist in cephalochordates. The purpose of this study was therefore threefold: First, to characterize the early expression of candidate amphioxus germline markers *nanos*, *piwil1*, *vasa*,

and Vasa protein in *B. lanceolatum* for comparative purposes with other cephalochordates; second, to determine whether germline markers in other taxa, including *pumilio*, *pufA*, *pufB*, *musashi*, *pl10*, *bruno2*, and *bruno4/6*, are associated with PGCs in amphioxus; and third, to analyse the late developmental expression of some of these candidates as a prelude to future regeneration studies. Given the considerable conservation in developmental gene expression patterning in cephalochordates demonstrated thus far (e.g., Somorjai et al., 2008; Wu et al., 2011; Zhang et al., 2013), we hypothesize that markers for PGCs and posterior somatic domains will show comparable gene expression profiles in *B. lanceolatum* to *B. belcheri*, *B. japonicum*, and *B. floridae*.

Here, we present the first analysis of putative germline and somatic stem cell markers in the European amphioxus, *B. lanceolatum*. We identify a core set of conserved PGC-associated markers in cephalochordates, including *piwil1*, *nanos*, and *vasa*, characterize Vasa protein distribution, and identify two new candidate germ cell markers in cephalochordates, *pl10*, and *bruno2*. We also characterize the amphioxus *musashi* ortholog, whose expression in the notochord represents a novelty in chordates. The highly conserved molecular expression data in the Branchiostomatidae support the view that cephalochordate evolution is strongly constrained, and show that data are broadly transposable across species, even in the context of germline formation. This study also provides the foundation for future studies of regeneration in the amphioxus *B. lanceolatum*.

MATERIALS AND METHODS

Embryos

Ripe adults were collected in Argelès-sur Mer (France) and spawned as previously described (Fuentes et al., 2007). Embryos were fixed at the relevant time points in 4% PFA in MOPS salts (0.1 M MOPS, 2 mM MgSO₄, 1 mM EGTA, and 0.5 M NaCl), and stored in 70% ETOH at -20°C. For phalloidin staining, embryos were stored in PBS at 4°C. Embryos were staged according to Hirakow and Kajita (1991, 1994), with modifications as per Zhang et al. (2013).

Phylogenetic Analysis

If not previously published, putative orthologous sequences were identified using a BLASTp search; reciprocal BLAST was used to confirm identity (Camacho et al., 2009). Protein sequences were aligned in Jalview version 2.8.2 (Waterhouse et al., 2009) using MAFFT on default settings, and checked manually. All positions with less than 95% site coverage were eliminated directly in MEGA5 prior to analysis. Evolutionary models considered to best describe the substitution pattern were identified as those with the lowest BIC (Bayesian Information Criterion) scores using MEGA5 (Tamura et al., 2011). Both neighbor joining (NJ) and maximum likelihood (ML) analyses were performed with 500 and 1000 bootstraps, respectively. The Nearest Neighbor Interchange method was used to infer trees in ML. Unless otherwise noted, and since concordant with results from the NJ method, only ML trees are shown. Model details for each analysis are included in the figure legends for ease of reference. All

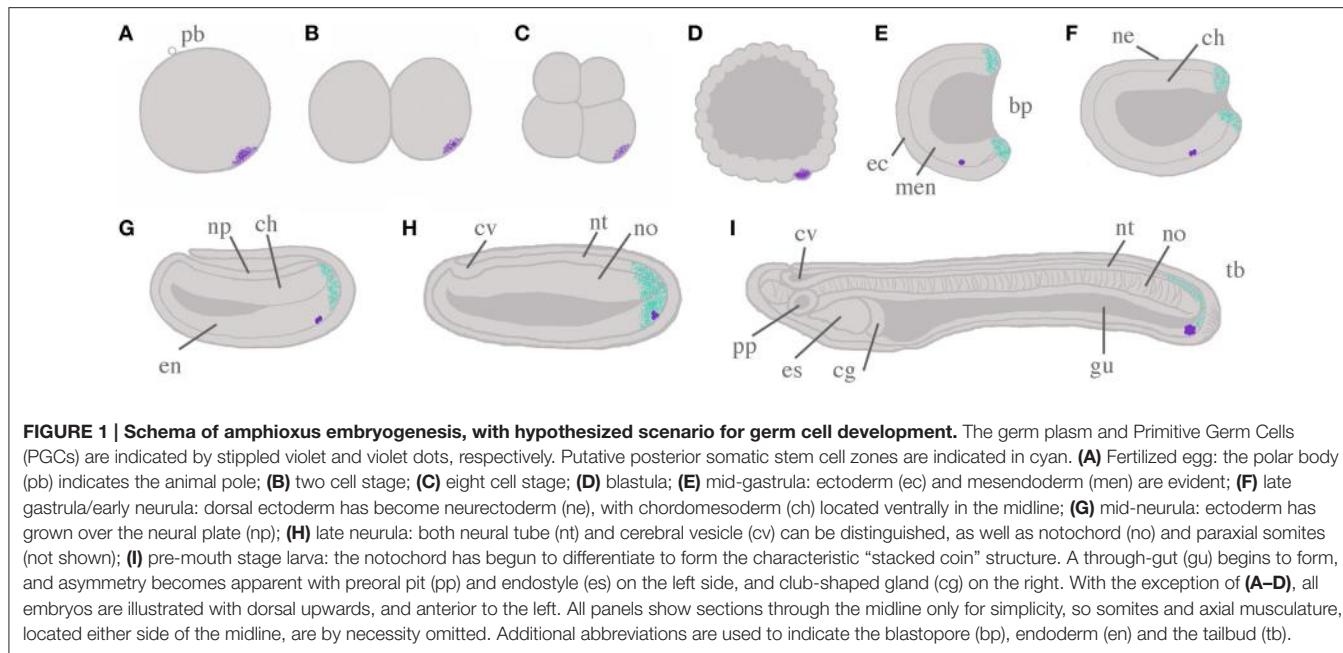


FIGURE 1 | Schema of amphioxus embryogenesis, with hypothesized scenario for germ cell development. The germ plasm and Primitive Germ Cells (PGCs) are indicated by stippled violet and violet dots, respectively. Putative posterior somatic stem cell zones are indicated in cyan. (A) Fertilized egg: the polar body (pb) indicates the animal pole; (B) two cell stage; (C) eight cell stage; (D) blastula; (E) mid-gastrula: ectoderm (ec) and mesendoderm (men) are evident; (F) late gastrula/early neurula: dorsal ectoderm has become neurectoderm (ne), with chordomesoderm (ch) located ventrally in the midline; (G) mid-neurula: ectoderm has grown over the neural plate (np); (H) late neurula: both neural tube (nt) and cerebral vesicle (cv) can be distinguished, as well as notochord (no) and paraxial somites (not shown); (I) pre-mouth stage larva: the notochord has begun to differentiate to form the characteristic “stacked coin” structure. A through-gut (gu) begins to form, and asymmetry becomes apparent with preoral pit (pp) and endostyle (es) on the left side, and club-shaped gland (cg) on the right. With the exception of (A–D), all embryos are illustrated with dorsal upwards, and anterior to the left. All panels show sections through the midline only for simplicity, so somites and axial musculature, located either side of the midline, are by necessity omitted. Additional abbreviations are used to indicate the blastopore (bp), endoderm (en) and the tailbud (tb).

sequences used for phylogenetic analyses, including associated accession numbers, are included in Supplementary File 1.

Cloning and Probe Synthesis

RNA was extracted from embryos and adult tissues using Trizol and phenol chloroform extraction; cDNA was generated using Tetro cDNA Synthesis kit (Bioline). Gene fragments for probe generation were amplified by PCR using gene-specific primers designed against the genome of *B. floridae* (Supplementary File 2), ligated into PGEMT-Easy (Promega) and transformed into XL10-Gold (Stratagene) or DH5 α (Invitrogen) strains of *E. coli* by heat shock using standard protocols. Selected clones were mini-prepped using pEqGOLD or Promega plasmid miniprep kits, and sequence verified. Template was generated by PCR on plasmids using Universal M13F (5' GTAAACGACGGCCAGT 3') and M13R (5' AACAGCTATGACCATG 3') primers. The band was gel-purified using either the QIAquick (Qiagen), GFX (Amersham), or Isolate II (Bioline) gel extraction kits following manufacturers guidelines. DIG-labeled (Roche) antisense probes were *in vitro* transcribed using T7, T3, and SP6 enzymes as appropriate following standard protocols. Probes were checked by agarose electrophoresis and purified using miniQuick Spin columns (Roche) or via precipitation with sodium acetate (3 M, pH 5.2) and ethanol.

Whole Mount *In situ* Hybridization (WMISH)

WMISH was performed as previously described (Somorjai et al., 2008). Briefly, fixed embryos were washed in PBT (0.1% Tween), and permeabilized using proteinase K (7.5 mg/ml) for empirically-tested periods based upon embryo stage and enzyme batch. Embryos were postfixed for 40 min in PFA, deacetylated in acetic acid in triethanolamine (0.1 M, pH 8), and pre-hybridized at least 2 h in hybridization solution. Embryos were incubated overnight with shaking at 60–65°C depending on probe. The first

post-hybridization washes were performed at the hybridization temperature, with subsequent washes at room temperature in decreasing concentrations of SSC. An RNase step was included (37°C). Embryos were incubated overnight in primary antibody (anti-DIG AP, Roche), pre-adsorbed at 1:3000, with rocking at 4°C. Copious washing in PBT was performed between each step. For the chromogenic reaction we used either BM Purple (Roche) or NBT/BCIP (Roche); embryos were postfixed in PFA for 20 min when the signal:background was deemed appropriate. At least three WMISH were performed for each gene, on 10–50 embryos per stage in total. Embryos were mounted in 80% glycerol/20% PBS, and photographed under a Leitz DMRB microscope (Leica Microsystems) with Normarski optics. Photographs were taken with the Retiga 2000R camera and the QCapture software suite (QImaging), and processed in Adobe Photoshop CS3.

Immunohistochemistry

Immunohistochemistry and Alexa-fluor 568-labeled phalloidin stainings for F-Actin (Invitrogen, 1:400) were carried out as per Somorjai et al. (2012a). Briefly, after fixation, embryos were washed in PBT (phosphate buffered saline plus 0.1% Tween, pH 7.6), and permeabilized in PBS with 0.2% Triton-X for 40 min. After copious washing in PBT, embryos were incubated overnight at 4°C in primary antibody. Embryos were again washed in PBT and incubated in secondary antibody or phalloidin for 2 h at room temperature, or overnight at 4°C. A specific *B. floridae* anti-Vasa antibody, generously donated by Dr Jr-Kai Yu, was used at 1:20,000 (Wu et al., 2011). Secondary antibodies were Alexa fluor 488 and 568 diluted at 1:400 (Molecular Probes). Embryos were mounted in Vectashield (VectorLabs) containing Hoescht 33342 dye to stain nuclei (1:2000 of 10 mg/ml). Confocal images were taken on a Leica TCS SP8 confocal microscope, and processed using NIH ImageJ 1.48 d and Adobe Photoshop CS3.

RESULTS

Identification of Candidate Germline and Somatic Stem Cell Markers in *B. lanceolatum*

We selected DEAD-box (*Vasa*, *Pl10*), Pumilio domain (Pumilio, PufA, PufB), PIWI domain (*Piwi1*, *Piwi2*) RRM (Musashi), CELF (Bruno2 and Bruno4/6), and Nanos families as candidate germline and somatic stem cell markers for analysis in *B. lanceolatum*. When *B. floridae* orthologs had not been previously characterized in the extensive phylogenetic analyses of Kerner et al. (2011), we identified putative stem cell markers using BLASTp searches against the genomes of *B. floridae* and *B. belcheri*, and confirmed the identity of our *B. lanceolatum* proteins by comparison with published sequences in other cephalochordates (Supplementary File 4), including transcriptomic data from *B. lanceolatum* in the NCBI TSA database (Oulion et al., 2012). We generated phylogenies that include, where possible, sequences from more than one amphioxus species to support the identity of these proteins (Supplementary File 5; and see below). We then cloned partial sequences of orthologs in *B. lanceolatum* using primers designed in its sister species *B. floridae* (Supplementary File 2). Using this approach, we successfully cloned 12 genes (including two *piwil1*; not shown) with known function in the germline or somatic stem cells (Table 1). While previous phylogenies show that the distinction among Piwi clades is unequivocal (Kerner et al., 2011), the evolutionary history of *piwi* genes in cephalochordates is more complex. We identified a single *piwil2* (*piwiA* in Kerner et al., 2011) and three *piwil1* (*piwiB* in Kerner et al., 2011) genes in the genomes of *B. belcheri* and *B. floridae*. The latter belong to an apparent tandem duplication cluster (not shown and Yue et al., 2015) that appears to be present in all *Branchiostoma*, as we successfully cloned two of the three paralogs of *piwil1* in *B. lanceolatum*. We also identified an ortholog of *piwiX* ("*piwilike*" in Zhang et al., 2013), but have been unable to clone the gene in *B. lanceolatum*. While EST data collected in NCBI and *B. floridae* EST databases (Yu et al., 2008) support the expression of *piwil1* and *piwil2* (Supplementary File 3), we have not identified any expression data for *piwiX* in any database, including our own tail regenerate transcriptome dataset (Dailey and Somorjai, unpublished).

We also cloned partial *pl10*, *vasa*, *nanos*, *bruno2* (*brunoB* or *CELF2* in Kerner et al., 2011), *bruno4/6* (*brunoA* or *CELF4/5/6* in Kerner et al., 2011), *pufA*, *pufB*, and *pumilio* sequences. The phylogenetic analyses broadly confirm previous studies (Kerner et al., 2011), though we could only confirm the existence of single A-type and B-type Bruno sequences. In most cases we could identify *B. belcheri* orthologs for the *B. floridae* proteins, in addition to several *B. lanceolatum* sequences (Supplementary Files 4, 5). EST data in *B. floridae* also supported the expression of these putative germline and somatic stem cell markers (Supplementary File 3).

As we were interested specifically in stem cell-related Musashi, and relationships among Musashi-related protein families are complex (Gasparini et al., 2011) we generated

phylogenies utilizing the available full length *B. floridae* and *B. belcheri* sequences, and included putative *Saccoglossus kowalevski* orthologs. We clearly identified sequences belonging to the TARDBP43 and hnRPAs/hnRPDs clades (Figure 2). The close relationship between Musashi-like and DAZAP proteins is also strongly supported by this analysis, although the branching order is unclear particularly within DAZAP sequences and in basal metazoans. Notably, we were unable to find an amphioxus sequence with convincing affinity to DAZAP/hnRP27 genes in either species. We did however identify a Musashi-like sequence in both *B. belcheri* and *B. floridae* (Figure 2, Supplementary File 4). In spite of the relatively low support for the Musashi clade, most likely due to the inclusion of

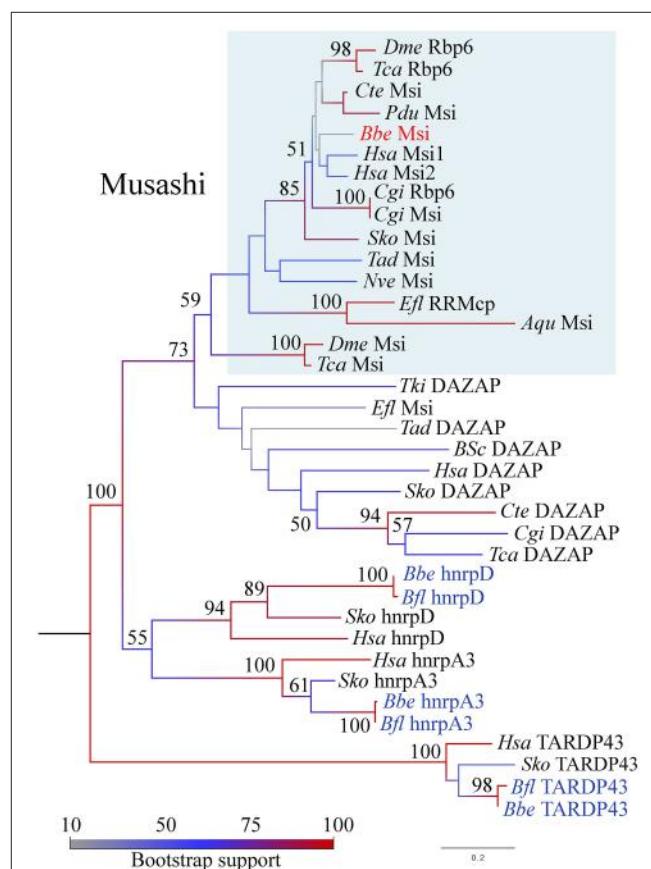


FIGURE 2 | Phylogenetic analysis of the RRM domain containing protein family in animals, including Musashi-like, DAZAP, hnRPD, hnRPAs, and TARDP43 clades. Maximum likelihood analysis was performed in MEGA5 with 1000 bootstrap replicates, indicated as a percentage at each node. The model used was rtREV + G with five rate categories on 173 sites. Branches are colored according to the level of node support; amphioxus species names are highlighted in blue and red, with the clade representing *musashi* genes boxed in blue. Protein names are taken directly from the literature where available. See text for details. Species name abbreviations are as follows: *Aqu*, *Amphimedon queenslandica*; *Bfl*, *Branchiostoma floridae*; *Bbe*, *Branchiostoma belcheri*; *Bsc*, *Botryllus schlosseri*; *Cgi*, *Crassostrea gigas*; *Cte*, *Capitella teleta*; *Dme*, *Drosophila melanogaster*; *Efl*, *Ephydatia fluviatilis*; *Hsa*, *Homo sapiens*; *Nve*, *Nematostella vectensis*; *Pdu*, *Platynereis dumerillii*; *Sko*, *Saccoglossus kowalevski*; *Tad*, *Trichoplax adhärens*; *Tca*, *Tribolium castaneum*; *Tki*, *Thelohanellus kitaei*; *Xtr*, *Xenopus tropicalis*.

TABLE 1 | Genes cloned in this study and accession numbers for all identified sequences in *Branchiostoma* species.

Gene name	<i>B. lanceolatum</i>	<i>B. floridae</i>	Identity Bla/Bfl%	Bla transcriptome	<i>B. belcheri</i>	<i>B. japonicum</i>
<i>bruno2</i>	KT354035.1	XP_002589327.1	64.91	JT864629.1	<i>Unpublished data</i> ¹	N.D.
<i>bruno4/6</i>	KT354036.1	XP_002588860.1	98.91	<i>Not found</i>	N.D.	N.D.
<i>musashi</i>	KT354038.1	XP_002590753.1	96.42	JT855672.1	N.D.	N.D.
				JT853755.1		
<i>nanos</i>	KT354040.1	ADM26639.1	96.64	<i>Not found</i>	AGI96000.1	AGI96005.1
<i>piwi-like 1</i>	KT354041.1	XP_002611937.1 AGI95996.1*	66.31	JT877108.1	AGI96002.1	AGI96006.1
<i>piwi-like 2</i>	KT354042.1	AGI95997.1XP_002596179.1*	85.46	JT858457.1	N.D.	N.D.
<i>pl10</i>	KT354043.1	XP_002593182.1	94.24	JT857642.1	ADD25830.1 <i>Unpublished data</i> ²	N.D.
<i>pufA</i>	KT354044.1	XP_002606015.1	95.08	JT902408.1	N.D.	N.D.
<i>pufB</i>	KT354045.1	XP_002607419.1	80	JT863417.1	N.D.	N.D.
<i>pumilio</i>	KT354046.1	XP_002601469.1	91.46	<i>Not found</i>	N.D.	N.D.
<i>vasa</i>	KT354047.1	ADM26640.1	84	JT881702.1	N.D.	AGI96004.1

Protein sequences predicted in genome or transcriptome assemblies are indicated by italicized accession numbers. Percentage identity of each *B. lanceolatum* clone is given relative to the most-complete available *B. floridae* protein (Bla/Bfl). Abbreviations: Bla, *B. lanceolatum*; Bfl, *B. floridae*; *, additional sequences; N.D., not determined. The sequence listed as "Unpublished data¹" is provided in Supplementary File 1 as "Bbe_Bruno2_076200F_001000_in," and "Unpublished data²" as "Bbe_PL10_173980F_003600_in."

non-bilaterian metazoan sequences and the divergent insect "Musashi" proteins, the amphioxus sequence groups with vertebrate and hemichordate sequences with strong support (85), in addition to the recently identified "real" *Drosophila* Musashi-related protein Rbp6 (Siddall et al., 2012). Insect "musashi" and *Rbp6* may therefore represent clade-specific duplications in this group from a *musashi*-like ancestor. We therefore propose that the *Rbp6/Msi* sequences be referred to as Musashi-like (blue boxed region in **Figure 2**), and all others outside the clade as DAZAP. Based on this nomenclature and the firm position of amphioxus *musashi* among deuterostome sequences, we are therefore confident that we identified a *musashi* gene orthologous to vertebrate *musashi1* and *musashi2*.

Candidate Marker Expression in Putative PGCs

Recently, expression patterns for putative germline markers have been described in three other species of amphioxus: *B. floridae*, *B. belcheri*, and *B. japonicum* (Wu et al., 2011; Zhang et al., 2013). We therefore performed WMISH for *piwil1*, *piwil2*, *vasa*, and *nanos* orthologs in early developmental stages of *B. lanceolatum*. **Figures 3A–D** show the characteristic expression in single "points" from the two cell stage to the gastrula stage in all four genes with the exception of *piwil2*. In some cases, the morulae or gastrulae contained up to three points (not shown). By the early neurula stages, the punctate distribution may be masked by the zygotic tailbud expression (discussed below).

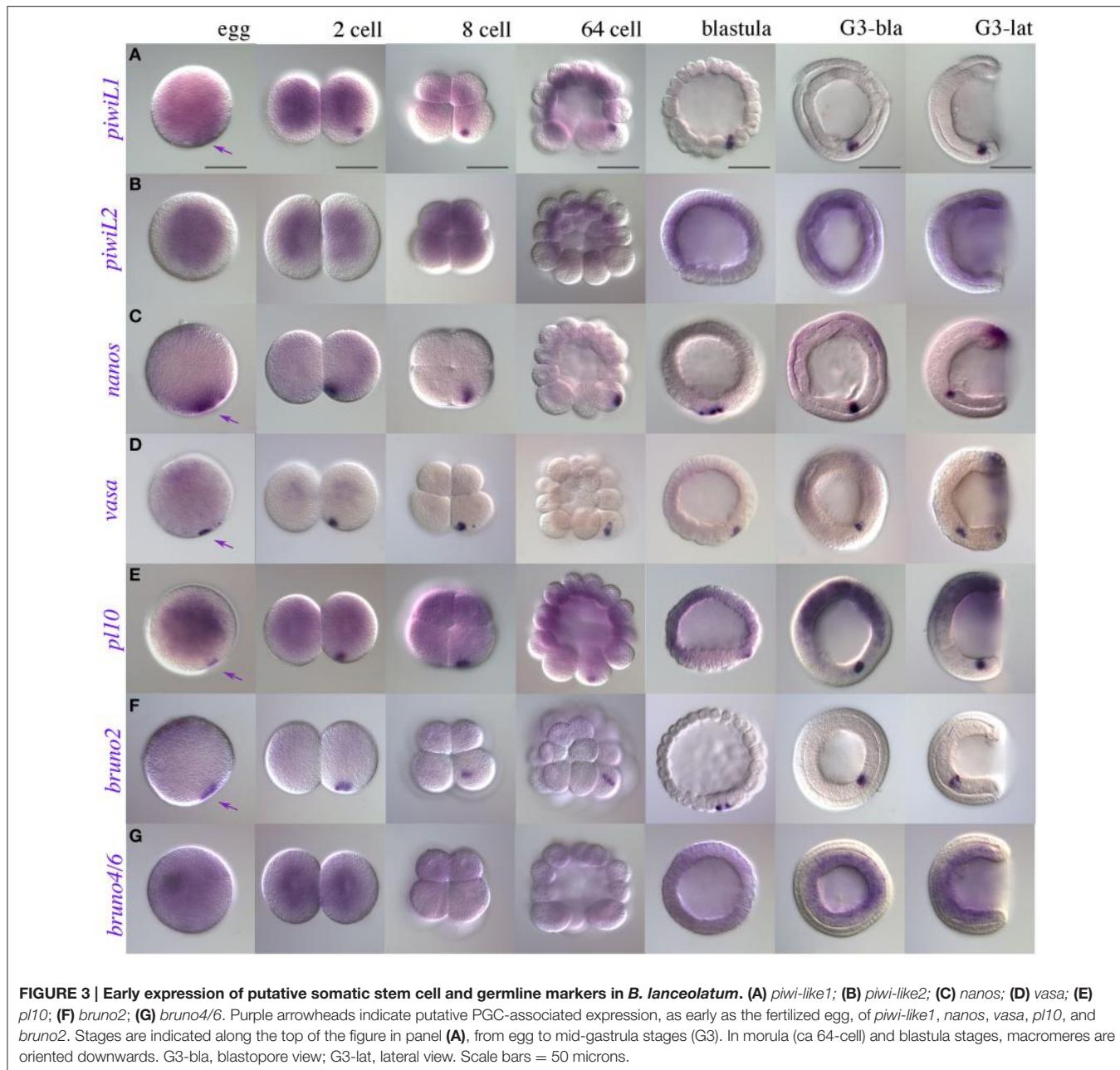
Another dead-box containing gene, *pl10*, has been implicated in germ cell specification, and in some cases regeneration, from sponges to annelids (Alié et al., 2011; Rebscher et al., 2012; Leininger et al., 2014; Kozin and Kostyuchenko, 2015). PL10 is closely related phylogenetically to the Vasa protein (Kerner et al., 2011), but expression of *pl10* has so far not been described in any cephalochordate. We therefore cloned a clear *pl10* ortholog in *B. lanceolatum* (Supplementary Files 4, 5) and determined its expression using WMISH. Like *vasa*, *pl10* is

expressed in a punctate pattern from fertilization until gastrula stages, consistent with a role in PGC specification or maintenance (**Figure 3E**).

We also determined early expression of members of three other classes of RNA-binding proteins that we might expect to have a stem cell association based on reports in other species: the Pumilio domain containing genes *pumilio*, *pufA*, and *pufB*; the CELF/Bruno genes *bruno2* and *bruno4/6*, and *musashi* (Gazave et al., 2013 and references therein). Up to gastrulation, *pumilio*, *pufA*, and *pufB* show no clear localization in the presumptive germline (Supplementary Files 6A–C). Interestingly, *pufA* ESTs are found in blastula-stage embryos, and we observe several independent but convincing instances in which *pufA* appeared to be expressed in a punctate distribution reminiscent of our other PGC-associated patterns in some cleavage stage embryos (Supplementary File 6A). Similarly to PUM domain containing genes, *bruno4/6* was absent in *B. floridae* EST databases, and showed no convincing expression until gastrulation (**Figure 3G**). In contrast, *bruno2* showed clear and strong localization to nuage or PGCs (**Figure 3F**). No other marker analyzed had specific expression in the presumptive PGCs (Supplementary File 6), including *musashi*, which had diffuse ubiquitous expression at early stages (Supplementary File 6D; see Supplementary File 7 for sense control).

Vasa Protein Distribution is Consistent with PGCs and Somatic Progenitor Cell Domains in *B. lanceolatum*

Along with transcript expression, localization of Vasa is a hallmark of primordial germ cells (PGCs) in multiple species. In order to confirm the identity of PGCs in *B. lanceolatum*, we took advantage of the recent generation of an antibody against *B. floridae* Vasa (Wu et al., 2011) to perform immunohistochemistry. Given its clear cross-reaction in several amphioxus species (Zhang et al., 2013), we reasoned that α -BfVasa should also label PGCs in the European amphioxus,



confirming our expression data. The protein distribution resembles that of *vasa* transcripts (Figure 3D), with a pattern reminiscent of germplasm in fertilized eggs and cleavage stages (Figures 4A–D). In the late gastrula/early neurula, the protein is perinuclear in small clusters of cells within the ventral endoderm (Figures 4E,F); although variable in number (or at least detection), we could clearly identify as many as eight cells by the careful analysis of series of confocal image z-sections (Figure 4F). Such clusters could be identified even in some mid-neurula stage embryos, either on one side in the ventral mesoderm (Figure 4G and inset), or in most cases posteriorly congruent with the zygotic tailbud domain (Figure 4H and inset). *Vasa* expression was however most conspicuous in the posterior neural

tube throughout neurulation (Figures 4I,J). Only in premouth stage and later larvae was it possible to again more easily identify posterior clusters of *Vasa*-expressing cells as distinct from posterior neural and tailbud expression (Figures 4K,L). *Vasa* also appeared to demarcate the posteriomost somites (not shown), similarly to *vasa* transcripts (Figure 5C, see below).

Candidate Stem Cell Marker Expression in Developing Somatic Tissues

We performed WMISH for selected genes from gastrulation onwards, reasoning that they should show expression patterns with possible roles in late developmental processes (Figure 5 and Supplementary File 8). We thus identified two classes:

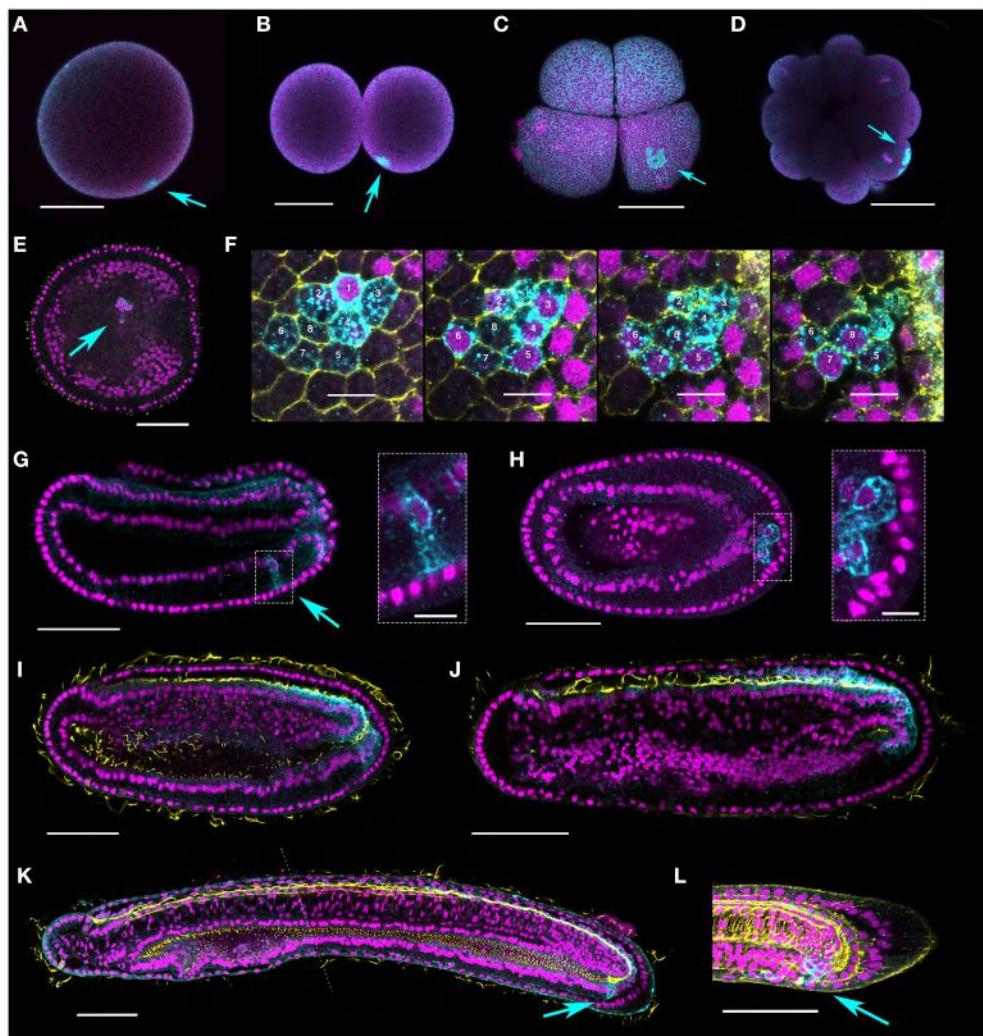


FIGURE 4 | Vasa protein expression during *B. lanceolatum* embryogenesis. Vasa is expressed in a single point in fertilized eggs (A), two cell (B), eight-cell (C), and morula (D) stages (arrows). (E,F) Perinuclear punctae are evident in late gastrula/early neurula (N0–N1) stages in up to eight cells (two different embryos shown, E and F). (E) shows a cluster of cells of which two clearly express Vasa (arrow). In (F), a series of z-stacks shows a cluster of eight cells expressing Vasa. Mid-neurula (N2) stage embryos (G,H) are variable, with some individuals showing a cluster of Vasa-positive cells in the ventral endoderm on one side (G, arrow and boxed inset); in others the putative germ cells are just ventral to and contiguous with the posterior Vasa expression domain (H, arrow and boxed inset). (I,J) germline expression is difficult to distinguish in late neurula (N3) stage embryos; zygotic expression is most apparent in the posterior neural tube and posteriomost ventral endoderm. (K,L) Clusters of cells are again evident in some pre-mouth stage and late larvae (L1, L2 arrows), and neural expression is much reduced. The dotted line in (K) indicates where two confocal images have been combined. Nuclei stained with Hoescht are false-colored magenta, and Vasa protein is shown in cyan. Acetylated tubulin marks axons and cilia in yellow in (E), (I–K). Phalloidin stains F-actin yellow in the membranes in (F) and (L). With the exception of scale bars in (F) and in boxed insets (G,H), which are 10 microns, all scale bars are 50 microns. Orientations are as follows: (E–L) anterior, left; (E,F,H), ventral views; (G,I–L), lateral, dorsal is to the top.

“tailbud-enriched” and, broadly speaking, “anterior endoderm-associated.” We found that *piwil1*, *nanos*, and *vasa* have strong tailbud expression throughout development (Figures 5A–C). *piwil1* and *nanos* also show clear posterior neural tube expression in N4 neurulae and L1 stage premouth larvae, as well as expression outlining the posterior somites (black arrowheads, Figure 5A; Supplementary File 8A). Though weaker, *piwil2* and *pl10* both show tailbud expression at later stages, and *pl10* is clearly expressed in the neural tube (Supplementary File 9).

In contrast, Pumilio domain containing genes appear enriched in anterior endoderm (Figures 5D,E and

Supplementary Files 8D,E). During gastrulation, *pumilio* shows weak expression around the blastopore. In early and mid-neurula stages, stronger expression is evident in the neural plate and anterior ventral endoderm, as well as anterior mesoderm. As neurulation proceeds, *pumilio* appears mostly restricted to the anterior endoderm, with expression much weaker in the last third of the embryo (Figure 5E). Expression continues to be strongest in the future pharyngeal domain until the pre-mouth larval stage. Weaker expression is evident in the rest of the endoderm, with some conspicuous staining in the mesoderm and endoderm of the tailbud region. Expression of

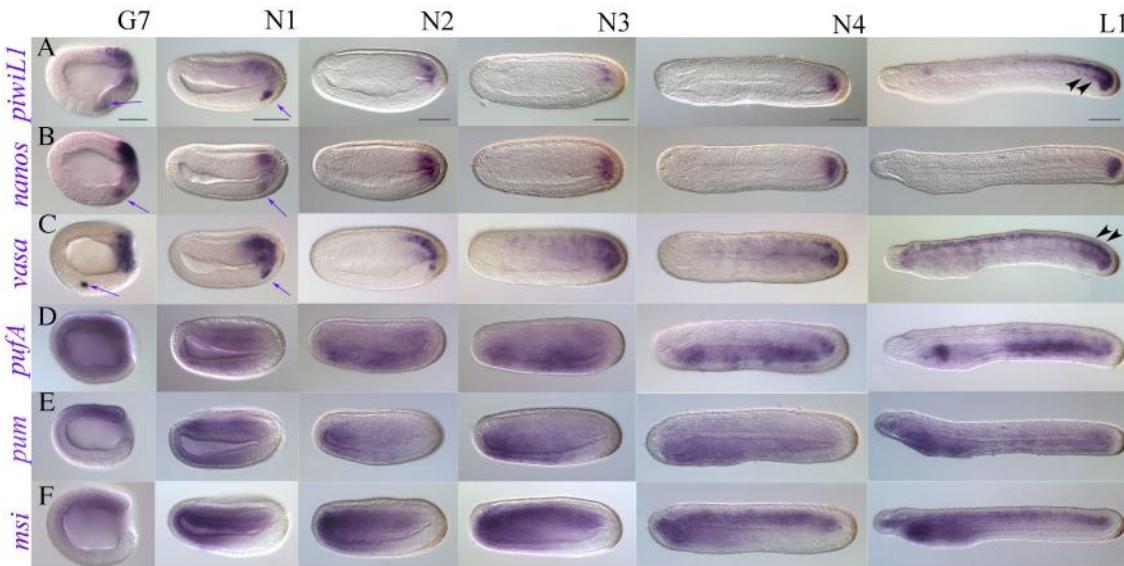


FIGURE 5 | Late expression of a representative subset of putative somatic stem cell and germline makers in *B. lanceolatum* (lateral views). (A) *piwi-like1*; (B) *nanos*; (C) *vasa*; (D) *pufA*; (E) *pumilio*; (F) *musashi*. Embryonic stages are indicated along the top of panel (A), from late gastrula/early neurula G7 to the premouth larval stage L1. Purple arrows indicate PGC-like expression that continues to be detectable for *piwi-like1*, *nanos*, and *vasa* into the early neurula stage N1. Black arrowheads in (A) and (C) indicate expression demarcating somite boundaries. All views are lateral, and all panels are oriented with anterior to the left and dorsal up. Scale bars = 50 microns.

pufA is broadly mesendodermal until N3 neurula stages, when it becomes stronger in an anterior domain that resolves into the club shaped gland in premouth L1 larval stages, as well as in most of the posterior endoderm (Figure 5D). *pufB* expression was very difficult to evaluate as long staining exposures were required for stages post-gastrulation, but quite closely matched that of *pufA* (not shown). In addition to its posterior expression, *pl10* shows diffuse but clear staining in anterior endoderm in N4 and L1 stages (Supplementary Files 9G–I), and clearly resolves to a domain encompassing the presumptive first gill slit in 2–3 day-old larvae (not shown).

We cloned the amphioxus *musashi* ortholog with the expectation that it would have neural expression. During early stages of development, *musashi* is ubiquitously expressed (Supplementary File 6D), paralleling *B. floridae* EST data (Supplementary File 3). However, in the gastrula stage, *musashi* resolves to a chordomesodermal domain of expression (Supplementary Files 6D, 7), which broadens in the early neurula N1 (Figure 5F). *musashi* is strongly expressed in the anteriormost endoderm and mesoderm from mid-neurula onwards, with weak expression in the neural floorplate and strong expression throughout the chordal plate. By the late neurula stage (30 h, N4) patches of expression can be seen in the neural tube as well as weakly in the cerebral vesicle. Expression is high and stable throughout the forming notochord as well as in the anterior endoderm. Strong notochordal and weak neural expression domains persist in the premouth L1 larva, with strongest expression in the anterior and posteriormost domains of the notochord. The presumptive pharynx also expresses *musashi*.

DISCUSSION

Germline-Associated Gene Expression Conservation in Cephalochordates

Recent work in *B. floridae*, *B. japonicum*, and *B. belcheri* has suggested that germline specification occurs by the asymmetric segregation of cytoplasmic determinants during cleavage, with expression of key conserved germline markers such as *vasa* and *nanos*, as well as *piwil1* and *tudor-related7*, in the germ plasm and PGCs (Wu et al., 2011; Zhang et al., 2013). We set out here to characterize the expression of germline-associated markers in the European amphioxus, *B. lanceolatum*, for which there were until now no data. Similarly to other species, our results also argue against an inductive mechanism for PGC specification: we demonstrate here that *B. lanceolatum* expresses *nanos*, *piwil1*, and *vasa* in the putative PGCs, as well as Vasa protein, suggesting the presence of a conserved core of germline-associated transcripts in cephalochordates. Stasis in developmental gene expression over millions of years of evolution is considered typical of *Branchiostoma* (Somorjai et al., 2008), paralleling the genus' relative genomic and morphological conservativeness. The apparent conservation in germline-associated gene expression in amphioxus species is in stark contrast to hypotheses derived in vertebrates that suggest that the evolution of germ plasm is coupled to increased speciation in this lineage (Johnson et al., 2011; Evans et al., 2014). Data in *Asymmetron*, the earliest diverging and most slowly evolving of the three extant amphioxus lineages (Kon et al., 2007; Yue et al., 2014), will be invaluable in evaluating the degree of conservation of germline specification mechanisms in cephalochordates.

Our research also identifies *pl10*, a DEAD-box gene related to *vasa*, and *bruno2* as putative PGC markers in amphioxus. Accumulating evidence suggests that *pl10* often plays a role in the germline in metazoans. In addition to *Drosophila*, *pl10* orthologs are expressed in the germinal cells or their derivatives in the annelid *Platynereis dumerilii* (Rebscher et al., 2007; Gazave et al., 2013), the platyhelminth *Dugesia japonicum* (Shibata et al., 1999; reported as *vasa*-related genes) several hydrozoan cnidarian species (Leclère et al., 2012; Siebert et al., 2015), the ctenophore *Mnemopsis leydii* (Alié et al., 2011), the sponge *Sycon ciliatum* (Leininger et al., 2014), and in the colonial urochordate *B. schlosseri* (Rosner et al., 2009). *vasa* is coexpressed with *pl10* in the latter, similarly to our results in *B. lanceolatum*. In contrast, data are sparse for the second gene identified, *bruno2*. Homologs of *bruno* are expressed in PGCs and/or germline derivatives in ctenophores (Alié et al., 2011), but not in *Platynereis* (Gazave et al., 2013). Interestingly, using RNAi, the Bruno-like gene *bruli* was shown to be required for maintenance of a subset of neoblasts in the asexual planarian *Schmidtea mediterranea* (Guo et al., 2006), but this gene is not homologous to canonical *bruno* genes. Confirmation of expression of *pl10* and *bruno2* in other species, and Tudor related *tdrd7* in *B. lanceolatum*, will further expand this repertoire.

We also identified several markers with weak ubiquitous expression during early development. For instance, *musashi*, *piwil2*, and genes of the Pumilio domain family do not appear to be associated specifically with PGCs in *B. lanceolatum* or *B. floridae* (this study; Yue et al., 2015). A possible exception is *pufA*, which we found to be concentrated in a PGC-like domain in some cleavage stage embryos (two-cell to morula) in several independent experiments. Given the variability of the expression observed, we hesitate to classify this as *bona fide* expression in PGCs. However, *pufA* is expressed in germ cells in other species, including zebrafish (Kuo et al., 2009) and *P. dumerilii* (Gazave et al., 2013). Interestingly, a global search of germline and reproduction-associated genes using the transcriptome and genome of *Asymmetron lucayanum* and *B. floridae*, respectively, identified a *pumilio/puf* gene with expression in oocytes (Yue et al., 2015). Similar studies in maturing gonads in *B. lanceolatum* may also reveal functions for some of our candidates during germline maturation.

Evolution of Musashi Related RRM-Containing Proteins and Novel Expression of Amphioxus *musashi*

The Musashi related proteins belong to a larger superfamily of RRM containing proteins, including Musashi, DAZAP, hnrp, and TARDBP clades. Although the evolutionary history of RRM domain containing proteins is complex, orthologs of *musashi*-related genes have been identified from sponge to human (Gasparini et al., 2011; Okamoto et al., 2012), including lancelets (Gasparini et al., 2011, this study). One of the principal findings of this study is that cephalochordates appear to have lost the ortholog of DAZAP, as we were unable to identify the gene in either the genomes of *B. floridae* or *B. belcheri*. Considerable confusion exists in the nomenclature in the literature due to

the difficulty in distinguishing between *musashi* and *DAZAP* related genes. This is particularly evident in basal metazoans, where phylogenetic signal is weak (Okamoto et al., 2012, this study). Gasparini et al. (2011) first suggested that previously identified *musashi*-like genes in *Halocynthia roretzi* and *Ciona intestinalis* (Kawashima et al., 2000) are in fact *DAZAP*. This gene is expressed in the brain and nerve cord, as might be expected from *musashi*-like genes (Kawashima et al., 2000). However, the *bona fide* *DAZAP1* in *Botryllus schlosseri* is expressed both during asexual (blastogenesis) and sexual (embryonic) development in many proliferating cell types, including the new growing vessels of the colonial circulatory system and the embryonic nerve cord, and is not restricted to neural stem cells as in other systems. Likewise, in the planarian *D. japonicum* *DAZAP/musashi*-like gene *Djdm1g* is expressed in differentiated tissues as well as X-ray sensitive neoblasts (Higuchi et al., 2008). In this context, it would be particularly interesting to determine whether the cephalochordate *musashi* is taking on any of the *DAZAP* functions, or whether a different functional homolog might be involved.

Our observation that neural cells within the developing CNS of amphioxus express *musashi* is broadly consistent with data in bilaterians. For instance, in the flatworm *Dugesia japonica*, three *musashi*-like genes have been identified with expression in the brain primordia (Higuchi et al., 2008). Similarly, in zebrafish, *musashi1* is expressed in neural tissues during early development, and knockdown by morpholino results in aberrant CNS formation (Shibata et al., 2012). Surprisingly however, amphioxus does not express *musashi* in a pattern consistent with a role in PGC specification or maintenance, in contrast to many other taxa. In *Drosophila*, *musashi* is required to maintain stem cell identity in GSCs (Siddall et al., 2006), and *Rbp6*, which is more closely related to vertebrate *musashi1/2* (Siddall et al., 2012; this study), also may play some function in the germline (Siddall et al., 2012). In mice, the *msi1* and *msi2* orthologs appear to have sub-functionalized such that *msi1* is required to maintain stem cell identity during early spermatogenesis, whereas *msi2* plays a role in differentiation (Siddall et al., 2006). The generation of specific antibodies will be critical to gaining an understanding of the distribution of Musashi protein during amphioxus development and stem cell regulation.

Given its known neural and germline functions, the finding that *musashi* is predominantly expressed in the developing notochord in amphioxus was unexpected. We are not aware of any data demonstrating a specific function for *musashi* in the notochord in any chordate. However, the ancestral function of these RRM containing proteins may simply be in the switch between undifferentiated/stem cell and differentiated cell types and in the regulation of proliferation (Potten et al., 2003; MacNicol et al., 2011; Hochgreb-Hägle et al., 2014). Supporting this, the anterior endoderm encompassing the zones that will form the mouth and gill slits in amphioxus larvae, which has conspicuous *musashi* expression, is a zone of extensive proliferation and remodeling (Holland and Holland, 2006). The expression in developing notochord described here, which is unique to amphioxus, might also reflect a role in differentiation

of this structure. Functional studies will help elucidate the role of Musashi in this and other structures.

Posterior Stem Cells and Implications for Amphioxus Regeneration

The zygotic expression of several markers in the tailbud, including *nanos*, *vasa*, *piwil1*, and *piwil2* among others, combined with circumstantial evidence that PGCs may migrate at the neurula stage toward the posterior (this study; Wu et al., 2011), suggest that the tailbud may be a source of progenitors or stem cells in larval amphioxus. Posterior elongation in amphioxus involves budding of somites directly from the tailbud, a source of Wnt ligand (Holland et al., 2000; Schubert et al., 2000, 2001; Somorjai et al., 2008). Although architecturally different, the tailbuds of vertebrates like mouse and chick are also sources of multipotent stem cells for embryonic elongation whose fate is Wnt signaling-dependent (Wilson et al., 2009; Garriock et al., 2015). The posterior growth zone may also act as a niche for progenitor cells even into adulthood, particularly in animals that add segments throughout their lives, such as many arthropods and most annelids (Bely and Wray, 2001; de Rosa et al., 2005; Seaver et al., 2005). The observation that the Vasa-positive PGCs lie within a stem cell marker-expressing posterior growth zone in amphioxus larvae (Wu et al., 2011; this study), representing a “mosaic” of PGCs and somatic stem cells, is however not unique to amphioxus. Gazave et al. (2013) have proposed the existence of an RNA binding protein signature for a new type of animal stem cell, termed “posterior stem cells,” in *P. dumerilii*. Lineage analysis and EdU labeling have also revealed that the 4 presumptive PGCs, which appear during gastrulation, are derived from a mesoderm posterior growth zone (MPGZ; Rebscher et al., 2007, 2012). While the mechanisms employed by these annelids and cephalochordates to specify the germline are somewhat different, the use of such techniques in amphioxus will be instrumental in elucidating the origin and fate of different cell types during posterior elongation.

The existence of a posterior stem cell in the tailbud, or any other resident stem cell population that could be activated following tail amputation, has clear implications for regeneration in amphioxus. Although it has recently been demonstrated that the European amphioxus has considerable regenerative ability, most notably of the tail (Somorjai et al., 2012a,b), we still know next to nothing about the molecular signature or function of the somatic stem cells/progenitor pools involved in the process. This study represents the first step toward identifying a putative posterior stem cell pool in *B. lanceolatum*. Our prediction is that somatic stem cell markers that are normally expressed during tailbud development, such as *vasa*, *nanos*, *piwil1*, or *piwil2*, will also be expressed during the adult tail regeneration process. We are currently analysing blastema transcriptomes and proteomes to test this hypothesis (Dailey and Somorjai, unpublished). We might also expect to find genes traditionally associated with the germline to be expressed during tail regeneration, if common expression of “stemness” markers in PGCs and somatic stem cells reflect broader roles in developmental regulation, as has recently been demonstrated for Vasa in the sea urchin (Yajima and Wessel, 2015). Although functional experiments are lacking,

comparative expression data in annelids are beginning to provide compelling evidence for this. In *P. dumerilii*, a number of RNA binding protein genes are expressed in PGCs as well as in putative posterior mesodermal and ectodermal stem cells during caudal regeneration, including *vasa*, *pl10*, *piwi*, *pufA*, *pufB*, *nanos*, and several *tudor* related genes (Rebscher et al., 2007; Gazave et al., 2013). Of these, several markers are also differentially expressed both in the germline and terminal growth zone during normal development and regeneration in the polychaetes *Alitta virens* and *Capitella sp I* (Dill and Seaver, 2008; Giani et al., 2011; Kozin and Kostyuchenko, 2015). However, the most striking example of a germline-independent redeployment of classic PGC markers in somatic tissues has been shown in the freshwater annelid *Pristina leidyi*, which reproduces exclusively asexually in the laboratory via paratomic fission. As might be expected, *nanos*, *piwil1*, and *vasa* are expressed in the posterior growth zone and developing (but unused) gonads. Notably, transcripts are also detected following amputation in the anterior blastema as well as the fission zone (Bely and Sikes, 2010; Özpolat and Bely, 2015), highlighting a more general role in tissues undergoing proliferation and remodeling. This phenomenon is not restricted to invertebrates or basal metazoans, as *piwil1* and *piwil2* are expressed in a complex spatiotemporal sequence during axolotl limb regeneration, with knockdown of either gene resulting in retardation of the regenerate outgrowth (Zhu et al., 2012). Future work in amphioxus will assess the tissue-specific expression pattern of some of the candidates identified here during adult tail regeneration. Development of knockdown tools and lineage analysis will be indispensable to elucidate their functional role during the regeneration process. Moreover, these methodologies will permit the comparative analyses of cellular and molecular processes necessary to understand the evolution of regeneration mechanisms in deuterostomes. More broadly, these types of studies should add to the growing body of literature aimed at understanding the link between soma and germline evolution.

AUTHOR CONTRIBUTIONS

SD, RF, and AR performed experiments. JG discussed experiments and contributed reagents. IS conceived the study, performed experiments, contributed reagents, analyzed the data and wrote the manuscript.

ACKNOWLEDGMENTS

This work was carried out sporadically over the course of several years in various countries. We would like to thank Jr-Kai Sky Yu (Academia Sinica, Taiwan) for contributing the Vasa antibody, and Irene Garcia for help in the laboratory (Barcelona). SD is funded through a MASTS PhD studentship (St Andrews, UK). IS gratefully acknowledges previous funding from Marie Curie IEF postdoctoral fellowship, FP7 People Programme (Barcelona, Spain); as well as MASTS (Marine Alliance for Science and Technology Scotland) laboratory start-up funds (St Andrews, UK). Embryo collection was made possible in part through the ASSEMBLE access programme (grant agreement

no. 227799). We thank the Laboratoire Océanologique de Banyuls-sur-Mer, and most especially Dr. Hector Escrivà and Dr. Stéphanie Bertrand for hosting us. The University of St Andrews Library fund for open access supported the article publishing fee.

REFERENCES

- Alié, A., Leclère, L., Jager, M., Dayraud, C., Chang, P., Le Guyader, H., et al. (2011). Somatic stem cells express Piwi and Vasa genes in an adult ctenophore: ancient association of “germline genes” with stemness. *Dev. Biol.* 350, 183–197. doi: 10.1016/j.ydbio.2010.10.019
- Bely, A. E., and Sikes, J. M. (2010). Latent regeneration abilities persist following recent evolutionary loss in asexual annelids. *Proc. Natl. Acad. Sci. U.S.A.* 107, 1464–1469. doi: 10.1073/pnas.0907931107
- Bely, A. E., and Wray, G. A. (2001). Evolution of regeneration and fission in annelids: insights from *engrailed-* and *orthodenticle*-class gene expression. *Development* 128, 2781–2791.
- Bertrand, S., and Escrivà, H. (2011). Evolutionary crossroads in developmental biology: amphioxus. *Development* 138, 4819–4830. doi: 10.1242/dev.066720
- Blackstone, N. W., and Jasker, B. D. (2003). Phylogenetic considerations of clonality, coloniality, and mode of germline development in animals. *J. Exp. Zool. B Mol. Dev. Evol.* 297, 35–47. doi: 10.1002/jezb.16
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., et al. (2009). BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. doi: 10.1186/1471-2105-10-421
- Chatfield, J., O'Reilly, M. A., Bachvarova, R. F., Ferjentsik, Z., Redwood, C., Walmsley, M., et al. (2014). Stochastic specification of primordial germ cells from mesoderm precursors in axolotl embryos. *Development* 141, 2429–2440. doi: 10.1242/dev.105346
- Crother, B. I., White, M. E., and Johnson, A. D. (2007). Inferring developmental constraint and constraint release: primordial germ cell determination mechanisms as examples. *J. Theor. Biol.* 248, 322–330. doi: 10.1016/j.jtbi.2007.05.035
- de Rosa, R., Prud'homme, B., and Balavoine, G. (2005). *Caudal* and *even-skipped* in the annelid *Platynereis dumerilii* and the ancestry of posterior growth. *Evol. Dev.* 7, 574–587. doi: 10.1111/j.1525-142X.2005.05061.x
- Dill, K. K., and Seaver, E. C. (2008). *Vasa* and *nanos* are coexpressed in somatic and germ line tissue from early embryonic cleavage stages through adulthood in the polychaete *Capitella* sp. I. *Dev. Genes Evol.* 218, 453–463. doi: 10.1007/s00427-008-0236-x
- Donoughe, S., Nakamura, T., Ewen-Campen, B., Green, D. A. I. I., Henderson, L., and Extavour, C. G. (2014). BMP signaling is required for the generation of primordial germ cells in an insect. *Proc. Natl. Acad. Sci. U.S.A.* 111, 4133–4138. doi: 10.1073/pnas.1400525111
- Evans, T., Wade, C. M., Chapman, F. A., Johnson, A. D., and Loose, M. (2014). Acquisition of germ plasm accelerates vertebrate evolution. *Science* 344, 200–203. doi: 10.1126/science.1249325
- Ewen-Campen, B., Schwager, E. E., and Extavour, C. G. (2010). The molecular machinery of germ line specification. *Mol. Reprod. Dev.* 77, 3–18. doi: 10.1002/mrd.21091
- Extavour, C. G. (2007). Evolution of the bilaterian germ line: lineage origin and modulation of specification mechanisms. *Integr. Comp. Biol.* 47, 770–785. doi: 10.1093/icb/icm027
- Extavour, C. G., and Akam, M. (2003). Mechanisms of germ cell specification across the metazoans: epigenesis and preformation. *Development* 130, 5869–5884. doi: 10.1242/dev.00804
- Fuentes, M., Benito, E., Bertrand, S., Paris, M., Mignardot, A., Godoy, L., et al. (2007). Insights into spawning behavior and development of the European amphioxus (*Branchiostoma lanceolatum*). *J. Exp. Zool. B Mol. Dev. Evol.* 308, 484–493. doi: 10.1002/jez.b.21179
- Garirock, R. J., Chalamalasetty, R. B., Kennedy, M. W., Canizales, L. C., Lewandoski, M., and Yamaguchi, T. P. (2015). Lineage tracing of neuromesodermal progenitors reveals novel Wnt-dependent roles in trunk progenitor cell maintenance and differentiation. *Development* 142, 1628–1638. doi: 10.1242/dev.11922
- Gasparini, F., Shimeld, S. M., Ruffoni, E., Burighel, P., and Manni, L. (2011). Expression of a Musashi-like gene in sexual and asexual development of the colonial chordate *Botryllus schlosseri* and phylogenetic analysis of the protein group. *J. Exp. Zool. B Mol. Dev. Evol.* 316, 562–573. doi: 10.1002/jez.b.21431
- Gazave, E., Béhague, J., Laplane, L., Guillou, A., Préau, L., Demilly, A., et al. (2013). Posterior elongation in the annelid *Platynereis dumerilii* involves stem cells molecularly related to primordial germ cells. *Dev. Biol.* 382, 246–267. doi: 10.1016/j.ydbio.2013.07.013
- Giani, V. C. Jr., Yamaguchi, E., Boyle, M. J., and Seaver, E. C. (2011). Somatic and germline expression of piwi during development and regeneration in the marine polychaete annelid *Capitella teleta*. *Evodevo* 2:10. doi: 10.1186/2041-9139-2-10
- Guo, T., Peters, A. H., and Newmark, P. A. (2006). A *bruno-like* gene is required for stem cell maintenance in planarians. *Dev. Cell* 11, 159–169. doi: 10.1016/j.devcel.2006.06.004
- Hallmann, A. (2011). Evolution of reproductive development in the volvocine algae. *Sex. Plant Reprod.* 24, 97–112. doi: 10.1007/s00497-010-0158-4
- Higuchi, S., Hayashi, T., Tarui, H., Nishimura, O., Nishimura, K., Shibata, et al. (2008). Expression and functional analysis of *musashi*-like genes in planarian CNS regeneration. *Mech. Dev.* 125, 631–645. doi: 10.1016/j.mod.2008.03.002
- Hirakow, R., and Kajita, N. (1991). Electron microscopic study of the development of amphioxus, *Branchiostoma belcheri tsingtauense*: the gastrula. *J. Morphol.* 207, 37–52. doi: 10.1002/jmor.1052070106
- Hirakow, R., and Kajita, N. (1994). Electron microscopic study of the development of amphioxus, *Branchiostoma belcheri tsingtauense*: the neurula and larva. *Acta Anat. Nippon* 69, 1–13.
- Hochgreb-Hägle, T., Koo, D. E., Das, N. M., and Bronner, M. E. (2014). Zebrafish stem/progenitor factor *msi2b* exhibits two phases of activity mediated by different splice variants. *Stem Cells* 32, 558–571. doi: 10.1002/stem.1583
- Holland, L. Z., and Holland, N. N. (1992). Early development in the lancelet (=Amphioxus) *Branchiostoma floridae* from sperm entry through pronuclear fusion: presence of vegetal pole plasma and lack of conspicuous ooplasmic segregation. *Biol. Bull.* 182, 77–96. doi: 10.2307/1542182
- Holland, L. Z., Holland, N. N., and Schubert, M. (2000). Developmental expression of *AmphiWnt1*, an amphioxus gene in the Wnt1/wingless subfamily. *Dev. Genes Evol.* 210, 522–524. doi: 10.1007/s004270000089
- Holland, N. D., and Holland, L. Z. (2006). Stage- and tissue-specific patterns of cell division in embryonic and larval tissues of amphioxus during normal development. *Evol. Dev.* 8, 142–149. doi: 10.1111/j.1525-142X.2006.00085.x
- Johnson, A. D., Crother, B. I., White, M. E., Patient, R., Bachvarova, R. F., Drum, M., et al. (2003b). Regulative germ cell specification in axolotl embryos: a primitive trait conserved in the mammalian lineage. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 358, 1371–1379. doi: 10.1098/rstb.2003.1331
- Johnson, A. D., Holland, L. Z., Holland, N. N., and Crother, B. I. (2003a). Evolution of predetermined germ cells in vertebrate embryos: implications for macroevolution. *Evol. Dev.* 5, 414–431. doi: 10.1046/j.1525-142X.2003.03048.x
- Johnson, A. D., Richardson, E., Bachvarova, R. F., and Crother, B. I. (2011). Evolution of the germ line-soma relationship in vertebrate embryos. *Reproduction* 141, 291–300. doi: 10.1530/REP-10-0474
- Juliano, C. E., Swartz, S. Z., and Wessel, G. M. (2010). A conserved germline multipotency program. *Development* 137, 4113–4126. doi: 10.1242/dev.047969
- Kawashima, T., Murakami, A. R., Ogasawara, M., Tanaka, K., Isoda, R., Sasakura, Y., et al. (2000). Expression patterns of *musashi* homologs of the ascidians, *Halocynthia roretzi* and *Ciona intestinalis*. *Dev. Genes Evol.* 210, 162–165. doi: 10.1007/s004270050024
- Kerner, P., Degnan, S. M., Marchand, L., Degnan, B. M., and Vervoort, M. (2011). Evolution of RNA-binding proteins in animals: insights from genome-wide analysis in the sponge *Amphimedon queenslandica*. *Mol. Biol. Evol.* 28, 2289–2303. doi: 10.1093/molbev/msr046

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fevo.2015.00156>

- Kon, T., Nohara, M., Yamanoue, Y., Fujiwara, Y., Nishida, M., and Nishikawa, T. (2007). Phylogenetic position of a whale-fall lancelet (Cephalochordata) inferred from whole mitochondrial genome sequences. *BMC Evol. Biol.* 7:127. doi: 10.1186/1471-2148-7-127
- Kozlin, V. V., and Kostyuchenko, R. P. (2015). *Vasa*, *PL10*, and *Piwi* gene expression during caudal regeneration of the polychaete annelid *Alitta virens*. *Dev. Genes Evol.* 225, 129–138. doi: 10.1007/s00427-015-0496-1
- Kuo, M. W., Wang, S. H., Chang, J. C., Chang, C. H., Huang, L. J., Lin, H. H., et al. (2009). A novel *puf-A* gene predicted from evolutionary analysis is involved in the development of eyes and primordial germ-cells. *PLoS ONE* 4:e4980. doi: 10.1371/journal.pone.0004980
- Leclerc, L., Jager, M., Barreau, C., Chang, P., Le Guyader, H., Manuel, M., et al. (2012). Maternally localized germ plasm mRNAs and germ cell/stem cell formation in the cnidarian *Clytia*. *Dev. Biol.* 364, 236–248. doi: 10.1016/j.ydbio.2012.01.018
- Leininger, S., Adamski, M., Bergum, B., Guder, C., Liu, J., Laplante, M., et al. (2014). Developmental gene expression provides clues to relationships between sponge and eumetazoan body plans. *Nat. Commun.* 5:3905. doi: 10.1038/ncomms4905
- MacNicol, M. C., Cragle, C. E., and MacNicol, A. M. (2011). Context-dependent regulation of Musashi-mediated mRNA translation and cell cycle regulation. *Cell Cycle* 10, 39–44. doi: 10.4161/cc.10.1.14388
- Okamoto, K., Nakatsukasa, M., Alié, A., Masuda, Y., Agata, K., and Funayama, N. (2012). The active stem cell specific expression of sponge Musashi homolog EflMsIA suggests its involvement in maintaining the stem cell state. *Mech. Dev.* 129, 24–37. doi: 10.1016/j.mod.2012.03.001
- Oulion, S., Bertrand, S., Belgacem, M. R., Le Petillon, Y., and Escrivà, H. (2012). Sequencing and analysis of the Mediterranean amphioxus (*Branchiostoma lanceolatum*) transcriptome. *PLoS ONE* 7:e36554. doi: 10.1371/journal.pone.0036554
- Özpolat, B. D., and Bely, A. E. (2015). Gonad establishment during asexual reproduction in the annelid *Pristina leidyi*. *Dev. Biol.* 405, 123–136. doi: 10.1016/j.ydbio.2015.06.001
- Potten, C. S., Booth, C., Tudor, G. L., Booth, D., Brady, G., Hurley, P., et al. (2003). Identification of a putative intestinal stem cell and early lineage marker; musashi-1. *Differentiation* 71, 28–41. doi: 10.1046/j.1432-0436.2003.700603.x
- Rebscher, N., Lidke, A. K., and Ackermann, C. F. (2012). Hidden in the crowd: primordial germ cells and somatic stem cells in the mesodermal posterior growth zone of the polychaete *Platynereis dumerillii* are two distinct cell populations. *EvoDevo* 3:9. doi: 10.1186/2041-9139-3-9
- Rebscher, N., Zelada-González, F., Banisch, T. U., Raible, F., and Arendt, D. (2007). Vasa unveils a common origin of germ cells and of somatic stem cells from the posterior growth zone in the polychaete *Platynereis dumerillii*. *Dev. Biol.* 306, 599–611. doi: 10.1016/j.ydbio.2007.03.521
- Rosner, A., Moiseeva, E., Rinkevich, Y., Lapidot, Z., and Rinkevich, B. (2009). Vasa and the germ line lineage in a colonial urochordate. *Dev. Biol.* 331, 113–128. doi: 10.1016/j.ydbio.2009.04.025
- Schubert, M., Holland, L. Z., and Holland, N. D. (2000). Characterization of an amphioxus wnt gene, *AmphiWnt11*, with possible roles in myogenesis and tail outgrowth. *Genesis* 27, 1–5. doi: 10.1002/1526-968X(200005)27:13.3.CO;2-V
- Schubert, M., Holland, L. Z., Stokes, M. D., and Holland, N. D. (2001). Three amphioxus Wnt genes (*AmphiWnt3*, *AmphiWnt5*, and *AmphiWnt6*) associated with the tail bud: the evolution of somitogenesis in chordates. *Dev. Biol.* 240, 262–273. doi: 10.1006/dbio.2001.0460
- Seaver, E. C., Thamm, K., and Hill, S. D. (2005). Growth patterns during segmentation in the two polychaete annelids, *Capitella* sp. I and *Hydrodries elegans*: comparisons at distinct life history stages. *Evol. Dev.* 7, 312–326. doi: 10.1111/j.1525-142X.2005.05037.x
- Shibata, N., Umesono, Y., Orii, H., Sakurai, T., Watanabe, K., and Agata, K. (1999). Expression of *vasa*(*vas*)-related genes in germline cells and totipotent somatic stem cells of planarians. *Dev. Biol.* 206, 73–87. doi: 10.1006/dbio.1998.9130
- Shibata, S., Umei, M., Kawahara, H., Yano, M., Makino, S., and Okano, H. (2012). Characterization of the RNA-binding protein Musashil1 in zebrafish. *Brain Res.* 1462, 162–173. doi: 10.1016/j.brainres.2012.01.068
- Siddall, N. A., Kalcina, M., Johanson, T. M., Monk, A. C., Casagrande, F., Been, R. P., et al. (2012). *Drosophila Rbp6* is an orthologue of vertebrate Msi-1 and Msi-2, but does not function redundantly with dMsi to regulate germline stem cell behaviour. *PLoS ONE* 7:e49810. doi: 10.1371/journal.pone.0049810
- Siddall, N. A., McLaughlin, E. A., Marriner, N. L., and Hime, G. R. (2006). The RNA-binding protein Musashi is required intrinsically to maintain stem cell identity. *Proc. Natl. Acad. Sci. U.S.A.* 103, 8402–8407. doi: 10.1073/pnas.0600906103
- Siebert, S., Goetz, F. E., Church, S. H., Bhattacharyya, P., Zapata, F., Haddock, S. H. D., et al. (2015). Stem cells in *Nanomia bijuga* (Siphonophora), a colonial animal with localized growth zones. *EvoDevo* 6, 1. doi: 10.1186/s13227-015-0018-2
- Somorjai, I., Bertrand, S., Camasses, A., Haguenauer, A., and Escrivà, H. (2008). Evidence for stasis and not genetic piracy in developmental expression patterns of *Branchiostoma lanceolatum* and *Branchiostoma floridae*, two amphioxus species that have evolved independently over the course of 200 Myr. *Dev. Genes Evol.* 218, 703–713. doi: 10.1007/s00427-008-0256-6
- Somorjai, I. M., Escrivà, H., and Garcia-Fernández, J. (2012b). Amphioxus makes the cut Again. *Commun. Integr. Biol.* 5, 499–502. doi: 10.4161/cib.21075
- Somorjai, I. M., Somorjai, R. L., Garcia-Fernández, J., and Escrivà, H. (2012a). Vertebrate-like regeneration in the invertebrate chordate amphioxus. *Proc. Natl. Acad. Sci. U.S.A.* 109, 517–522. doi: 10.1073/pnas.1100045109
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2-a multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033
- Wilson, V., Olivera-Martinez, I., and Storey, K. G. (2009). Stem cells, signals and vertebrate body axis extension. *Development* 136, 1591–1604. doi: 10.1242/dev.021246
- Wu, H. R., Chen, Y. T., Su, Y. H., Luo, Y. J., Holland, L. Z., and Yu, J. K. (2011). Asymmetric localization of germline markers *Vasa* and *Nanos* during early development in the amphioxus *Branchiostoma floridae*. *Dev. Biol.* 353, 147–159. doi: 10.1016/j.ydbio.2011.02.014
- Yajima, M., and Wessel, G. M. (2015). Essential elements for translation: the germline factor *Vasa* functions broadly in somatic cells. *Development* 142, 1960–1970. doi: 10.1242/dev.118448
- Yu, J. K., Wang, M. C., Shin, I. T., Kohara, Y., Holland, L. Z., Satoh, N., et al. (2008). A cDNA resource for the cephalochordate amphioxus *Branchiostoma floridae*. *Dev. Genes Evol.* 218, 723–727. doi: 10.1007/s00427-008-0228-x
- Yue, J. X., Li, K. L., and Yu, J. K. (2015). Discovery of germline-related genes in Cephalochordate amphioxus: a genome wide survey using genome annotation and transcriptome data. *Mar. Genomics* 24, 147–157. doi: 10.1016/j.margen.2015.03.010
- Yue, J. X., Yu, J. K., Putnam, N. H., and Holland, L. Z. (2014). The transcriptome of an amphioxus, *Asymmetron lucayanum*, from the Bahamas: a window into chordate evolution. *Genome Biol. Evol.* 6, 2681–2696. doi: 10.1093/gbe/evu212
- Zhang, Q. J., Luo, Y. J., Wu, H. R., Chen, Y. T., and Yu, J. K. (2013). Expression of germline markers in three species of amphioxus supports a preformation mechanism of germ cell development in cephalochordates. *EvoDevo* 4:17. doi: 10.1186/2041-9139-4-17
- Zhu, W., Pao, G. M., Satoh, A., Cummings, G., Monaghan, J. R., Harkins, T. T., et al. (2012). Activation of germline-specific genes is required for limb regeneration in the Mexican axolotl. *Dev. Biol.* 370, 42–51. doi: 10.1016/j.ydbio.2012.07.021
- Conflict of Interest Statement:** 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 © 2016 Dailey, Febrero Planas, Rossell Espier, Garcia-Fernández and Somorjai. 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) or licensor 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.**

Pax2/5/8 and Pax6 alternative splicing events in basal chordates and vertebrates: a focus on paired box domain

Peter Fabian, Iryna Kozmikova, Zbynek Kozmik and Chrysoula N. Pantzartzi *

Department of Transcriptional Regulation, Institute of Molecular Genetics, Prague, Czech Republic

OPEN ACCESS

Edited by:

Hector Escrivá,

Centre National de la Recherche
Scientifique, France

Reviewed by:

Manuel Irimia,

Centre for Genomic Regulation, Spain
Simona Candiani,
University of Genoa, Italy

*Correspondence:

Chrysoula N. Pantzartzi,

Department of Transcriptional
Regulation, Institute of Molecular
Genetics, Videnska 1083,
Prague 14220, Czech Republic
chrysoula.pantzartzi@img.cas.cz

Specialty section:

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

Received: 28 May 2015

Accepted: 15 June 2015

Published: 02 July 2015

Citation:

Fabian P, Kozmikova I, Kozmik Z and
Pantzartzi CN (2015) Pax2/5/8 and
Pax6 alternative splicing events in
basal chordates and vertebrates: a
focus on paired box domain.
Front. Genet. 6:228.
doi: 10.3389/fgene.2015.00228

Paired box transcription factors play important role in development and tissue morphogenesis. The number of *Pax* homologs varies among species studied so far, due to genome and gene duplications that have affected PAX family to a great extent. Based on sequence similarity and functional domains, four Pax classes have been identified in chordates, namely Pax1/9, Pax2/5/8, Pax3/7, and Pax4/6. Numerous splicing events have been reported mainly for *Pax2/5/8* and *Pax6* genes. Of significant interest are those events that lead to Pax proteins with presumed novel properties, such as altered DNA-binding or transcriptional activity. In the current study, a thorough analysis of *Pax2/5/8* splicing events from cephalochordates and vertebrates was performed. We focused more on *Pax2/5/8* and *Pax6* splicing events in which the paired domain is involved. Three new splicing events were identified in *Oryzias latipes*, one of which seems to be conserved in Acanthomorphata. Using representatives from deuterostome and protostome phyla, a comparative analysis of the *Pax6* exon-intron structure of the paired domain was performed, during an attempt to estimate the time of appearance of the *Pax6(5a)* mRNA isoform. As shown in our analysis, this splicing event is characteristic of Gnathostomata and is absent in the other chordate subphyla. Moreover, expression pattern of alternative spliced variants was compared between cephalochordates and fish species. In summary, our data indicate expansion of alternative mRNA variants in paired box region of *Pax2/5/8* and *Pax6* genes during the course of vertebrate evolution.

Keywords: Pax2/5/8, Pax6, alternative splicing, paired domain, splice variants

Introduction

Transcription factors encoded by genes of the paired box (PAX) family are highly conserved throughout metazoan phyla and hold a vital role in embryonic development. The association of different PAX subfamilies with organ and tissue morphogenesis, such as the thymus, central nervous system (CNS), enteric nervous system, kidneys, ear, thyroid, neural crest, vertebrae, and midbrain-hindbrain boundary (MHB) formation has been the object of various studies (summarized in Noll, 1993; Chi and Epstein, 2002; Paixao-Cortes et al., 2013; Blake and Ziman, 2014). Certain members of the PAX family are characterized as the master control genes for eye morphogenesis (Gehring, 1996, 2002, 2012; Kozmik, 2008; Klimova and Kozmik, 2014) and genetic defects are linked to the onset of eye-related diseases, e.g., small eye in mouse or aniridia in human.

Mutations in *Pax* genes are also correlated with diseases of the kidney and the CNS, but various types of cancer as well (see Chi and Epstein, 2002; Lang et al., 2007; Paixao-Cortes et al., 2013; Blake and Ziman, 2014 and references therein).

Gene and genome duplications, followed by gene losses, helped shape PAX gene family, leading to a varying number of *Pax* homologs in metazoan phyla studied so far (reviewed in Noll, 1993; Breitling and Gerber, 2000; Hoshiyama et al., 2007; Paixao-Cortes et al., 2013). Different subfamilies have been identified and are classified according to the similarity in sequence, functional domains they possess as well as their expression patterns (see Stuart et al., 1994; Blake and Ziman, 2014 and references therein). *PaxB* is apparently the oldest member and is present in sponges and cnidarians (Kozmik et al., 2003; Hoshiyama et al., 2007; Hill et al., 2010). *Pox neuro* and single genes from Pax1/9, Pax2/5/8, Pax4/6, and Pax3/7 classes are present in cephalochordates (Short and Holland, 2008; Takatori et al., 2008). *Pox neuro* is present in *Drosophila* (Bopp et al., 1989), but is lost in the lineages of tunicates and vertebrates. Two *Pax2/5/8* genes are present in urochordates, due to a duplication prior to ascidian and larvacean diversification (Wada et al., 2003; Canestro et al., 2005). As a result of two rounds of whole-genome duplication (Escrivá et al., 2002; Putnam et al., 2008) and subsequent gene losses, the coelacanth *Latimeria chalumnae* possesses nine *Pax* genes, i.e., discrete gene copies from each of the Pax1/9, Pax2/5/8, Pax3/7, and Pax4/6 classes, suggesting that all members of the PAX family were present in the ancestor that gave rise to the tetrapod lineage (Paixao-Cortes et al., 2013). More than nine genes are present in teleost fishes (reviewed in Ravi et al., 2013), due to the so-called third round of genome duplication (Jaillon et al., 2004; Van De Peer, 2004).

All *Pax* proteins contain a DNA-binding domain in their N-terminus, known as the paired domain (PD), as well as transactivation and inhibitory domains in their C-terminus. The PD consists of 128 aminoacids and is made up by two helix-turn-helix (HTH) subdomains, known as PAI and RED or N-terminal and C-terminal respectively, joined through a linker (Czerny et al., 1993; Xu et al., 1999). The Pax4/6 class contains an additional DNA-binding homeodomain, while the Pax2/5/8 possesses a partial homeodomain and an octapeptide motif, the latter known to interact with members of the Groucho family of co-repressors (Eberhard et al., 2000; Kreslova et al., 2002). Classes Pax1/9 and Pax3/7 both contain the octapeptide motif, yet the former lacks the homeodomain (Chi and Epstein, 2002). *Pax* loci which encode for proteins with truncated paired domain have been identified in *Drosophila*, *C. elegans*, as well as representatives of Hemichordata and Echinodermata (Chisholm and Horvitz, 1995; Cinar and Chisholm, 2004; Howard-Ashby et al., 2006; Friedrich and Caravas, 2011; Ravi et al., 2013).

Gene/genome duplication is a driving force for evolution (Bergthorsson et al., 2007; Maere and Van De Peer, 2010) and this could be nicely exemplified by the well-studied PAX gene family, where many duplicates were preserved in the genome and obtained new functions and new domains of expression (neofunctionalization), or original gene functions were partitioned (subfunctionalization) between duplicates

(Pfeffer et al., 1998; Bassham et al., 2008; Kleinjan et al., 2008, reviewed in Holland and Short, 2010).

At posttranscriptional level, alternative splicing is also known to promote evolution, protein diversity and development of novel functions in eukaryotic genomes (reviewed in Nilsen and Graveley, 2010; Chen et al., 2012; Kelemen et al., 2013). In fact, it has been suggested that in the case of *Pax* genes the impact of alternative splicing on functional motifs is more intense than gene duplication and subsequent divergence of the duplicated genes (Short and Holland, 2008). It has been shown that alternative splicing usually takes place in a tissue or developmental stage-specific manner (Wang et al., 2008a; Kelemen et al., 2013). Depending on which exonic segments are cut-out and whether intronic regions are retained in the transcripts, splicing events can be clustered into four major groups, namely (1) exon skipping, (2) alternative 3'-, (3) alternative 5'-splice sites, and (4) intron-inclusion (Koralewski and Krutovsky, 2011). These four types of events can occur independently or in combination with other incidents, such as mutually exclusive exons, alternative initiation and alternative polyadenylation (Wang et al., 2008a; Koralewski and Krutovsky, 2011; Kelemen et al., 2013).

Alternative splicing of *Pax* genes has been observed in various species, from protostomes (Fu and Noll, 1997; Cinar and Chisholm, 2004) to cephalochordates (Glardon et al., 1998; Kozmik et al., 1999; Short and Holland, 2008; Holland and Short, 2010; Short et al., 2012) and vertebrates (Kozmik et al., 1993, 1997; Poleev et al., 1995; Heller and Brandli, 1997, 1999; Lun and Brand, 1998; Short et al., 2012), where multiple incidents from all major groups of splicing events were present. In principle, splice forms seem to have diverged between lineages, some of them are species- or genus-specific (Heller and Brandli, 1999; Short et al., 2012), nevertheless several splice isoforms seem to be evolutionary conserved (Kwak et al., 2006; Short and Holland, 2008; Short et al., 2012; Ravi et al., 2013). The majority of reported splice events regards the transactivation and inhibitory domain in the C-terminal part of the *Pax* proteins (Kozmik et al., 1993; Ward et al., 1994; Nornes et al., 1996; Tavassoli et al., 1997; Kreslova et al., 2002; Robichaud et al., 2004), still there is an increasing number of events affecting paired domain and consequently DNA binding capacity (Kozmik et al., 1993, 1997; Zwollo et al., 1997; Short and Holland, 2008; Short et al., 2012). One such example is the *Pax6(5a)* isoform, where inclusion of exon 5a (Walther and Gruss, 1991; Glaser et al., 1992; Puschel et al., 1992) leads to a protein with interrupted paired domain that recognizes an altered DNA binding sequence (Epstein et al., 1994). Apparently, this event is quite conserved among vertebrate lineages, with differences in size and peptide sequence of exon 5a between fish and tetrapods (Ravi et al., 2013). In some cases, alternatively spliced *Pax* isoforms exhibit temporally and spatially differentiated expression patterns (Kozmik et al., 1993, 1997; Heller and Brandli, 1997; Short and Holland, 2008) and have been associated with cancer and genetic disorders (reviewed in Wang et al., 2008b; Holland and Short, 2010).

Previous studies have shown that any insertion in the conserved paired domain, no matter if it is a single-aminoacid extension or a whole exon cassette, modifies DNA binding

capacity and attributes differentiated functions to the isoforms bearing the insertion (Kozmik et al., 1997; Azuma et al., 2005).

In the present study, we sought to identify splicing events in *Pax2/5/8* and *Pax6* classes that affect the paired domain and study the expression patterns of these alternative spliced transcripts. We identified three new splicing events in *Oryzias latipes Pax2* genes, one of which seems to be highly conserved in Acanthomorphata. We detected a re-occurring splicing event in *O. latipes* and *Danio rerio* *Pax6* genes generating the exon 5a insertion. Using our data set we tried to elucidate the time point at which the exon 5a-insertion appeared and the extent of its conservation in various phylogenetic groups.

Materials and Methods

Data Collection and *de novo* Gene Annotation

Nucleotide and aminoacid sequences for annotated *Pax2/5/8* and *Pax6* genes were obtained using proper keywords, through NCBI GenBank (Benson et al., 2013), ENSEMBL release 78 (Cunningham et al., 2015), the UCSC Genome Browser database (Karolchik et al., 2014), the SpBase (Sea Urchin Genome Database, Cameron et al., 2009) and the JGI (Grigoriev et al., 2012). The retrieved *Pax* genes were crosschecked using GENSCAN (Burge and Karlin, 1997), BLASTx and version 0.9 of the NNSPLICE splice predictor (Reese et al., 1997).

For various taxonomic groups (e.g., Chondrichtyes: Holopocephali) there are available genomes, but no *Pax* genes are annotated in public databases. In order to include representatives from these groups, we conducted BLAST searches against the NCBI GenBank and wgs subdivision (Trace archive), using known homologs from Deuterostomia species. Where required, small contigs or scaffolds were fused using Merger of the EMBOSS software suite (Rice et al., 2000) and gene structure was defined using GENSCAN (Burge and Karlin, 1997), BLASTx and splice predictor (Reese et al., 1997). ScanProsite (De Castro et al., 2006) was used to detect conserved functional domains in newly identified genes. In addition, adjacent genes of the *de novo* predicted *Pax* genes were also predicted/annotated and gene order was compared to known *Pax* syntenic regions through the Genomicus website v78.01 (Louis et al., 2013). In all cases, the NNSPLICE was used for the prediction of possible alternative acceptor and donor sites.

PipMaker (Schwartz et al., 2000) was used along with BLAST, in order to locate putative sequence conservation among species. Alignment of *Pax6* paired domains from various species was performed using the MUSCLE algorithm (Edgar, 2004) included in Mega version 5.0 (Tamura et al., 2011).

Expressed Sequenced Tags (ESTs) Retrieval and Analysis

In order to validate already annotated or newly predicted *Pax* homologs, BLAST searches were performed against the ESTs subdivision. Collected ESTs were aligned with predicted coding sequences from genome analyses and mRNA sequences—if available—in order to detect putative splicing events not recognized so far.

Animal Collection

Specimens of *Branchiostoma floridae* were collected from Old Tampa Bay, Florida, USA. Gametes were obtained and embryos were raised, as previously described (Holland and Yu, 2004). *B. lanceolatum* adults were collected in Banyuls-sur-Mer, France, prior to summer breeding season and raised in the lab until spawning. The spawning of males and females was induced by temperature shift (Fuentes et al., 2007). *B. lanceolatum* and *B. floridae* embryos were developed at 16°C and 26°C, respectively. Embryos of inbred strains of *Oryzias latipes* (Cab) and *Danio rerio* (AB) were used for all experiments. *O. latipes* and *D. rerio* embryonic stages were determined according to Iwamatsu (2004) and Kimmel et al. (1995). Housing of animals and *in vivo* experiments were performed after approval by the Animal Care Committee of the Institute of Molecular Genetics (study ID#36/2007) and in compliance with national and institutional guidelines (ID#12135/2010-17210).

RNA Isolation and Reverse Transcription

Total RNA was isolated from embryos using the Trizol reagent (Ambion). Random-primed cDNA was prepared in a 20 µl reaction from 500 ng of total RNA using SuperScript VILO cDNA Synthesis kit (Invitrogen).

Screen for Alternative Splicing and RT-PCR Analysis

cDNA was subjected to PCR using DreamTaq polymerase (Thermo Scientific) for 30 cycles under the following conditions: 1 min at 98°C, 30 s at 60°C, 30 s at 72°C. Primers for this analysis are provided in Table 1. PCR products were analyzed on 2.5% agarose gel and bands of interest were eluted, cloned to pCR-Blunt II (Invitrogen) and sequenced (GATC Biotech sequencing service, Germany).

Results

Pax2/5/8 Splicing Events in Chordates

Exhaustive search through databases and literature, in combination with *de novo* analysis of available ESTs and mRNA sequences (Table S1), revealed numerous splicing events in chordate members of the *Pax2/5/8* class (Figure 1). Some of these events seem to characterize specific orthologs and are present in cephalochordates, fish and mammals (e.g., exon 2 of *Pax5* gene), while others are much less conserved (e.g., exon 3a of mouse *Pax5*). *Branchiostoma floridae* appears to experience the largest number of splicing events, however no event of insertion in the paired domain has been reported so far and no such event could be predicted using splice prediction software or available ESTs/mRNA sequences.

A single *Oryzias latipes Pax2* gene, namely *OlPax2.2*, located on chromosome 19 (NC_019877), has been used in previous studies (Paixao-Cortes et al., 2013). Our search through NCBI revealed an annotated Pax-2a-like gene on chromosome 15 (NW_004088010.1). It must be noted that the aminoacid sequence encoded by the first half of this gene exhibits no similarity to the paired domain of other *Pax* genes and it is not supported by ESTs, a fact that indicates an erroneous gene

TABLE 1 | Summary of primers used in RT-PCR reactions in the present study.

Species	Gene	Sequence 5'→3'
<i>Branchiostoma</i>	Pax258	F: AATGGGTGCGTGTCCAAGATT R: AACGCTGGGATGTTCTCATT
	Pax46	F: GTCCCACGGCTGTGTCAGTAA R: TCGTTGTACCAAATGCCTTCC
<i>Oryzias latipes</i>	Pax2.1	F: GCAGCGGATCGTGGAGCTT R: GAACATAGTGGGTTTGGCGC R*: CTCTGAAATGCCCTCTGATAG
	Pax2.2	F: GCAGCGGATTGTGGAGCTG R: AAACATTGTGGGTTTGTCTT
	Pax6.1	F: CCACCAGGAGAAAATAGTGGAACTT R: ATCTGCTCACGCAGCCGTTGGAT
	Pax6.3	F: AACCAGCTGGGGGGGTATTTGT R: CTCCAATGGCCCCGGGGACGG
<i>Danio rerio</i>	Pax2.1	F: ACCAGCTAGGAGGGGTGTT R: CCAGGCGAACATTGTAGGAT
	Pax2.2	F: CGACAGCTGGGGTCAGTC R: CGAACATGGGGGGATTTGT
	Pax6.1a	F: CCCGACTCCACGAGACAGAAAATAGTT R: ACCCAAGATTACTCACGCAGCCGTTG
	Pax6.1b	F: CCGGACTCCACAGACAGAAAGATCGTC R: CCAGAACATTGCTCACGCACCCATTC

For *Branchiostoma* species the same set of primers was used for each gene class amplified. *Oryzias latipes* Pax2.1 reverse primer marked with asterisk was used to amplify alternative exon 2a along with Pax2.1 forward primer.

prediction, caused by a non-sequenced area in this genomic region. We assume that the first coding exon as well as the two exons coding for the paired domain of *Olpax2.1* are located in this non-sequenced region. The record of an unplaced scaffold (NW_004093539, Table S1) was retrieved through BLAST. It apparently corresponds partly to the non-sequenced region of chromosome 15 and contains the 5' UTR, the first exon of *Olpax2.1* gene and part of the first intron.

In order to retrieve more information on *Olpax2.1* and to elucidate how many *Olpax2* (either *Pax2.1* or *Pax2.2*) transcripts exist, we searched for different ESTs and mRNA sequences using *Olpax2.2* and *Danio rerio Pax2.1* as queries. There are only two ESTs (AM320053 and AM321390) for *Olpax2.1* that contain the first coding exon, as well as the complete exon 2 and part of exon 3, encoding for N- and C- paired subdomains, respectively. Using the genomic scaffolds and available ESTs collectively (Table S1), an almost complete *Olpax2.1* gene was reconstructed (Figure 1). For *Olpax2.2* gene, one cDNA sequence and four ESTs were retrieved (Table S1), comparison of which revealed both 5' and 3' alternatively spliced parts of exon 2 encoding for the N-terminal of paired domain (see Figure 1).

Exon-to-exon comparison of *Pax2/5/8* genes between *O. latipes* and *D. rerio*, shows that in principle there is conservation in the sequence, number, size, and borders of exons and some indication for alternatively spliced exons in *Olpax2* genes (light gray boxes in Figure 1), which are not suggested by the available

ESTs. Retrieved ESTs encoding both *Olpax2.1* and *Olpax2.2* support a 5' alternative splicing donor site in exon 2 (Figure 1, Table S1); the insertion is 12 bp long, and results to four additional aminoacids, exactly at the beginning of the paired domain. The same insertion has been reported for *D. rerio Pax2.1* (Lun and Brand, 1998), as well.

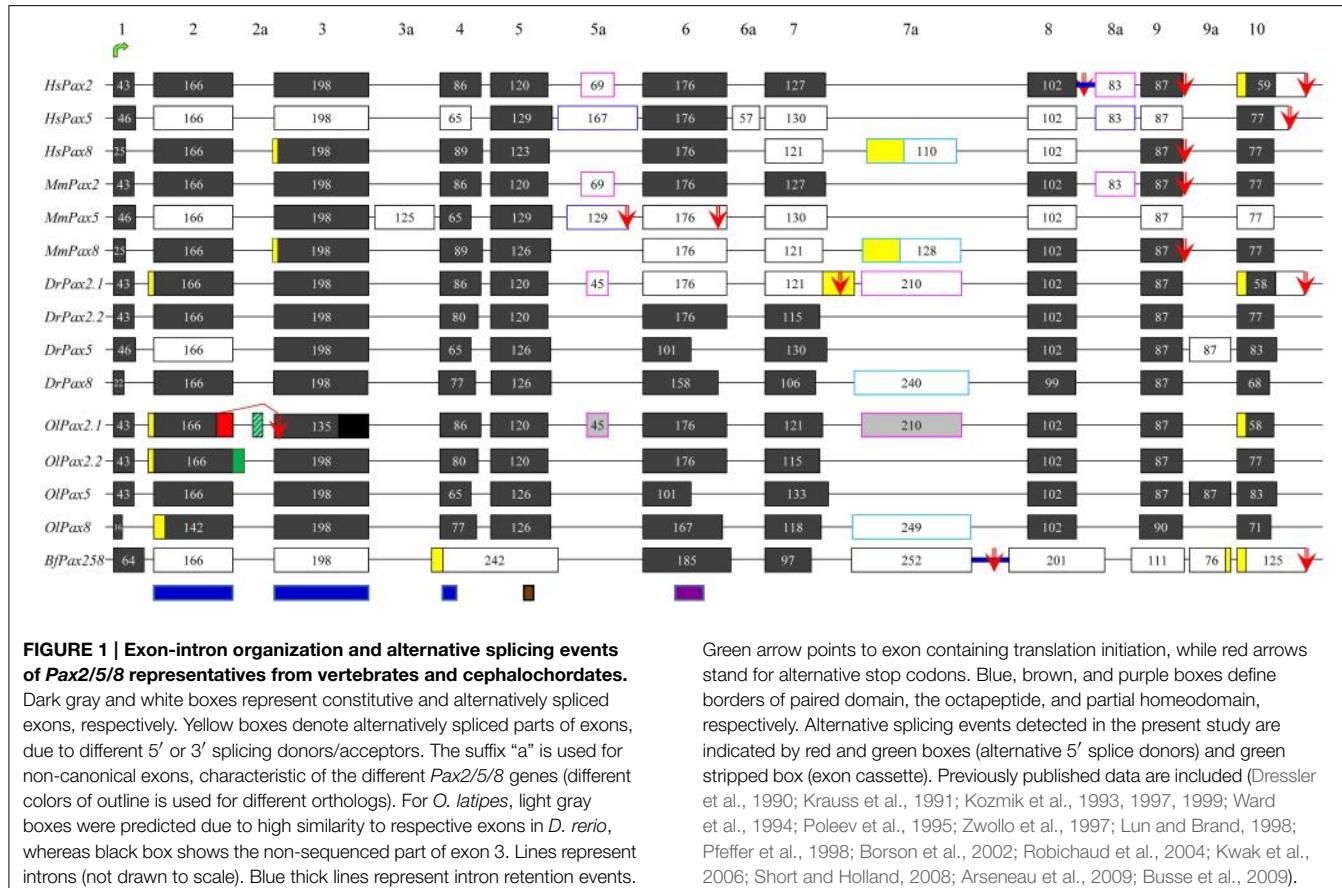
In the present study we identified two splicing events that lead to insertion of extra aminoacids in the paired domain of *Olpax2* genes. More specifically, an alternatively spliced 21-bp exon was detected between exons 2 and 3 of fish *Pax2.1* genes, which is annotated in some species (e.g., *Poecilia reticulata* and *Maylandia zebra*). This exon could not be detected *in silico* in the genome of *O. latipes*, due to the fact that the intron between exons 2 and 3 is not sequenced. Through BLAST searches and *de novo* analysis of *Pax2.1* genes we spotted this exon in numerous representatives from different orders of Acanthomorphata (Table S2), while PipMaker alignment reveals a high degree of sequence conservation among compared species (Figure 2). A putative exon with proper splice sites has been identified in the respective genomic region of three Cyprinidae species (*D. rerio*, *Pimephales promelas*, and *Cyprinus carpio*). Even though this exon is highly conserved in these species, the encoded aminoacids are quite dissimilar from those of the Acanthomorphata 21-bp exon (Figure 2). In both cases, inclusion of this exon leads to an alternative transcript, which incorporates seven extra aminoacids toward the end of the a3 helix of the PAI subdomain (Figure 2).

In regard to *Olpax2.2*, the available mRNA sequence in GenBank and our analysis revealed a 24-bp in-frame extension at the 3' end of exon 2 (Figures 1, 3, Table S1), which does not alter the downstream translation (Figure 3). This isoform, to which we will refer as *Olpax2.2(ext24+)*, is due to an alternative splicing donor downstream the canonical one (Figure 3). In *D. rerio*, a similar isoform is neither supported by splicing prediction software nor by available mRNA sequences.

It should be noted that splicing prediction analysis of *Olpax2.1* gene revealed the presence of an alternative donor site in exon 2, upstream the canonical one. Deletion of 35 bp at the 3' end of exon 2 causes a frameshift and leads consequently to a premature stop codon at the beginning of exon 3 (Figure S1). A similar donor site was not *in silico* identified in *D. rerio* (data not shown).

Comparative Analysis of 5a-exon Insertion in Pax6 Paired Domain

In teleosts and tetrapods studied so far, the major part of Pax6 paired domain is encoded by two exons, responsible for the N- and C-subdomains, with a size of 131 and 216 bp, respectively. In Tetrapoda and in the *Pax6.1* copy of teleosts, a small exon of varying size (36–42 bp), namely 5a, has been shown to be included in alternative transcripts, causing an in-frame insertion in the paired domain (Ravi et al., 2013). We wanted to identify at which point of evolutionary history this exon appeared and investigate a putative correlation of the appearance of this exon with the exon/intron organization of *Pax6* homologs. For this reason, already annotated *Pax6* homologs were collected from public databases and available genomes and EST sequences from non-jawed vertebrates,



cephalochordates, tunicates, hemichordates, echinoderms, as well as *Drosophila* and *C. elegans* were analyzed (Table S3).

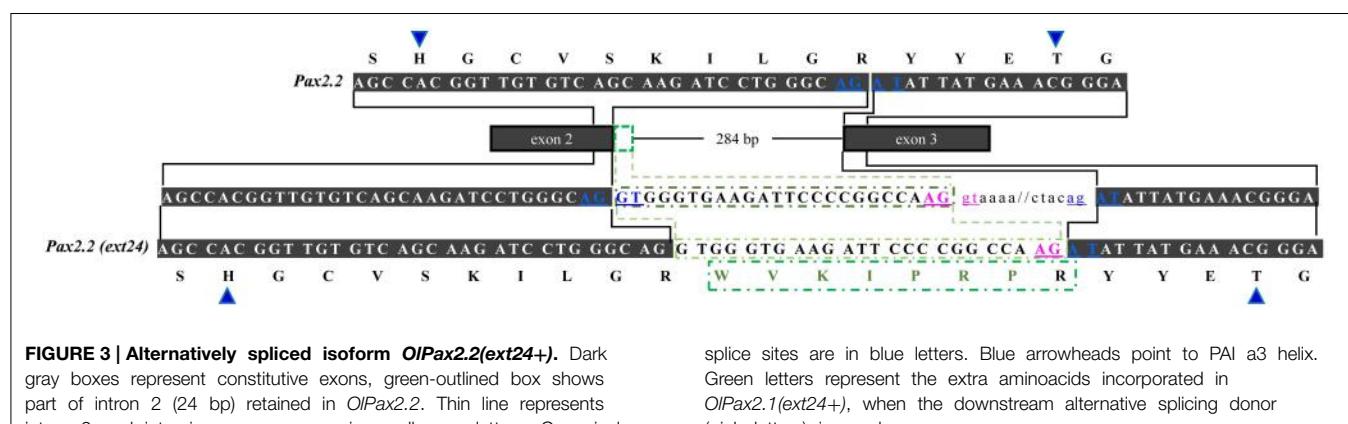
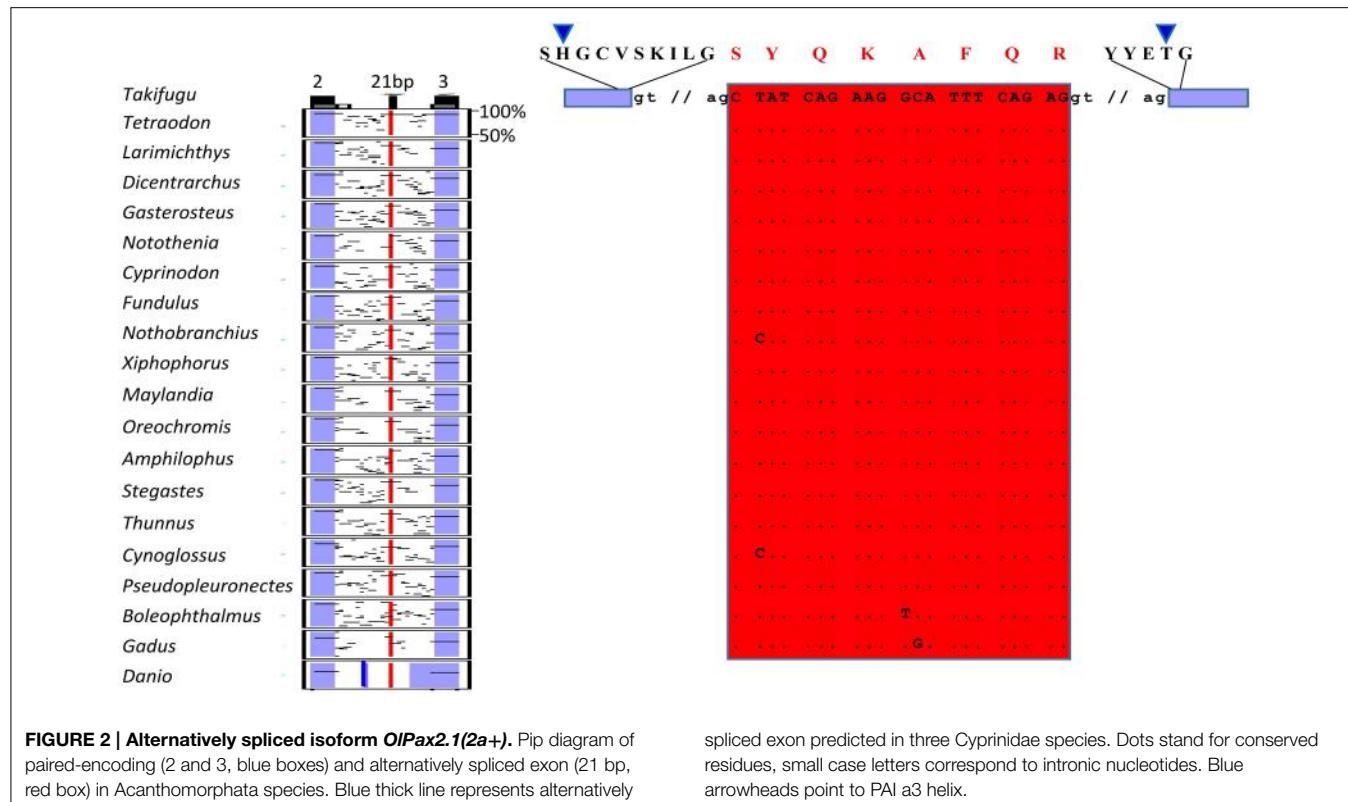
Analysis of a genomic scaffold from *Leucoraja erinacea*, containing the *Pax6* ortholog, provides evidence that besides Holocephali (Ravi et al., 2013), the 5a exon is also present in Elasmonbranchii, the second subclass of Chondrichthyes. Unfortunately, no genomes from hagfishes are publicly available, yet the two *Pax6* mRNA sequences that were retrieved from *Eptatretus bergeri* (Table S3), do not provide any indication of an exon5a-like insertion in the paired domain.

In regard to Hypoartia, genomic scaffolds containing parts of the *Pax6* genes from *Petromyzon marinus* and *Lethenteron japonicum* were retrieved and analyzed, along with two *Pax6* mRNA from the species *L. japonicum* and *Lampetra fluviatilis*. Apparently, there are more than one *Pax6* genes in the *L. japonicum* genome, yet the low genome coverage for both *P. marinus* and *L. japonicum* (5 \times and 20 \times , respectively) does not allow for safer conclusions. In all cases, there is no evidence for the existence of the exon 5a in Hypoartia.

Existing models for *Stonylocentrotus purpuratus* predict two truncated neighboring *Pax6* proteins that contain either the paired or the homeobox domain. Taking into account the provided information in SpBase mentioning that the gene models are incomplete and an intervening sequence appears to be missing in a scaffold gap in the Spur3.1 assembly (Howard-Ashby

et al., 2006; Cameron et al., 2009), we re-evaluated the prediction and tried to re-construct the *SpPax6* homolog.

Our focus was on the exon-intron structure in the region of paired domain (Figure 4). It is apparent, from our analysis, that the size and borders of the two exons encoding the main part of paired domain underwent various changes in different taxonomic groups (Figure 4). More specifically, in basal deuterostomes, such as Echinodermata and Hemichordata, there is one large exon with a size of 347 bp encoding for the first 115 aminoacids of the paired domain. Therefore, there is no intervening non-coding region in the respective position to vertebrate's intron, or in other words "space" for insertion of an alternatively spliced exon in the paired domain. In cephalochordates and tunicates, this large exon has split into two exons, first of which has a size of 166 bp, still larger than the N-subdomain encoding exon in vertebrates, and the second one is 181 bp long, slightly smaller than the respective exon in vertebrates. The 166 bp exon encodes a peptide including the first four aminoacids of a3 helix of the N-subdomain (Xu et al., 1995) and ends shortly after the position where the exon 5a is inserted, i.e., downstream of a3 helix. Thorough *in silico* search in the intronic sequence flanked by the paired-encoding exons in tunicates and cephalochordates did not reveal a putative alternatively spliced exon cassette similar to the exon 5a (Figure 4, Figure S2). It seems that paired-encoding



exons obtained fixed borders and size of 131 and 216 bp for PAI and RED, respectively before diversification of cyclostomes and preserved them throughout vertebrates, nonetheless “birth” of exon 5a probably appeared in Gnathostomata (**Figure 4**, Figure S2). It should be noted that the major re-arrangements concern the genomic region encoding for the PAI a3 helix, in contrast to the high conservation observed toward the C-subdomain.

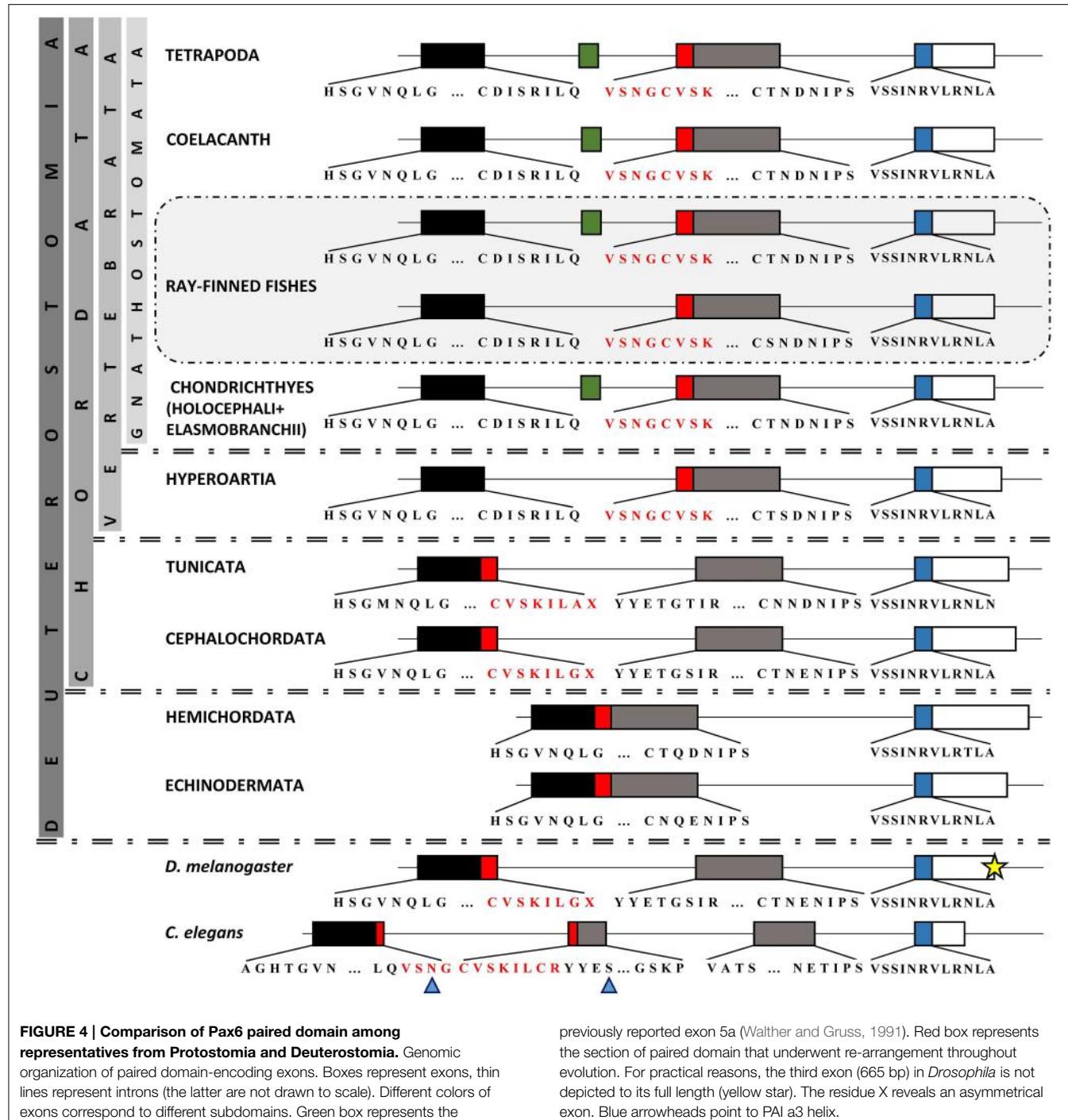
Developmental Expression of PAI-RED Isoforms

To verify *in silico* predicted alternative splice isoforms (**Figures 2–4**, Figure S1) and compare their expression across various developmental stages, we performed RT-PCR using

RNA from different embryonic stages of *Branchiostoma lanceolatum*, *B. floridae*, *Oryzias latipes*, and *Danio rerio* (**Figure 5**).

In agreement with our *in silico* analysis, RT-PCR (**Figure 5**) and DNA sequencing (data not shown) using primers located in exons encoding PAI and RED domains revealed that *B. lanceolatum* and *B. floridae* express single *Pax258* and *Pax46* isoforms.

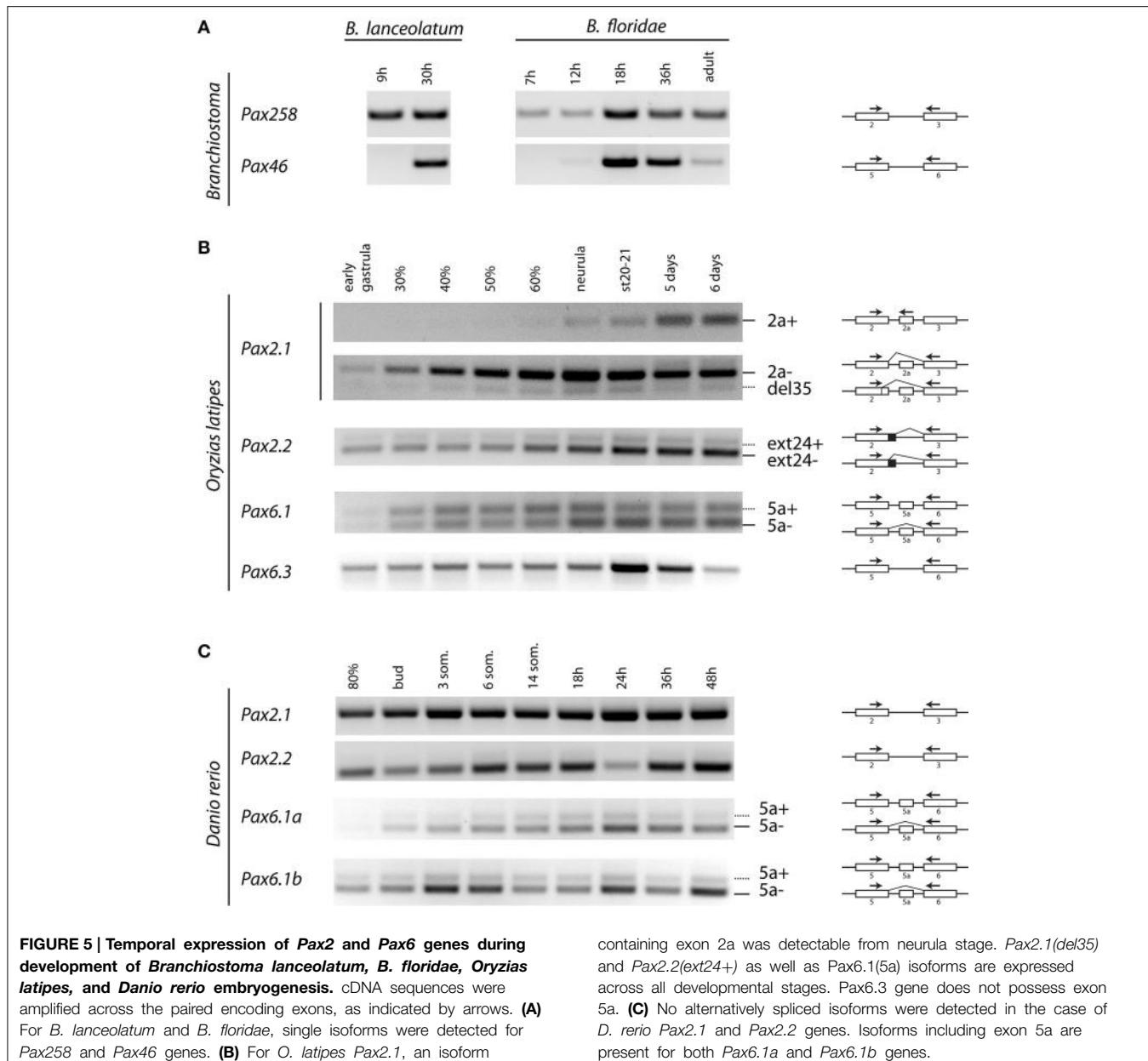
Three splice isoforms of *OIPax2* genes have been *in silico* predicted. Lack of information concerning intron 2 of *OIPax2.1* prohibited the *in silico* detection of the 21-bp exon, characteristic of Acanthomorphata. However, the presence of 21-bp exon in the *O. latipes* genome was experimentally verified, using a proper



set of primers (Table 1), one of which was specifically designed on the 21-bp exon characteristic of Acanthomorphata (Figure 2). Expression of the alternatively spliced *OIPax2.1(2a+)* isoform was detectable from neurula and later developmental stages (Figure 5). In contrast, the truncated isoform *OIPax2.1(del35)*, which results from a deletion in exon 2 and a premature stop codon in the 3rd exon (Figure S1), was present across all developmental stages (Figure 5). *OIPax2.1(del35)* was expressed at much lower level than dominant *OIPax2.1(2a-)* isoform.

Sequencing of this isoform revealed that it makes use of the alternative donor site in exon 2, as predicted by *in silico* analysis (Figure S1). The extended isoform *OIPax2.2(ext24+)* (Figure 3) was present at detectable level throughout the examined stages (Figure 5). In the case of *D. rerio* *Pax2.1* and *Pax2.2* genes, no alternative splice variants were detected, in agreement with splicing prediction analysis.

The alternatively spliced isoform *OIPax6.1(5a-)* was expressed approximately at the same level as isoform *OIPax6.1(5a+)*



(Figure 5). In agreement with previous studies (Ravi et al., 2013), *O. latipes* *Pax6.3* gene does not possess the equivalent of exon 5a. Variants bearing exon 5a were observed for both *D. rerio* *Pax6.1a* and *Pax6.1b* genes, and in both cases, expression level of isoform 5a- was relatively higher than isoform 5a+.

In silico and experimental data, collectively, demonstrate increased complexity of splicing events in vertebrate paired domain of *Pax* genes in comparison to cephalochordates.

Discussion

Paired box (*Pax*) genes encode for transcription factors that are considered key players in organogenesis and embryonic development. The presence of *Pax* genes in a variety of organisms

and the evolution of the PAX family has been the object of various studies (Hill et al., 2010; Paixao-Cortes et al., 2013; Ravi et al., 2013). Whole-genome duplications as well as lineage-specific gene duplications provide additional possibilities for diversified evolution and/or speciation (Bergthorsson et al., 2007; Maere and Van De Peer, 2010). These processes are considered to have played important role in shaping the number of *Pax* homologs in various taxonomic groups (see Paixao-Cortes et al., 2013; Ravi et al., 2013 and references therein), but same applies for alternative splicing, a posttranslational mechanism that also promoted evolution and complexity of *Pax* proteins (Glardon et al., 1998; Short et al., 2012).

In the present study, we wanted to evaluate the degree of alternative splicing taking place in various lineages, as well as

to identify splicing events that are either evolutionary conserved or characteristic of cephalochordates and not vertebrates or vice versa. For this purpose, we collected annotated homologs from Pax2/5/8 and Pax4/6 classes from public databases (NCBI, Ensembl, UCSC, JGI and SpBase). Furthermore, we analyzed *de novo* genomes, ESTs and mRNA sequences from different species, in order to enrich our dataset with taxonomic groups not present in previous studies. Our second focus was *Pax* isoforms from cephalochordates and vertebrates, that differ in the paired domain, and their expression patterns across different developmental stages.

Apart from partially reconstructing the *Oryzias latipes* *Pax2.1*, using available scaffolds and EST sequences, we identified three new splicing events in the *Pax2* genes of *O. latipes*. The *OlPax2.1(2a+)* isoform is reminiscent of the *5a* isoform found in *Pax6* homologs (Ravi et al., 2013; this study), as it incorporates a 21-bp in-frame-exon located in the intron between the two exons encoding for the paired domain. Our analysis showed that this exon is present in numerous species from various orders of Acanthomorphata and exhibits a high degree of conservation among compared species (Figure 2). We presume that sequence conservation of this mRNA splice form over a wide phylogenetic distance also implies conservation of this isoform's function. A similar exon in terms of location, yet quite divergent in terms of sequence, was *in silico* predicted only in three Cyprinidae species (*D. rerio*, *Pimephales promelas*, and *Cyprinus carpio*).

The second alternatively spliced isoform, namely *OlPax2.2(ext24+)*, results from the use of an alternative splice donor downstream the canonical one at the end of *OlPax2.2* exon 2. In this case, extra aminoacids are incorporated in the middle of a3 helix of the PAI subdomain, with no influence on the downstream sequence. A similar isoform could not be detected in *D. rerio*, neither experimentally nor *in silico*. The sequences surrounding the normal splice junctions of exon 2-intron 2 are highly conserved between *D. rerio* and *O. latipes* *Pax2.2* genes, yet there is no proper donor-acceptor site in the region of *D. rerio* (AG-AG) that corresponds to the alternative splice site of *O. latipes*.

Both *OlPax2.1(2a+)* and *OlPax2.2(ext24+)* transcripts bear an insertion in the recognition a3 helix of PAI subdomain. Previous studies on insertions in the paired domain of *Pax* genes have proven that, regardless of the number of the inserted aminoacids, disruption of this helix, which is responsible for all major groove DNA contacts of the N-terminal subdomain (Xu et al., 1995, 1999) is expected to inactivate the DNA-binding function of the N-terminal HTH motif, which subsequently leads to severe restriction in the DNA-binding sequence specificity of the paired domain (Kozmik et al., 1997).

The importance of alternative splicing as a mechanism for divergent evolution is established. In the case of *Pax* genes, the fact that insertions in the paired domain may preferentially guide *Pax* proteins, namely *Pax6(5a)* and *Pax8(S)*, to the control region of genes containing a modified binding site (5aCON-like sequence, Kozmik et al., 1997), in other words insertions add new target-genes in the repertoire of genes controlled by *Pax* genes,

may be indicative of a mechanism through which alternative splicing contributes to the increase of complexity at the level of protein function.

The isoform *OlPax2.1(del35)* makes use of an alternative 5' splicing donor, upstream of the normal splicing site in exon 2 (N-terminal of paired domain). As mentioned before, the exact junction sequence between exon 2 and intron 2 is not known, nevertheless, the sequence at the normal end of exon 2 (CAG) is in agreement with the optimal consensus for 5' splice sites (Stephens and Schneider, 1992), in contrast to the sequences at the alternative upstream 5' splice donor (CGG/GT, Figure S1). As it has been observed before for the *Pax8* gene (Kozmik et al., 1997), there is a higher abundance and constitutive splicing of the *Pax2* mRNA relative to the alternative transcript (Figure 5), a fact that could be attributed to different affinities by which the spliceosomes may recognize the two 5' donor sites. A similar truncated isoform could not be detected neither during *Pax2.1* transcript analysis of *D. rerio*, nor by *in silico* analysis.

The alternative isoform *OlPax2.1(del35)*, lacks the greater part of a3 helix of PAI domain and ends at a premature stop codon exactly at the beginning of exon 3. Truncated isoforms are not a rare phenomenon, given the fact that approximately 35% of alternatively spliced human transcripts have been found to contain a premature termination codon, rendering them as candidates for non-sense-mediated decay (Green et al., 2003; Lewis et al., 2003). It has been proposed that most low copy number alternative isoforms produced in human cells are likely to be non-functional, therefore we assume that this is also the case for *OlPax2.1(del35)*. Deletion of a3 helix has been observed in one of the *Pax6* isoforms in *B. floridae* (Glardon et al., 1998), yet this deletion does not influence downstream translation and hence its functionality. A 32 bp deletion in mouse is responsible for splotch phenotype in mouse (Epstein et al., 1991). In addition, there are accumulating reports about heterozygous deletions of parts of PAI subdomain in general or a3 helix in specific, most of which cause a frame shift and a premature stop codon (Schimmenti et al., 1997; Fletcher et al., 2005) and are correlated with diseases in human (e.g., renal-coloboma syndrome, oligomeganephronia).

In regard to the *Pax6* class and the alternative splice isoform *Pax6(5a)*, our analysis showed that an important re-arrangement of coding and non-coding sequences in the region of paired domain took place during evolution. Although conservation of the position of introns has been noted between highly divergent eukaryotes, the number and placement of the majority of introns are dynamically fluctuating during evolution (Hartung et al., 2002; Rogozin et al., 2003). In Hemichordata and Echinodermata, the exon-intron organization does not allow for any type of insertions in the paired domain. In other lineages compared, paired domain is encoded by exons disrupted by one or more introns. Incidents of intron gain and loss as well as intron sliding have been reported for various genes (Hartung et al., 2002), whereas the intron density, i.e., the average number of intron per gene does not necessarily coincide with the position of the genome on the evolutionary tree (Jeffares et al., 2006). We assume that the 5a insertion is characteristic of Gnathostomata.

Introns are required for alternative splicing and alternative splicing increases the size of the proteome, thus increasing the level of complexity in higher eukaryotes. Moreover, introns have been found to harbor many conserved non-coding elements, necessary for gene regulation (Irvine et al., 2008; Bhatia et al., 2014).

Unique isoforms were detected during expression pattern study of *Branchiostoma Pax2/8* and *Pax4/6* genes. This is in agreement with *in silico* analysis, during which no splicing events involving the paired domain were predicted. In contrast, new alternative spliced variants were identified for fish species. Previous studies have shown that there is no developmental regulation of paired domain alternative splice forms of *Pax6* and *Pax8*, as opposed to splicing events affecting the C-terminal sequences of *Pax8* protein (Kozmík et al., 1993, 1997). In principle, non-constitutive *Olpax2* isoforms are expressed at low levels, therefore at this stage, it is not easy to conclude as to the regulation of these isoforms. Nonetheless there is an indication of a temporal

regulation of *Olpax2.1(2a+)* isoform, which requires further investigation.

Acknowledgments

This study was supported by grant LH12047 from the Ministry of Education, Youth and Sports of the Czech Republic and the project “BIOCEV—Biotechnology and Biomedicine Centre of the Academy of Sciences and Charles University” (CZ.1.05/1.1.00/02.0109). The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fgene.2015.00228>

References

- Arseneau, J. R., Laflamme, M., Lewis, S. M., Maicas, E., and Ouellette, R. J. (2009). Multiple isoforms of *PAX5* are expressed in both lymphomas and normal B-cells. *Br. J. Haematol.* 147, 328–338. doi: 10.1111/j.1365-2141.2009.07859.x
- Azuma, N., Tadokoro, K., Asaka, A., Yamada, M., Yamaguchi, Y., Handa, H., et al. (2005). The *Pax6* isoform bearing an alternative spliced exon promotes the development of the neural retinal structure. *Hum. Mol. Genet.* 14, 735–745. doi: 10.1093/hmg/ddi069
- Bassham, S., Canestro, C., and Postlethwait, J. (2008). Evolution of developmental roles of *Pax2/5/8* paralogs after independent duplication in urochordate and vertebrate lineages. *BMC Biol.* 6:35. doi: 10.1186/1741-7007-6-35
- Benson, D. A., Cavanaugh, M., Clark, K., Karsch-Mizrachi, I., Lipman, D. J., Ostell, J., et al. (2013). GenBank. *Nucleic Acids Res.* 41, D36–D42. doi: 10.1093/nar/gks1195
- Bergthorsson, U., Andersson, D. I., and Roth, J. R. (2007). Ohno's dilemma: evolution of new genes under continuous selection. *Proc. Natl. Acad. Sci. U.S.A.* 104, 17004–17009. doi: 10.1073/pnas.0707158104
- Bhatia, S., Monahan, J., Ravi, V., Gautier, P., Murdoch, E., Brenner, S., et al. (2014). A survey of ancient conserved non-coding elements in the *PAX6* locus reveals a landscape of interdigitated cis-regulatory archipelagos. *Dev. Biol.* 387, 214–228. doi: 10.1016/j.ydbio.2014.01.007
- Blake, J. A., and Ziman, M. R. (2014). *Pax* genes: regulators of lineage specification and progenitor cell maintenance. *Development* 141, 737–751. doi: 10.1242/dev.091785
- Bopp, D., Jamet, E., Baumgartner, S., Burri, M., and Noll, M. (1989). Isolation of two tissue-specific *Drosophila* paired box genes, *Pox meso* and *Pox neuro*. *EMBO J.* 8, 3447–3457.
- Borson, N. D., Lacy, M. Q., and Wettstein, P. J. (2002). Altered mRNA expression of *Pax5* and *Blimp-1* in B cells in multiple myeloma. *Blood* 100, 4629–4639. doi: 10.1182/blood.V100.13.4629
- Breitling, R., and Gerber, J. K. (2000). Origin of the paired domain. *Dev. Genes Evol.* 210, 644–650. doi: 10.1007/s004270000106
- Burge, C., and Karlin, S. (1997). Prediction of complete gene structures in human genomic DNA. *J. Mol. Biol.* 268, 78–94. doi: 10.1006/jmbi.1997.0951
- Busse, A., Rietz, A., Schwartz, S., Thiel, E., and Keilholz, U. (2009). An intron 9 containing splice variant of *PAX2*. *J. Transl. Med.* 7:36. doi: 10.1186/1479-736
- Cameron, R. A., Samanta, M., Yuan, A., He, D., and Davidson, E. (2009). SpBase: the sea urchin genome database and web site. *Nucleic Acids Res.* 37, D750–D754. doi: 10.1093/nar/gkn887
- Canestro, C., Bassham, S., and Postlethwait, J. (2005). Development of the central nervous system in the larvacean *Oikopleura dioica* and the evolution of the chordate brain. *Dev. Biol.* 285, 298–315. doi: 10.1016/j.ydbio.2005.06.039
- Chen, L., Tovar-Corona, J. M., and Urrutia, A. O. (2012). Alternative splicing: a potential source of functional innovation in the eukaryotic genome. *Int. J. Evol. Biol.* 2012:596274. doi: 10.1155/2012/596274
- Chi, N., and Epstein, J. A. (2002). Getting your pax straight: pax proteins in development and disease. *Trends Genet.* 18, 41–47. doi: 10.1016/S0168-9525(01)02594-X
- Chisholm, A. D., and Horvitz, H. R. (1995). Patterning of the *Caenorhabditis elegans* head region by the *Pax-6* family member *vab-3*. *Nature* 377, 52–55. doi: 10.1038/377052a0
- Cinar, H. N., and Chisholm, A. D. (2004). Genetic Analysis of the *Caenorhabditis elegans pax-6* Locus: roles of paired domain-containing and nonpaired domain-containing isoforms. *Genetics* 168, 1307–1322. doi: 10.1534/genetics.104.031724
- Cunningham, F., Amode, M. R., Barrell, D., Beal, K., Billis, K., Brent, S., et al. (2015). Ensembl 2015. *Nucleic Acids Res.* 43, D662–D669. doi: 10.1093/nar/gku1010
- Czerny, T., Schaffner, G., and Busslinger, M. (1993). DNA sequence recognition by pax proteins: bipartite structure of the paired domain and its binding site. *Genes Dev.* 7, 2048–2061. doi: 10.1101/gad.7.10.2048
- De Castro, E., Sigrist, C. J. A., Gattiker, A., Bulliard, V., Langendijk-Genevaux, P. S., Gasteiger, E., et al. (2006). ScanProsite: detection of PROSITE signature matches and ProRule-associated functional and structural residues in proteins. *Nucleic Acids Res.* 34, W362–W365. doi: 10.1093/nar/gkl124
- Dressler, G. R., Deutsch, U., Chowdhury, K., Nornes, H. O., and Gruss, P. (1990). *Pax2*, a new murine paired-box-containing gene and its expression in the developing excretory system. *Development* 109, 787–795.
- Eberhard, D., Jimenez, G., Heavey, B., and Busslinger, M. (2000). Transcriptional repression by *Pax5* (BSAP) through interaction with corepressors of the Groucho family. *EMBO J.* 19, 2292–2303. doi: 10.1093/emboj/19.10.2292
- Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* 32, 1792–1797. doi: 10.1093/nar/gkh340
- Epstein, D. J., Vekemans, M., and Gros, P. (1991). *Splotch (sp^{2H})*, a mutation affecting development of the mouse neural tube, shows a deletion within the paired homeodomain of *Pax-3*. *Cell* 67, 767–774. doi: 10.1016/0092-8674(91)90071-6
- Epstein, J. A., Glaser, T., Cai, J., Jepeal, L., Walton, D. S., and Maas, R. L. (1994). Two independent and interactive DNA-binding subdomains of the *Pax6* paired

- domain are regulated by alternative splicing. *Genes Dev.* 8, 2022–2034. doi: 10.1101/gad.8.17.2022
- Escrivà, H., Manzon, L., Youson, J., and Laudet, V. (2002). Analysis of lamprey and hagfish genes reveals a complex history of gene duplications during early vertebrate evolution. *Mol. Biol. Evol.* 19, 1440–1450. doi: 10.1093/oxfordjournals.molbev.a004207
- Fletcher, J., Hu, M., Berman, Y., Collins, F., Grigg, J., Mciver, M., et al. (2005). Multicystic Dysplastic kidney and variable phenotype in a family with a novel deletion mutation of *PAX2*. *J. Am. Soc. Nephrol.* 16, 2754–2761. doi: 10.1681/ASN.2005030239
- Friedrich, M., and Caravas, J. (2011). New insights from hemichordate genomes: prebilaterian origin and parallel modifications in the paired domain of the pax gene eyegone. *J. Exp. Zool. B Mol. Dev. Evol.* 316, 387–392. doi: 10.1002/jez.b.21412
- Fu, W., and Noll, M. (1997). The Pax2 homolog sparkling is required for development of cone and pigment cells in the *Drosophila* eye. *Genes Dev.* 11, 2066–2078. doi: 10.1101/gad.11.16.2066
- Fuentes, M., Benito, E., Bertrand, S., Paris, M., Mignardot, A., Godoy, L., et al. (2007). Insights into spawning behavior and development of the European amphioxus (*Branchiostoma lanceolatum*). *J. Exp. Zool. B Mol. Dev. Evol.* 308, 484–493. doi: 10.1002/jez.b.21179
- Gehring, W. J. (1996). The master control gene for morphogenesis and evolution of the eye. *Genes Cells* 1, 11–15. doi: 10.1046/j.1365-2443.1996.11011.x
- Gehring, W. J. (2002). The genetic control of eye development and its implications for the evolution of the various eye-types. *Int. J. Dev. Biol.* 46, 65–73.
- Gehring, W. J. (2012). The evolution of vision. *Wiley Interdiscip. Rev. Dev. Biol.* 3, 1–40. doi: 10.1002/wdev.96
- Glardon, S., Holland, L. Z., Gehring, W. J., and Holland, N. D. (1998). Isolation and developmental expression of the amphioxus *Pax-6* gene (*AmphiPax-6*): insights into eye and photoreceptor evolution. *Development* 125, 2701–2710.
- Glaser, T., Walton, D. S., and Maas, R. L. (1992). Genomic structure, evolutionary conservation and aniridia mutations in the human *PAX6* gene. *Nat. Genet.* 2, 232–239. doi: 10.1038/ng1192-232
- Green, R. E., Lewis, B. P., Hillman, R. T., Blanchette, M., Lareau, L. F., Garnett, A. T., et al. (2003). Widespread predicted nonsense-mediated mRNA decay of alternatively-spliced transcripts of human normal and disease genes. *Bioinformatics* 19(Suppl. 1), i118–i121. doi: 10.1093/bioinformatics/btg1015
- Grigoriev, I. V., Nordberg, H., Shabalov, I., Aerts, A., Cantor, M., Goodstein, D., et al. (2012). The genome portal of the department of energy joint Genome Institute. *Nucleic Acids Res.* 40, D26–D32. doi: 10.1093/nar/gkr947
- Hartung, F., Blattner, F. R., and Puchta, H. (2002). Intron gain and loss in the evolution of the conserved eukaryotic recombination machinery. *Nucleic Acids Res.* 30, 5175–5181. doi: 10.1093/nar/gkf649
- Heller, N., and Brandli, A. W. (1997). *Xenopus Pax-2* displays multiple splice forms during embryogenesis and pronephric kidney development. *Mech. Dev.* 69, 83–104. doi: 10.1016/S0925-4773(97)00158-5
- Heller, N., and Brandli, A. W. (1999). *Xenopus Pax-2/5/8* orthologues: novel insights into *Pax* gene evolution and identification of *Pax-8* as the earliest marker for otic and pronephric cell lineages. *Dev. Genet.* 24, 208–219. doi: 10.1002/(sici)1520-6408(1999)24:3/4<208::aid-dvg4>3.0.co;2-j
- Hill, A., Boll, W., Ries, C., Warner, L., Osswalt, M., Hill, M., et al. (2010). Origin of *Pax* and *Six* gene families in sponges: single *PaxB* and *Six1/2* orthologs in *Chalinula loosanoffi*. *Dev. Biol.* 343, 106–123. doi: 10.1016/j.ydbio.2010.03.010
- Holland, L. Z., and Short, S. (2010). Alternative splicing in development and function of chordate endocrine systems: a focus on *Pax* genes. *Integr. Comp. Biol.* 50, 22–34. doi: 10.1093/icb/icq048
- Holland, L. Z., and Yu, J. K. (2004). Cephalochordate (amphioxus) embryos: procurement, culture, and basic methods. *Methods Cell. Biol.* 74, 195–215. doi: 10.1016/S0091-679X(04)74009-1
- Hoshiyama, D., Iwabe, N., and Miyata, T. (2007). Evolution of the gene families forming the *Pax/Six* regulatory network: isolation of genes from primitive animals and molecular phylogenetic analyses. *FEBS Lett.* 581, 1639–1643. doi: 10.1016/j.febslet.2007.03.027
- Howard-Ashby, M., Materna, S. C., Brown, C. T., Chen, L., Cameron, R. A., and Davidson, E. H. (2006). Identification and characterization of homeobox transcription factor genes in *Strongylocentrotus purpuratus*, and their expression in embryonic development. *Dev. Biol.* 300, 74–89. doi: 10.1016/j.ydbio.2006.08.039
- Irvine, S. Q., Fonseca, V. C., Zompa, M. A., and Antony, R. (2008). *Cis*-regulatory organization of the *Pax6* gene in the ascidian *Ciona intestinalis*. *Dev. Biol.* 317, 649–659. doi: 10.1016/j.ydbio.2008.01.036
- Iwamatsu, T. (2004). Stages of normal development in the medaka *Oryzias latipes*. *Mech. Dev.* 121, 605–618. doi: 10.1016/j.mod.2004.03.012
- Jaillon, O., Aury, J. M., Brunet, F., Petit, J. L., Stange-Thomann, N., Mauceli, E., et al. (2004). Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431, 946–957. doi: 10.1038/nature03025
- Jeffares, D. C., Mourier, T., and Penny, D. (2006). The biology of intron gain and loss. *Trends Genet.* 22, 16–22. doi: 10.1016/j.tig.2005.10.006
- Karolchik, D., Barber, G. P., Casper, J., Clawson, H., Cline, M. S., Diekhans, M., et al. (2014). The UCSC Genome Browser database: 2014 update. *Nucleic Acids Res.* 42, D764–D770. doi: 10.1093/nar/gkt1168
- Kelemen, O., Convertini, P., Zhang, Z., Wen, Y., Shen, M., Falaleeva, M., et al. (2013). Function of alternative splicing. *Gene* 514, 1–30. doi: 10.1016/j.gene.2012.07.083
- Kimmel, C. B., Ballard, W. W., Kimmel, S. R., Ullmann, B., and Schilling, T. F. (1995). Stages of embryonic development of the zebrafish. *Dev. Dyn.* 203, 253–310. doi: 10.1002/aja.1002030302
- Kleinjan, D. A., Bancewicz, R. M., Gautier, P., Dahm, R., Schonthaler, H. B., Damante, G., et al. (2008). Subfunctionalization of duplicated zebrafish *pax6* genes by *cis*-regulatory divergence. *PLoS Genet.* 4:e29. doi: 10.1371/journal.pgen.0040029
- Klimova, L., and Kozmík, Z. (2014). Stage-dependent requirement of neuroretinal *Pax6* for lens and retina development. *Development* 141, 1292–1302. doi: 10.1242/dev.098822
- Koralewski, T. E., and Krutovsky, K. V. (2011). Evolution of exon-intron structure and alternative splicing. *PLoS ONE* 6:e18055. doi: 10.1371/journal.pone.0018055
- Kozmík, Z., Czerny, T., and Busslinger, M. (1997). Alternatively spliced insertions in the paired domain restrict the DNA sequence specificity of *Pax6* and *Pax8*. *EMBO J.* 16, 6793–6803. doi: 10.1093/emboj/16.22.6793
- Kozmík, Z., Daube, M., Frei, E., Norman, B., Kos, L., Dishaw, L. J., et al. (2003). Role of *Pax* genes in eye evolution: a cnidarian *PaxB* gene uniting *Pax2* and *Pax6* functions. *Dev. Cell.* 5, 773–785. doi: 10.1016/S1534-5807(03)00325-3
- Kozmík, Z., Holland, N. D., Kalousová, A., Paces, J., Schubert, M., and Holland, L. Z. (1999). Characterization of an amphioxus paired box gene, *AmphiPax2/5/8*: developmental expression patterns in optic support cells, nephridium, thyroid-like structures and pharyngeal gill slits, but not in the midbrain-hindbrain boundary region. *Development* 126, 1295–1304.
- Kozmík, Z., Kurzbauer, R., Dorfler, P., and Busslinger, M. (1993). Alternative splicing of *Pax-8* gene transcripts is developmentally regulated and generates isoforms with different transactivation properties. *Mol. Cell. Biol.* 13, 6024–6035. doi: 10.1128/MCB.13.10.6024
- Kozmík, Z. (2008). The role of *Pax* genes in eye evolution. *Brain Res. Bull.* 75, 335–339. doi: 10.1016/j.brainresbull.2007.10.046
- Krauss, S., Johansen, T., Korzh, V., and Fjose, A. (1991). Expression of the zebrafish paired box gene *pax[zf-b]* during early neurogenesis. *Development* 113, 1193–1206.
- Kreslova, J., Holland, L. Z., Schubert, M., Burgdorf, C., Benes, V., and Kozmík, Z. (2002). Functional equivalence of amphioxus and vertebrate *Pax2/5/8* transcription factors suggests that the activation of mid-hindbrain specific genes in vertebrates occurs via the recruitment of *Pax* regulatory elements. *Gene* 282, 143–150. doi: 10.1016/S0378-1119(01)00840-X
- Kwak, S. J., Vemaraju, S., Moorman, S. J., Zeddis, D., Popper, A. N., and Riley, B. B. (2006). Zebrafish *pax5* regulates development of the utricular macula and vestibular function. *Dev. Dyn.* 235, 3026–3038. doi: 10.1002/dvdy.20961
- Lang, D., Powell, S. K., Plummer, R. S., Young, K. P., and Ruggeri, B. A. (2007). *PAX* genes: roles in development, pathophysiology, and cancer. *Biochem. Pharmacol.* 73, 1–14. doi: 10.1016/j.bcp.2006.06.024
- Lewis, B. P., Green, R. E., and Brenner, S. E. (2003). Evidence for the widespread coupling of alternative splicing and nonsense-mediated mRNA decay in humans. *Proc. Natl. Acad. Sci. U.S.A.* 100, 189–192. doi: 10.1073/pnas.0136770100

- Louis, A., Muffato, M., and Roest Crolius, H. (2013). Genomicus: five genome browsers for comparative genomics in eukaryota. *Nucleic Acids Res.* 41, D700–D705. doi: 10.1093/nar/gks1156
- Lun, K., and Brand, M. (1998). A series of *no isthmus (noi)* alleles of the zebrafish *pax2.1* gene reveals multiple signaling events in development of the midbrain-hindbrain boundary. *Development* 125, 3049–3062.
- Maere, S., and Van De Peer, Y. (2010). “Duplicate retention after small- and large-scale duplications,” in *Evolution after Gene Duplication*, eds K. Dittmar and D. Liberles (Hoboken, NJ: John Wiley & Sons, Inc.), 31–56.
- Nilsen, T. W., and Graveley, B. R. (2010). Expansion of the eukaryotic proteome by alternative splicing. *Nature* 463, 457–463. doi: 10.1038/nature08909
- Noll, M. (1993). Evolution and role of *Pax* genes. *Curr. Opin. Genet. Dev.* 3, 595–605. doi: 10.1016/0959-437X(93)90095-7
- Nornes, S., Mikkola, I., Krauss, S., Delghandi, M., Perander, M., and Johansen, T. (1996). Zebrafish *Pax9* encodes two proteins with distinct C-terminal transactivating domains of different potency negatively regulated by adjacent N-terminal sequences. *J. Biol. Chem.* 271, 26914–26923. doi: 10.1074/jbc.271.43.26914
- Paixao-Cortes, V. R., Salzano, F. M., and Bortolini, M. C. (2013). Evolutionary history of chordate PAX genes: dynamics of change in a complex gene family. *PLoS ONE* 8:e73560. doi: 10.1371/journal.pone.0073560
- Pfeffer, P. L., Gerster, T., Lun, K., Brand, M., and Busslinger, M. (1998). Characterization of three novel members of the zebrafish *Pax2/5/8* family: dependency of *Pax5* and *Pax8* expression on the *Pax2.1 (noi)* function. *Development* 125, 3063–3074.
- Poleev, A., Wendler, F., Fickenscher, H., Zannini, M. S., Yaginuma, K., Abbott, C., et al. (1995). Distinct functional properties of three human paired-box-protein, PAX8, isoforms generated by alternative splicing in thyroid, kidney and Wilms' tumors. *Eur. J. Biochem.* 228, 899–911. doi: 10.1111/j.1432-1033.1995.0899m.x
- Puschel, A. W., Gruss, P., and Westerfield, M. (1992). Sequence and expression pattern of *pax-6* are highly conserved between zebrafish and mice. *Development* 114, 643–651.
- Putnam, N. H., Butts, T., Ferrier, D. E., Furlong, R. F., Hellsten, U., Kawashima, T., et al. (2008). The amphioxus genome and the evolution of the chordate karyotype. *Nature* 453, 1064–1071. doi: 10.1038/nature06967
- Ravi, V., Bhatia, S., Gautier, P., Loosli, F., Tay, B. H., Tay, A., et al. (2013). Sequencing of *Pax6* loci from the elephant shark reveals a family of *Pax6* genes in vertebrate genomes, forged by ancient duplications and divergences. *PLoS Genet.* 9:e1003177. doi: 10.1371/journal.pgen.1003177
- Reese, M. G., Eeckman, F. H., Kulp, D., and Haussler, D. (1997). Improved splice site detection in Genie. *J. Comput. Biol.* 4, 311–323. doi: 10.1089/cmb.1997.4.311
- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOSS: the European molecular biology open software suite. *Trends Genet.* 16, 276–277. doi: 10.1016/S0168-9525(00)02024-2
- Robichaud, G. A., Nardini, M., Laflamme, M., Cuperlovic-Culf, M., and Ouellette, R. J. (2004). Human Pax-5 C-terminal isoforms possess distinct transactivation properties and are differentially modulated in normal and malignant B cells. *J. Biol. Chem.* 279, 49956–49963. doi: 10.1074/jbc.M407171200
- Rogozin, I. B., Wolf, Y. I., Sorokin, A. V., Mirkin, B. G., and Koonin, E. V. (2003). Remarkable interkingdom conservation of intron positions and massive, lineage-specific intron loss and gain in eukaryotic evolution. *Curr. Biol.* 13, 1512–1517. doi: 10.1016/S0960-9822(03)00558-X
- Schimmenti, L. A., Cunliffe, H. E., Mcnue, L. A., Ward, T. A., French, M. C., Shim, H. H., et al. (1997). Further delineation of renal-coloboma syndrome in patients with extreme variability of phenotype and identical *PAX2* mutations. *Am. J. Hum. Genet.* 60, 869–878.
- Schwartz, S., Zhang, Z., Frazer, K. A., Smit, A., Riemer, C., Bouck, J., et al. (2000). PipMaker—a web server for aligning two genomic DNA sequences. *Genome Res.* 10, 577–586. doi: 10.1101/gr.10.4.577
- Short, S., and Holland, L. Z. (2008). The evolution of alternative splicing in the *Pax* family: the view from the Basal chordate amphioxus. *J. Mol. Evol.* 66, 605–620. doi: 10.1007/s00239-008-9113-5
- Short, S., Kozmik, Z., and Holland, L. Z. (2012). The function and developmental expression of alternatively spliced isoforms of amphioxus and *Xenopus laevis* *Pax2/5/8* genes: revealing divergence at the invertebrate to vertebrate transition. *J. Exp. Zool. B Mol. Dev. Evol.* 318, 555–571. doi: 10.1002/jez.b.22460
- Stephens, R. M., and Schneider, T. D. (1992). Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites. *J. Mol. Biol.* 228, 1124–1136. doi: 10.1016/0022-2836(92)90320-J
- Stuart, E. T., Kioussi, C., and Gruss, P. (1994). Mammalian *Pax* genes. *Annu. Rev. Genet.* 28, 219–236. doi: 10.1146/annurev.ge.28.120194.001251
- Takatori, N., Butts, T., Candiani, S., Pestarino, M., Ferrier, D. K., Saiga, H., et al. (2008). Comprehensive survey and classification of homeobox genes in the genome of amphioxus, *Branchiostoma floridae*. *Dev. Genes Evol* 218, 579–590. doi: 10.1007/s00427-008-0245-9
- Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., and Kumar, S. (2011). MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* 28, 2731–2739. doi: 10.1093/molbev/msr121
- Tavassoli, K., Ruger, W., and Horst, J. (1997). Alternative splicing in PAX2 generates a new reading frame and an extended conserved coding region at the carboxy terminus. *Hum. Genet.* 101, 371–375. doi: 10.1007/s004390050644
- Van De Peer, Y. (2004). *Tetraodon* genome confirms *Takifugu* findings: most fish are ancient polyploids. *Genome Biol.* 5:250. doi: 10.1186/gb-2004-5-12-250
- Wada, S., Tokuoka, M., Shoguchi, E., Kobayashi, K., Di Gregorio, A., Spagnuolo, A., et al. (2003). A genomewide survey of developmentally relevant genes in *Ciona intestinalis*. II. Genes for homeobox transcription factors. *Dev. Genes Evol.* 213, 222–234. doi: 10.1007/s00427-003-0321-0
- Walther, C., and Gruss, P. (1991). *Pax-6*, a murine paired box gene, is expressed in the developing CNS. *Development* 113, 1435–1449.
- Wang, E. T., Sandberg, R., Luo, S., Khrebtukova, I., Zhang, L., Mayr, C., et al. (2008a). Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476. doi: 10.1038/nature07509
- Wang, Q., Fang, W. H., Krupinski, J., Kumar, S., Slevin, M., and Kumar, P. (2008b). *Pax* genes in embryogenesis and oncogenesis. *J. Cell. Mol. Med.* 12, 2281–2294. doi: 10.1111/j.1582-4934.2008.00427.x
- Ward, T. A., Nebel, A., Reeve, A. E., and Eccles, M. R. (1994). Alternative messenger RNA forms and open reading frames within an additional conserved region of the human *PAX-2* gene. *Cell Growth Differ.* 5, 1015–1021.
- Xu, H. E., Rould, M. A., Xu, W., Epstein, J. A., Maas, R. L., and Pabo, C. O. (1999). Crystal structure of the human Pax6 paired domain-DNA complex reveals specific roles for the linker region and carboxy-terminal subdomain in DNA binding. *Genes Dev.* 13, 1263–1275. doi: 10.1101/gad.13.10.1263
- Xu, W., Rould, M. A., Jun, S., Desplan, C., and Pabo, C. O. (1995). Crystal structure of a paired domain-DNA complex at 2.5 Å resolution reveals structural basis for Pax developmental mutations. *Cell* 80, 639–650. doi: 10.1016/0092-8674(95)90518-9
- Zwollo, P., Arrieta, H., Ede, K., Molinder, K., Desiderio, S., and Pollock, R. (1997). The *Pax-5* gene is alternatively spliced during B-cell development. *J. Biol. Chem.* 272, 10160–10168. doi: 10.1074/jbc.272.15.10160

Conflict of Interest Statement: 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 © 2015 Fabian, Kozmikova, Kozmik and Pantzartzi. 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) or licensor 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.



OPEN ACCESS

Edited by:

Naiara Rodriguez-Ezpeleta,
AZTI-Tecnalia, Spain

Reviewed by:

Ingo Braasch,
University of Oregon, USA
Brian Frank Eames,
University of Saskatchewan, Canada

***Correspondence:**

Sylvain Marcellini,
Department of Cell Biology, Faculty of
Biological Sciences, Universidad de
Concepción, Barrio Universitario s/n,
Casilla 160-C, Concepción, Chile
smarcellini@udec.cl;
Mélanie Debiais-Thibaud,
Institut des Sciences de l'Evolution de
Montpellier, UMR5554, Université
Montpellier, Centre National de la
Recherche Scientifique, IRD, EPHE,
Eugène Bataillon, cc064 - 34000
Montpellier, France
mdebiais@univ-montp2.fr

†Present Address:

Willian T. A. F. Silva,
Department of Ecology and
Genetics/Evolutionary Biology,
Evolutionary Biology Center, Uppsala
University, Uppsala, Sweden

[‡]Co-first authors.

Specialty section:

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

Received: 30 April 2015

Accepted: 24 August 2015

Published: 15 September 2015

Citation:

Enault S, Muñoz DN, Silva WTA, Borday-Birraux V, Bonade M, Oulion S, Ventéo S, Marcellini S and 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. *Front. Genet.* 6:283.
doi: 10.3389/fgene.2015.00283

Molecular footprinting of skeletal tissues in the catshark *Scyliorhinus canicula* and the clawed frog *Xenopus tropicalis* identifies conserved and derived features of vertebrate calcification

Sébastien Enault^{1†}, David N. Muñoz^{2‡}, Willian T. A. F. Silva^{1†}, Véronique Borday-Birraux^{3,4}, Morgane Bonade³, Silvan Oulion¹, Stéphanie Ventéo⁵, Sylvain Marcellini^{2*} and Mélanie Debiais-Thibaud^{1*}

¹ Institut des Sciences de l'Evolution de Montpellier, UMR5554, Université Montpellier, Centre National de la Recherche Scientifique, IRD, EPHE, Montpellier, France, ²Laboratory of Development and Evolution, Department of Cell Biology, Faculty of Biological Sciences, Universidad de Concepción, Concepción, Chile, ³Laboratoire EGCE UMR Centre National de la Recherche Scientifique 9191, IRD247, Université Paris Sud, Gif-sur-Yvette, France, ⁴Université Paris Diderot, Sorbonne Paris Cité, Paris, France, ⁵Institute for Neurosciences of Montpellier, Institut National de la Santé et de la Recherche Médicale U1051, Montpellier, France

Understanding the evolutionary emergence and subsequent diversification of the vertebrate skeleton requires a comprehensive view of the diverse skeletal cell types found in distinct developmental contexts, tissues, and species. To date, our knowledge of the molecular nature of the shark calcified extracellular matrix, and its relationships with osteichthyan skeletal tissues, remain scarce. Here, based on specific combinations of expression patterns of the *Col1a1*, *Col1a2*, and *Col2a1* fibrillar collagen genes, we compare the molecular footprint of endoskeletal elements from the chondrichthyan *Scyliorhinus canicula* and the tetrapod *Xenopus tropicalis*. We find that, depending on the anatomical location, *Scyliorhinus* skeletal calcification is associated to cell types expressing different subsets of fibrillar collagen genes, such as high levels of *Col1a1* and *Col1a2* in the neural arches, high levels of *Col2a1* in the tesserae, or associated to a drastic *Col2a1* downregulation in the centrum. We detect low *Col2a1* levels in *Xenopus* osteoblasts, thereby revealing that the osteoblastic expression of this gene was significantly reduced in the tetrapod lineage. Finally, we uncover a striking parallel, from a molecular and histological perspective, between the vertebral cartilage calcification of both species and discuss the evolutionary origin of endochondral ossification.

Keywords: fibrillar collagens, vertebrate skeletogenesis, bone, cartilage, *Scyliorhinus canicula*, *Xenopus tropicalis*

Introduction

The evolutionary origin and diversification of the skeleton remains one of the most intriguing issue in vertebrate biology. Solving this problem requires a comprehensive view of the diversity of skeletal cell types found in distinct developmental contexts, tissues, anatomical locations, and

species, as has been emphasized in a recent synthesis of existing skeletal terminologies (Dahdul et al., 2012). In mammals, the chondrocytes produce the extracellular matrix of the fibrous, elastic and hyaline cartilage, while osteoblasts and osteocytes are involved in bone formation (Benjamin and Evans, 1990; Hartmann, 2009; Zhang et al., 2009; Long, 2011). Yet, an intermediate type of chondroid bone, exhibiting characteristics of both bone and cartilage, has been described in rodents as well as teleosts, leading some authors to propose that, in fact, bone and cartilage represent two extreme forms of a skeletal tissue continuum (Huysseune and Verraes, 1986; Huysseune and Sire, 1990; Mizoguchi et al., 1997; Kranenborg et al., 2005; Estêvão et al., 2011). In addition, chondrichthyans display a series of heavily calcified skeletal tissues such as the cartilaginous tesserae of the jaws (with no obvious homologs in osteichthyans, see Dean et al., 2005; Dean and Summers, 2006; Dean et al., 2009; Omelon et al., 2014), the vertebral body developing around the notochord (Peignoux-Deville et al., 1982; Dean and Summers, 2006; Eames et al., 2007; Porter et al., 2007; Fleming et al., 2015) and the perichondrium of the neural arches laying on each side of the neural tube (Peignoux-Deville et al., 1982; Eames et al., 2007). In summary, while developmental and paleontological studies have revealed the versatile nature of skeletal tissues characterizing the vertebrate skeleton (Donoghue and Sansom, 2002; Janvier and Arsenault, 2002; Dahdul et al., 2012; Janvier, 2015), the molecular identity and the evolutionary relationships of the distinct vertebrate skeletal cell types remain an open question.

The comparison of expression patterns represents a powerful approach to examine cell type evolution and, for instance, has shed light on the origin of sensory neurons in animals (Arendt, 2008). Here, we have explored the possibility that combinations of expression patterns of genes coding for crucial components of the skeletal matrix can serve as useful molecular footprints to compare the identity of skeletal cell types between chondrichthyan and osteichthyan representatives. We chose to focus on the *Col1a1*, *Col1a2*, and *Col2a1* genes, belonging to the Clade A of the fibrillar collagen family, because they are known to contribute to biomineralization and because they are intimately associated to skeletal development and evolution (Wada et al., 2006; Rychel and Swalla, 2007; Zhang and Cohn, 2008; Landis and Silver, 2009; Eyre and Weis, 2013; Veis and Dorvee, 2013). *Col1a1* and *Col1a2* (Type I collagen) are robustly expressed in osteichthyan osteoblasts (Kobayashi and Kronenberg, 2005; Li et al., 2009; Albertson et al., 2010; Estêvão et al., 2011; Eames et al., 2012). By contrast, the osteoblastic expression of *Col2a1* (Type II collagen) is more variable and has been reported in developing bones of gar and teleosts (Benjamin and Ralphs, 1991; Albertson et al., 2010; Eames et al., 2012), at low levels in some scattered mouse osteoblasts (Hilton et al., 2007), and transiently in chick preosteoblasts (Abzhanov et al., 2007). In addition, *Col2a1* displays a conserved expression pattern in chondrocytes of immature hyaline cartilage whose proliferation drives the growth of endochondral bones (Benjamin and Ralphs, 1991; Nah et al., 2001; Kerney and Hanken, 2008; Hartmann, 2009; Albertson et al., 2010; Estêvão et al., 2011; Eames et al., 2012). *Col2a1* expression becomes progressively downregulated as the hyaline cartilage matures and calcifies its extracellular matrix

(Eames et al., 2003; Hartmann, 2009). Of particular relevance for this study, *Col2a1*-negative mature cartilage calcification usually occurs at levels that are too weak to robustly stain with Alizarin red, a reagent commonly used to specifically detect the calcifying bone matrix of vertebrate embryos (Kirsch et al., 1997; Khanarian et al., 2014), with some exceptions reported in the swell shark vertebrae and the domestic fowl trachea (Hogg, 1982; Eames et al., 2007). Possibly due to lineage-specific duplications, lamprey and hagfish (cyclostomes) exhibit one or two *Col2a1* orthologs (and no *Col1a1* or *Col1a2* genes) expressed in broad regions encompassing mesenchymal cells and some, but not all, cartilaginous elements (Zhang and Cohn, 2006, 2008; Zhang et al., 2006; Ota and Kuratani, 2010; Cattell et al., 2011). In shark, immunohistochemistry allowed the clear detection of Type II collagen fibers in cartilage matrix, while the weaker reactivity of the anti-Type I collagen antibody suggested a perichondral expression, without allowing the discrimination of cells secreting *Col1a1* and/or *Col1a2* proteins (Eames et al., 2007).

In order to identify skeletal cell types sharing a specific molecular identity between remotely related jawed vertebrates, we compared the endoskeletal expression patterns of the *Col1a1*, *Col1a2*, and *Col2a1* fibrillar collagen genes in the chondrichthyan *Scyliorhinus canicula* (*S.c.*) and the tetrapod *Xenopus tropicalis* (*X.t.*). We find that, depending on the anatomical location, skeletal calcification in *S.c.* occurs in the vicinity of cell types expressing distinct combinations of fibrillar collagen genes. In particular, calcification is associated to perichondral cells expressing high levels of *Col1a1* and *Col1a2* in the neural arches, and to chondrocytes expressing high levels of *Col2a1* in the tesserae or experiencing a drastic *Col2a1* downregulation in the centrum. In *X.t.*, the moderate expression of *Col2a1* in some osteoblasts differs from the situation described in actinopterygians and amniotes, suggesting that the osteoblastic expression of this gene was significantly reduced in the tetrapod lineage. Finally, we observe a striking parallel between the internal calcification of the vertebral cartilage of *X.t.* and *S.c.* and discuss the evolutionary origins of endochondral ossification.

Materials and Methods

Scyliorhinus canicula Biological Material

Scyliorhinus canicula embryos were obtained at the Station Méditerranéenne de l'Environnement Littoral (SMEL, Sète, France) and raised in the laboratory at 18°C. Embryos were euthanized by overdose of MS-222 (Sigma) following all animal-care specifications of the European ethics legislation. Whole embryos were fixed 48 h in PFA 4% in PBS 1× at 4°C and then transferred in ethanol at -20°C for storage. Dissected body parts (jaws or trunk sections) were rehydrated and transferred to a 25% sucrose bath for cryosection at 14 μm thickness, and stored at -20°C on alternative slides to get comparable sections on each slide. These sections were used for *in situ* hybridizations and Alizarin red-Alcian blue histological staining (see following sections). Dissected body parts were decalcified in MORSE (sodium citrate 10% and formic acid 20%) solution for 5 days before being transferred to paraplast blocs and sectioned at 10 μm thickness.

These sections were used for Hematoxylin-Eosin-Safran (HES) histological staining and anti-Col2 immunofluorescence. To perform immunofluorescence, dissected trunk slices from 6.7 cm-long embryos and dissected jaw from 9 cm-long embryo were demineralized for 3 h in MORSE solution at room temperature prior to dehydratation, embedded in paraplast and cut at 10–12 µm thickness.

Histological and Immunological Stainings

The same histological procedures were performed for the catshark and clawed frog samples. Histological Alizarin red/Alcian blue double staining was performed by rehydrating samples 1 min in phosphate-buffered saline (PBS) 1X, incubating 30 s in a 0.005% Alizarin red S solution (in KOH 0.5%), washing once with PBS 1X, incubating for 2 min in a 0.02% Alcian blue 8G in solution (in 8:2 ethanol/glacial acid acetic), and washing once in EtOH 100% and once in PBS 1X. The slides were then mounted in mowiol. Hematoxilin-Eosin-Safran (HES) histological staining was performed following standard protocols. Col2 immunofluorescence was performed using a 1/200 dilution of a primary anti-collagen II (II-II6B3; Developmental Studies Hybridoma Bank, Iowa City, IA, USA) and a 1/1500 dilution of a secondary Goat polyclonal anti-Mouse IgG—AlexaFluor 594 (Abcam ab150116). For epitope retrieval, sections were treated with trypsin 0.05% (Sigma) in 0.1% CaCl₂ buffer at pH7.8 buffer during 10 min at 37°C. Cell nuclei were counterstained with Hoechst.

Scyliorhinus canicula Collagen Clones

Plasmids containing partial or complete collagen cDNA sequences were obtained through screening of a cDNA library of embryo RNA extracts (Oulion et al., 2010). Specific clones were identified by BLAST as *Scyliorhinus canicula* (Sc-) *Collagen1a1* gene (Sc-*Col1a1*, NCBI accession numbers EU241868.1 and KT261785), *Collagen1a2* gene (Sc-*Col1a2*, NCBI accession numbers EU241869.1 and KT261784), and *Collagen2a1* gene (Sc-*Col2a1*, NCBI accession number EU241867.1). The sequences and details of the clones are provided in the Data Sheet 1. The phylogenetic relationships between proteic sequences were inferred by using the Maximum Likelihood method based on the Le_Gascuel_2008 model (Le and Gascuel, 2008). Initial tree(s) for the heuristic search were obtained by applying the Neighbor-Joining method to a matrix of pairwise distances estimated using a JTT model. A discrete Gamma distribution was used to model evolutionary rate differences among sites [4 categories (+G, parameter = 0.7935)]. The analysis involved 16 amino acid sequences. All positions containing gaps and missing data were eliminated. There were a total of 544 positions in the final dataset. Evolutionary analyses were conducted in MEGA6 (Tamura et al., 2013).

Scyliorhinus canicula and *Xenopus tropicalis* Probes

PCR products from specific amplification on Sc-*Col1a2* and Sc-*Col2a1* cDNA inserts were ligated into the pGEM-Teasy vector using the TA cloning kit (Promega). Sc-*Col1a1* was directly amplified from the original cDNA clone. *Xenopus tropicalis* (Xt-

Xt-*Col1a1* (NM_001011005.1), Xt-*Col1a2* (NM_001079250.1), and Xt-*Col2a1* (NM_203889) were amplified from stage NF60 hindlimb cDNA containing both bone and cartilage and blunt-cloned into the pBluescript vector. The PCR primers used in this study are given in Supplementary Table 1. Antisense DIG riboprobes were synthesized using the DIG RNA labeling mix (Roche) and the T3, T7 or Sp6 RNA polymerase (Promega) following the manufacturer's instructions. DIG-labeled riboprobes were purified on MicroSpin G50 column (GE Healthcare).

In situ Hybridization on *Scyliorhinus canicula* Sections

DIG-labeled probes were hybridized at 70°C overnight, sections were washed twice in 50% formamide, 1× SSC, 0.1% Tween-20 for 1 h at 70°C, twice in MABT buffer for 30 min before blocking in blocking buffer (MABT, 2% blocking reagent from Roche, 20% inactivated sheep serum) for 2 h at room temperature. Sections were then exposed overnight to a 1:2000 dilution of anti-DIG-AP conjugate antibody (Roche) at 4°C. After washing, slides were incubated with NBT-BCIP (Roche) staining solution according to the manufacturer's instructions and the reaction stopped by washing in water. Images of *in situ* hybridizations and histological stainings were taken under a Hamamatsu NanoZoomer 2.0-HT Slide Scanner (40× objective). Sense probe negative *in situ* hybridization results are shown in Data Sheet 2.

Xenopus tropicalis Animal Care and *In situ* Hybridization Procedure

Adult frogs are routinely maintained at the University of Concepcion following standard protocols established for *Xenopus tropicalis*. Embryos and tadpoles were obtained by natural mating and staged according to the Nieuwkoop and Faber developmental table (Nieuwkoop and Faber, 1967). Tadpoles were anesthetized with a solution of 200 mg/mL of MS-222 (Sigma) and subsequently decapitated, in agreement with international bioethical recommendations (Close et al., 1996; Ramlochansingh et al., 2014). The Ethics Committee of the University of Concepcion (Concepcion, Chile) approved all experimental procedures carried out 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). Sense probe negative *in situ* hybridization results are shown in Data Sheet 2. *In situ* hybridizations on paraffin sections were performed as previously described (see Data Sheet 3 and Espinoza et al., 2010; Aldea et al., 2013).

Results

Skeletal Expression of the Major Fibrillar Collagen Genes in *Scyliorhinus canicula* Fins and Jaws

The Sc-*Col1a1*, Sc-*Col1a2*, and Sc-*Col2a1* protein sequences were unambiguously associated to their respective orthology

groups by phylogenetic analyses (**Data Sheets 4, 5**). We examined calcification patterns by Alizarin red, Alcian blue, and HES stainings as well as the expression of *Sc-Col1a1*, *Sc-Col1a2*, and *Sc-Col2a1* in developing *S.c.* fins and jaws (**Figure 1**).

Alizarin red is specific for high levels of calcium ions and will therefore stain calcified extracellular matrix, while Alcian blue has a strong affinity for glycosaminoglycans of the cartilage matrix. The HES staining classically allows the location of nuclei (dark purple), cytoplasms (pink), and densely organized collagen

fibers (orange-pink). Both Safran and the acid aniline dye Eosin will stain the mineralized matrix more intensely than the non-mineralized matrix. Transverse sections through the pectoral fins showed that cartilaginous radials are devoid of calcification both in 7 cm long embryos (**Figures 1A–C'**) and 9 cm long embryos (not shown). By contrast, longitudinal sections of Meckel's cartilage from 9 cm long embryos allowed the detection of tesserae calcification at the cartilage periphery (**Figures 1H–J**). Tesserae calcification is associated to a darker HES staining of the hyaline matrix surrounding clusters of chondrocytes, and occurs

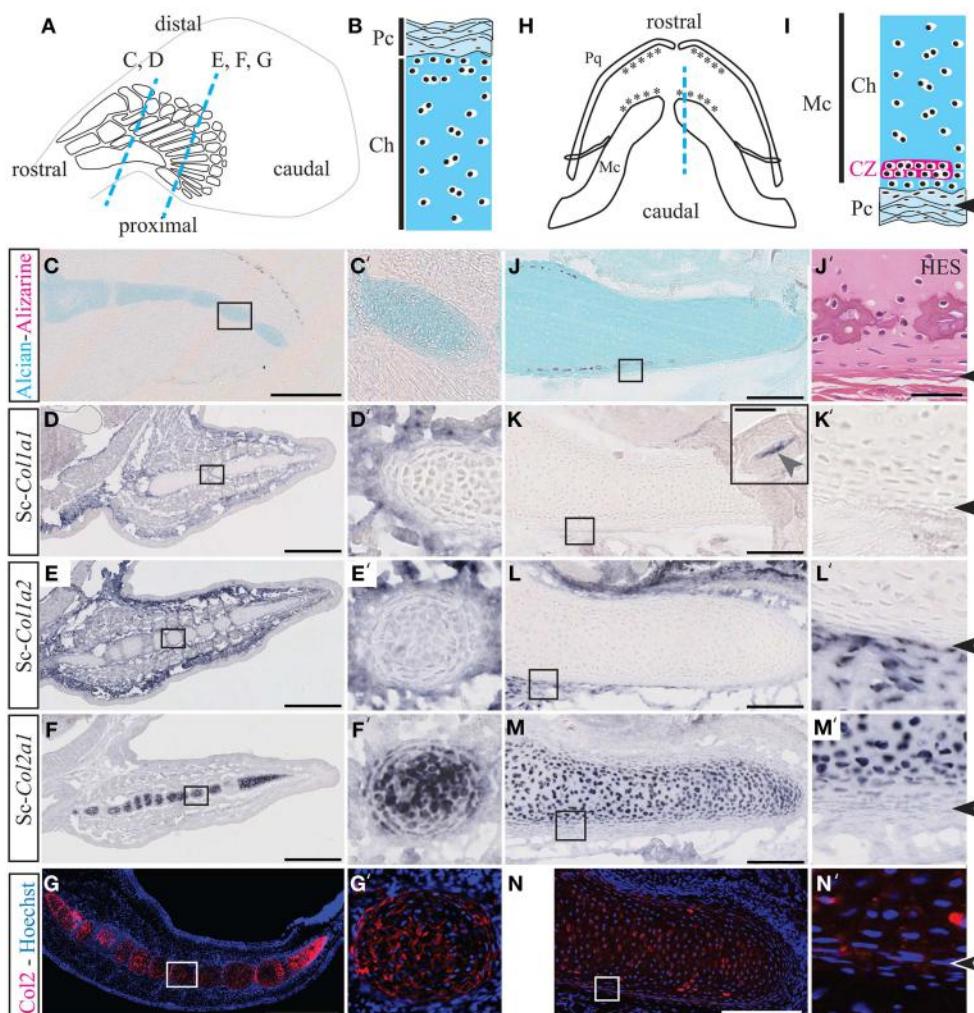


FIGURE 1 | Cartilage calcification and collagen expression in *Scyliorhinus canicula* radials and Meckel's cartilage. **(A)** Schematic drawing of the pectoral fin anatomy from 7 cm long *S.c.* embryos and of the orientation of the paraffin sections shown in C–G (blue dotted lines). Rostral and caudal refer to the embryonic axis. **(B)** General histology of pectoral skeletal elements, with the center of the cartilaginous element located at the bottom. **(C, C')** Alcian red and Alcian blue double staining. **(D–F)** Gene expression patterns in the pectoral fin for *Sc-Col1a1* **(D)**, *Sc-Col1a2* **(E)**, and *Sc-Col2a1* **(F)**. **(G)** Immunofluorescence using an anti-Type II collagen (Col2) antibody specifically marks the pectoral fin cartilaginous condensations. **(H)** Schematic drawing of the jaw anatomy from 9 cm-long *S.c.* embryos (ventral view) and of the orientation of the paraffin sections shown in **(J–N)** (blue dotted line). **(I)** General histology of Meckel's cartilage, with the center of the cartilaginous element located at the top. The arrowheads in **(I, J'–N')** demarcate the fibrous perichondrium from the cartilage. **(J)** Alcian red and Alcian blue double staining. **(J')** Higher magnification of a tesserae located in a similar region as the area boxed in **(J)** and stained with HES. **(K–M)** Gene expression patterns in the jaw for *Sc-Col1a1* [the inset in **(K)** shows a *Sc-Col1a1* positive dermal denticle from the same section], *Sc-Col1a2* **(L)** and *Sc-Col2a1* **(M)**. **(N)** Immunofluorescence using an anti-Type II collagen (Col2) antibody specifically marks the cartilaginous condensations of Meckel's cartilage. Insets in **(C–N)** are shown at higher magnification in **(C'–N')**, respectively. CZ, calcification zone of the tesserae; Ch, chondrocytes; Fb, fibroblasts; Pc, perichondrium; Pq, palatoquadrate. Scale bars: **(C–G)**, 250 μ m; **(J–N)**, 100 μ m.

within the cartilaginous scaffold, one or two cell diameters away from the fibrous perichondrium (**Figure 1J'**).

In the pectoral fin, Sc-Col1a1 and Sc-Col1a2 are expressed in the fibrous perichondrium and the connective tissue surrounding the cartilaginous elements (**Figures 1D–E'**), and Sc-Col2a1 is expressed in the chondrocytes of the cartilaginous matrix of the radials (**Figures 1F,F'**). In the jaw, we failed to detect Sc-Col1a1 at the level of Meckel's cartilage, albeit an intense staining was observed in dermal denticles located on the same section and serving as an internal positive control (**Figures 1K,K'**). Sc-Col1a2 and Sc-Col2a1 transcripts were detected, respectively, in the fibrous perichondrium of Meckel's cartilage (**Figures 1L,L'**) and in the chondrocytes of the cartilaginous element (**Figures 1M,M'**). Immunofluorescence experiments performed on developing *S.c.* fins and jaws further confirmed the cartilage-specific expression of the Sc-Col2a1 protein (**Figures 1G,G';N,N'**). The punctuated localization of Sc-Col2a1 around the cell body of fin and jaw chondrocytes might result from low levels of expression, and is consistent with the concentration of this protein in the pericellular matrix, as reported in other species (Benjamin and Ralphs, 1991; Mizoguchi et al., 1997; Nah et al., 2001). Taken together, our results support the idea that *S.c.* tesserae growth and calcification occur within a Type I-negative and Type II-positive collagenous microenvironment (**Figures 1J–N'**).

Skeletal Expression of the Major Fibrillar Collagen Genes in *Scyliorhinus canicula* Vertebrae

The transverse sections of 6 cm embryos shown in **Figures 2A–D** reveal that the *S.c.* vertebrae are cartilaginous, devoid of calcification, and express Sc-Col2a1 (in chondrocytes of the centrum and the neural arches) and Sc-Col1a1 and Sc-Col1a2 (in the perichondrium surrounding all vertebral elements). In the vertebral column of 7 cm-long embryos, Alcian blue stains the cartilaginous vertebral body and the neural arches (**Figures 2E–I'**). Alizarin red specifically stains the fibrous perichondrium of the neural arches as well as an internal calcification ring located within the centrum and surrounding the notochord, as reported in other chondrichthyan species (see **Figures 2E–K** and Peignoux-Deville et al., 1982; Eames et al., 2007). Histologically, the calcified ring of the vertebral body exhibits darker HES staining of the matrix surrounding large cells of chondrocytic appearance (**Figures 2G,J**). By contrast, cells located in the calcifying extracellular matrix of the neural arches are thin with reduced amount of cytoplasm (**Figures 2H,K**).

The expression of Sc-Col1a1 and Sc-Col1a2 was evident in the fibrous perichondrium and the connective tissue surrounding all vertebral elements (**Figures 2L–M'**) as well as in scattered cells embedded in the calcified layer of the neural arches (arrowheads in **Figures 2L,M'**). Nor Sc-Col1a1 nor Sc-Col1a2 were detected in the calcified layer of the vertebral body (the lighter ring-shaped signal in **Figures 2L–M'** is identical to the background observed in negative controls, see **Data Sheet 2**). While Sc-Col2a1 is expressed in most vertebral chondrocytes, it is significantly downregulated in cells embedded within the calcifying layer of the vertebral body (**Figures 2N,N'**).

Likewise, an anti-type II collagen antibody intensely stained the cartilaginous, non-calcified, vertebral cartilage of the neural arches, and the centrum, as well as a thin layer surrounding the notochord (**Figures 2O,O'**). In agreement with the *in situ* hybridization results, the calcifying regions of the neural arches and of the vertebral body displayed a much fainter reaction to the Type II collagen antibody (arrowheads in **Figures 2O,O'**). Taken together, these observations reveal a negative correlation between Sc-Col2a1 expression and extracellular matrix calcification. By contrast, Sc-Col1a1 and Sc-Col1a2 are expressed in all perichondral cells of the vertebrae, regardless of their calcification degree.

Skeletal Expression of the Major Fibrillar Collagen Genes in the *Xenopus tropicalis* Limb

We examined the expression of Xt-Col1a1, Xt-Col1a2, and Xt-Col2a1 in the diaphysis and epiphysis of *X.t.* hindlimbs both before (stage NF54, **Figures 3A–C**) and after (stage NF60, **Figures 3M–O**) ossification. At stage NF54, Xt-Col1a1, and Xt-Col1a2 are most strongly expressed in perichondrial cells of developing long bones (**Figures 3D–I**). At stage NF60, Xt-Col1a1, and Xt-Col1a2 transcripts are robustly detected in osteoblasts and in some osteocytes, albeit more weakly (**Figures 3P–U**). Finally, Xt-Col2a1 is expressed in all chondrocytes of NF54 non-calcified cartilaginous elements (**Figures 3J–L**), and is restricted to the epiphyseal chondrocytes at stage NF60 (**Figures 3V–X**).

Histology of the Developing *Xenopus tropicalis* Vertebrae

Because of the complex shape of the *X.t.* vertebrae, transverse sections either run through the lateral (**Figures 4A,D–F,K–P**) or the dorsal (**Figures 4A,G–I,Q–V**) region of the non-calcified (stage NF54, see **Figures 4B–I**) and calcified (stage NF57, see **Figures 4J–V**) neural arches protecting the neural tube. At stage NF57, the cartilage matrix is abundant (**Figures 4L,R**) and displays a pronounced HES and Alizarin red staining co-localizing at the level of the dorsal region underlying the notochord, and within the lateral and dorsal neural arches (see **Figures 4K,M,N,Q,S,T**). In addition, cartilage calcification and periosteal bone develop in contact to each other (**Figures 4O,P,U,V**).

Skeletal Expression of the Major Fibrillar Collagen Genes in the *Xenopus tropicalis* Vertebrae

Xt-Col1a1, Xt-Col1a2, and Xt-Col2a1 expression patterns were examined in the lateral and dorsal neural arch regions of the vertebrae (see **Figures 4E,F,H,I,O,P,U,V**). At stage NF54, Xt-Col1a1, and Xt-Col1a2 are expressed in scattered cells of mesenchymal appearance located in the vicinity of the cartilage (**Figures 5A,B**), as well as in a thin layer of perichondrium surrounding the dorsal neural arch (**Figures 5D,E**). At this early stage, Xt-Col2a1 is expressed in all chondrocytes and is also evident in the perichondrium of the dorsal neural arch (**Figures 5C,F**). At stage NF57, Xt-Col1a1, and Xt-Col1a2 are robustly expressed in osteoblasts lying onto the calcified

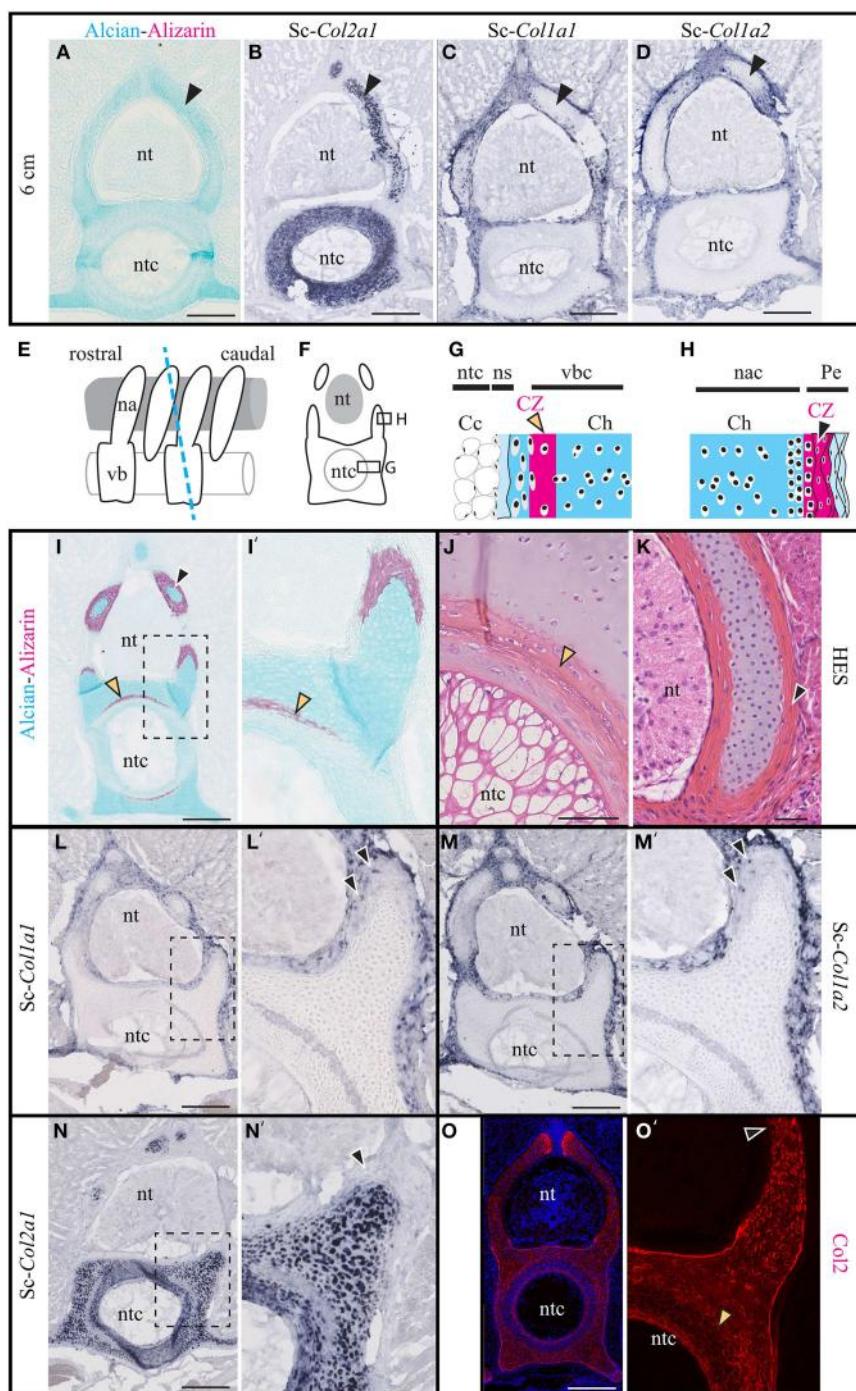


FIGURE 2 | Cartilage calcification and collagen expression in *Scyliorhinus canicula* vertebrae. **(A–D)** Transverse sections of the vertebrae of 6 cm-long embryos (black arrowheads show the hyaline cartilage of the neural arches). **(A)** Alcian blue and Alizarin red double staining revealing the distribution of the hyaline cartilage and the absence of detectable calcification. **(B–D)** *In situ* hybridizations showing the expression of *Sc-Col2a1*, *Sc-Coll1a1*, and *Sc-Coll1a2*, as indicated. **(E)** Schematic drawing of the vertebral anatomy from 9 cm-long *S.c.* embryos (lateral view) and of the orientation of the transverse sections (blue dotted line) represented in **(F)** and shown in **(I–O')**. **(G)** General histology of the centrum. **(H)** General histology of the neural arches. **(I)** Alcian red and Alcian blue double staining. **(J,K)** HES staining of the centrum and of the neural arch. **(L–N)** *In situ* hybridizations showing the expression of *Sc-Col2a1*, *Sc-Coll1a1*, and *Sc-Coll1a2*, as indicated. Arrowheads in **(L',M')** indicate scattered *Sc-Coll1a1* and *Sc-Coll1a2* positive cells embedded in the calcified layer of the neural arches. **(O)** Immunofluorescence using an anti-Type II collagen (Col2) specific antibody. Higher magnifications of **(I,L–O)** are shown in **(I',L–O')** respectively. Orange and black arrowheads show the calcifying matrix of the centrum and neural arches, respectively. **(O')** Col2

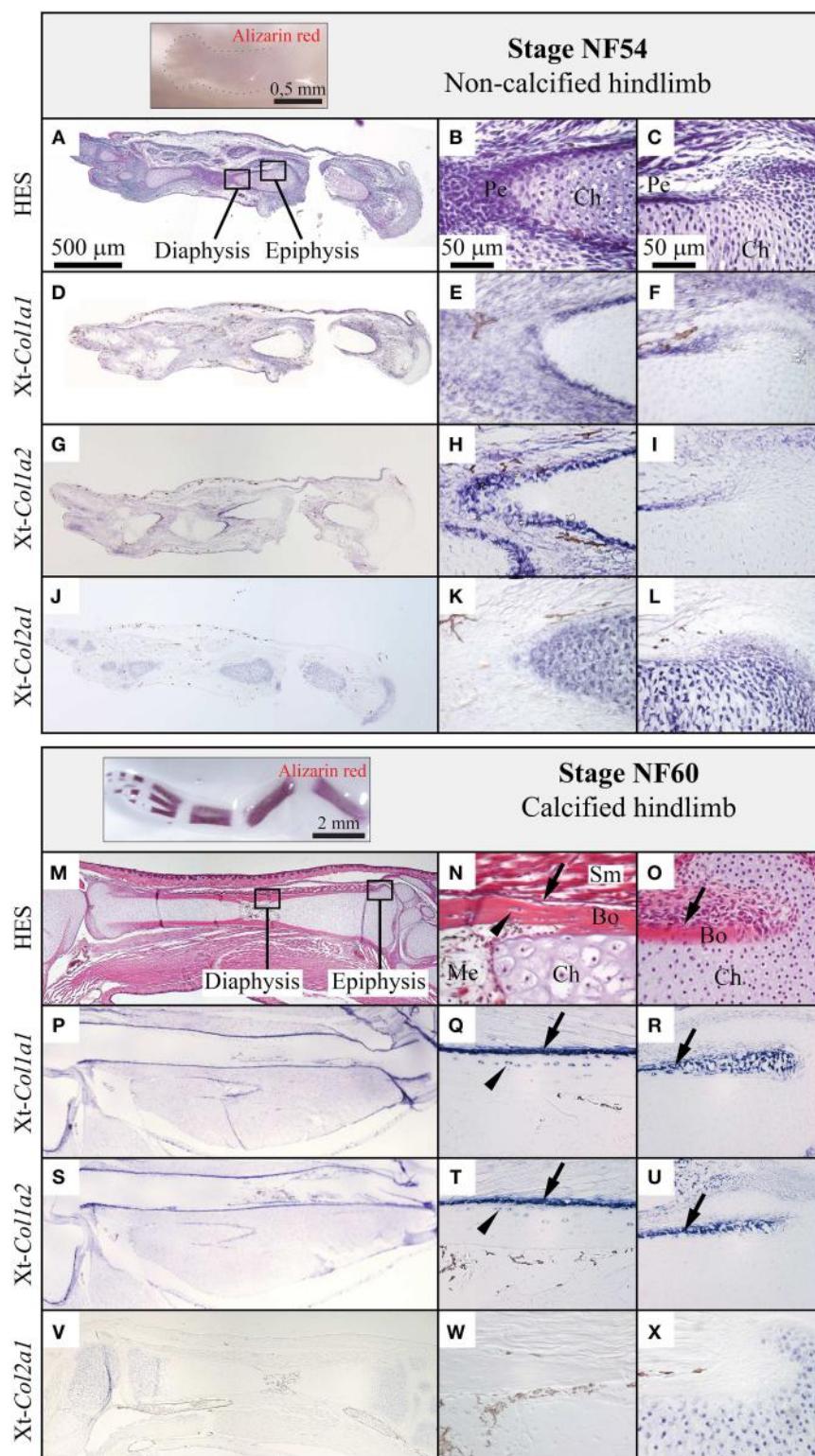


FIGURE 3 | Comparison of the *Col1a1*, *Col1a2*, and *Col2a1* expression patterns during *Xenopus tropicalis* hindlimb development. Stage NF54 (top panel) or NF60 (bottom panel) hindlimbs were examined by whole mount Alizarin red staining (insets), sectioned along the proximo-distal axis and stained with HES, (A–C, M–O) or processed by *in situ* hybridization for the *Xt-Col1a1*, *Xt-Col1a2*, and *Xt-Col2a1* probes, (D–L, P–X). Results are shown for the whole skeletal element (left column, scale bar: 500 μ m) and higher magnifications of the diaphysis (middle column, scale bar: 50 μ m) and epiphysis (right column, scale bar: 50 μ m). Arrows and arrowheads show osteoblasts and osteocytes, respectively. *In situ* hybridization signal is light to dark blue, and brown endogenous *X.t.* pigment cells are visible on most sections. Legend: Bo, bone; Ch, chondrocytes; Me, medulla; Pe, perichondrium; Sm, striated muscles.

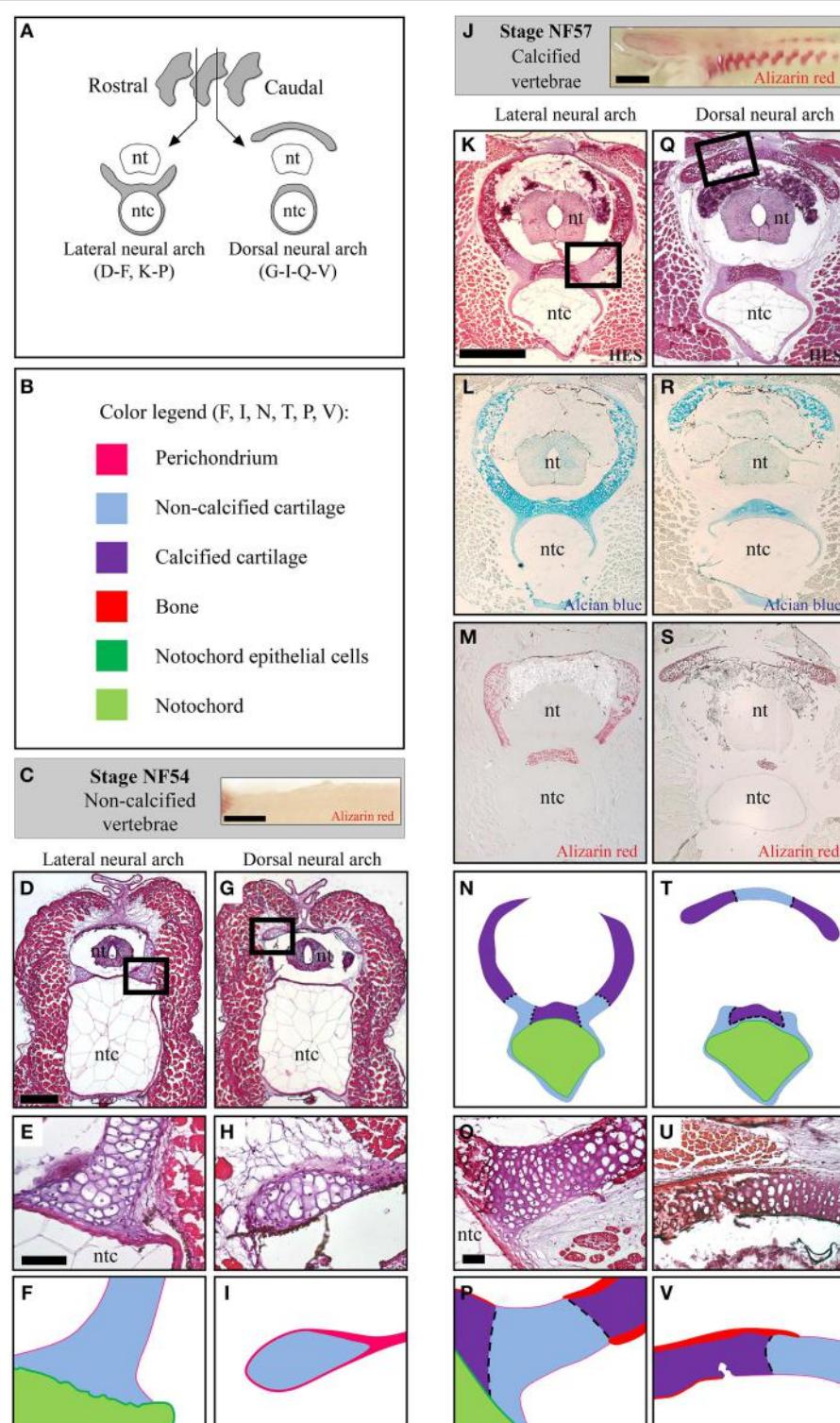


FIGURE 4 | Histology of the developing *Xenopus tropicalis* vertebrae. (A) Schematic drawing of transverse sections running through the lateral or the dorsal region of the neural arches. **(B)** Color code used to represent the distinct skeletal tissues of the *X.t.* vertebrae in **(F,I,N,T,P,V)**. **(C)** Whole mount Alizarin red staining of stage NF54 vertebral column (lateral view, anterior to the left). **(D–I)** Histology of the stage NF54 vertebrae examined with HES **(D,E,G,H)**. **(J)** Whole mount Alizarin red staining of stage NF57 vertebral columns (lateral view, anterior to the left). **(K–V)** Histology of the stage NF57 vertebrae examined with HES **(K,O,Q,U)**, Alcian blue **(L,R)** and Alizarin red **(M,S)**. Insets in **D, G, K** and **Q** are shown in **F, I, P** and **V**, respectively. Panels **E, H, K, O, Q** and **U** are schematized in **F, I, N, P, T** and **V**, respectively. Abbreviations: nt, neural tube; ntc, notochord. Scale bars: 1 mm in **(C,J)**; 250 μ m in **(D,G)**; 50 μ m in **(E,F)** and **(H,I)**; 500 μ m in **(K–N)** and **(Q–T)**; and 50 μ m in **(O,P,U,V)**.

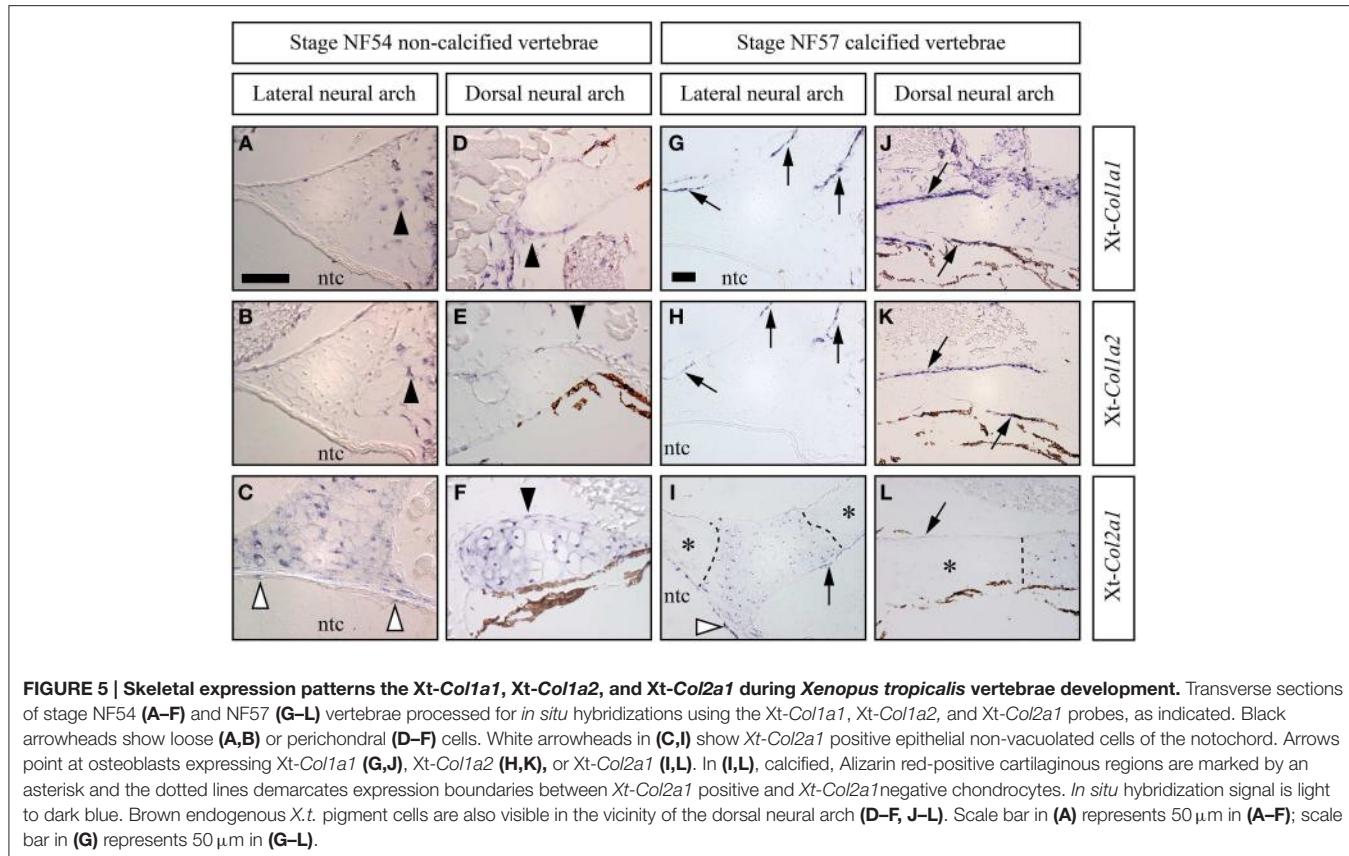


FIGURE 5 | Skeletal expression patterns the *Xt-Col1a1*, *Xt-Col1a2*, and *Xt-Col2a1* during *Xenopus tropicalis* vertebrae development. Transverse sections of stage NF54 (A–F) and NF57 (G–L) vertebrae processed for *in situ* hybridizations using the *Xt-Col1a1*, *Xt-Col1a2*, and *Xt-Col2a1* probes, as indicated. Black arrowheads show loose (A,B) or perichondral (D–F) cells. White arrowheads in (C,I) show *Xt-Col2a1* positive epithelial non-vacuolated cells of the notochord. Arrows point at osteoblasts expressing *Xt-Col1a1* (G,J), *Xt-Col1a2* (H,K), or *Xt-Col2a1* (I,L). In (I,L), calcified, Alizarin red-positive cartilaginous regions are marked by an asterisk and the dotted lines demarcates expression boundaries between *Xt-Col2a1* positive and *Xt-Col2a1* negative chondrocytes. *In situ* hybridization signal is light to dark blue. Brown endogenous *X.t.* pigment cells are also visible in the vicinity of the dorsal neural arch (D–F, J–L). Scale bar in (A) represents 50 µm in (A–F); scale bar in (G) represents 50 µm in (G–L).

bone matrix of the vertebrae (arrows in **Figures 5G,H,J,K**). These osteoblasts also express *Xt-Col2a1*, albeit more weakly than hypertrophic chondrocytes (**Figures 5I,L**). In chondrocytes, *Xt-Col2a1* is excluded from the Alizarin red-positive regions (asterisk in **Figures 5I,L**), forming sharp expression boundaries between calcified and non-calcified cartilage (dotted line in **Figures 4P,V, 5I,L**). In addition, at stages NF54 and NF57, we detected a strong *Xt-Col2a1* staining in the epithelial non-vacuolated cells of the notochord (arrowheads in **Figures 5C,I**), a known site of *Col2a1* expression in cyclostomes and teleosts (Ota and Kuratani, 2010; Yamamoto et al., 2010).

Discussion

Conserved Early Molecular Patterning of the Hyaline Cartilage and Non-calcified Perichondrium

In non-calcified *S.c.* skeletal elements, the expression patterns of the *Col1a1/Col1a2* (perichondrium) and *Col2a1* (cartilage) genes do not overlap. By contrast, in actinopterygians, *Col2a1* orthologs are expressed in the perichondrium, albeit at lower levels than in cartilage (Albertson et al., 2010; Eames et al., 2012). Likewise, our results in *X.t.* reveal a faint *Xt-Col2a1* expression in the non-calcified perichondrium of the dorsal neural arch at stage NF54. It is likely that more sensitive techniques will help assess the expression levels of *Xt-Col2a1* in the perichondrium

of the *X.t.* lateral neural arch or hindlimb, two sites where it was not detected by *in situ* hybridization. Interestingly, Clade A fibrillar collagen members from lamprey and hagfish are expressed both in perichondral cells and in chondrocytes, while the amphioxus ortholog is expressed in chondrocytes and in the mesenchyme located at the tip of regenerating cirri (Zhang and Cohn, 2006, 2008; Zhang et al., 2006; Ota and Kuratani, 2010; Cattell et al., 2011; Kaneto and Wada, 2011). Altogether, these data suggest that the largely complementary expression patterns of *Col1a1/Col1a2* (exclusively in the fibrous perichondrium) and *Col2a1* (preferentially in the hyaline cartilage) represent a synapomorphy of non-calcified skeletal elements in jawed vertebrates. It is therefore tempting to propose that the Clade A precursor was expressed in chondrocytes and perichondral cells, and that the functional partitioning of ancestral enhancers was involved in this expression divergence (Force et al., 1999; Zhang and Cohn, 2008). According to this scenario, after the genomic duplications that gave rise to the complete set of Clade A members, the *Col1a1* and *Col1a2* genes would have rapidly lost their cartilage-specific enhancers, while the activity of perichondral *Col2a1* enhancers would have been dramatically reduced, or abolished, in distinct jawed vertebrate lineages.

Col2a1 Osteoblastic Expression was Significantly Reduced in the Tetrapod Lineage

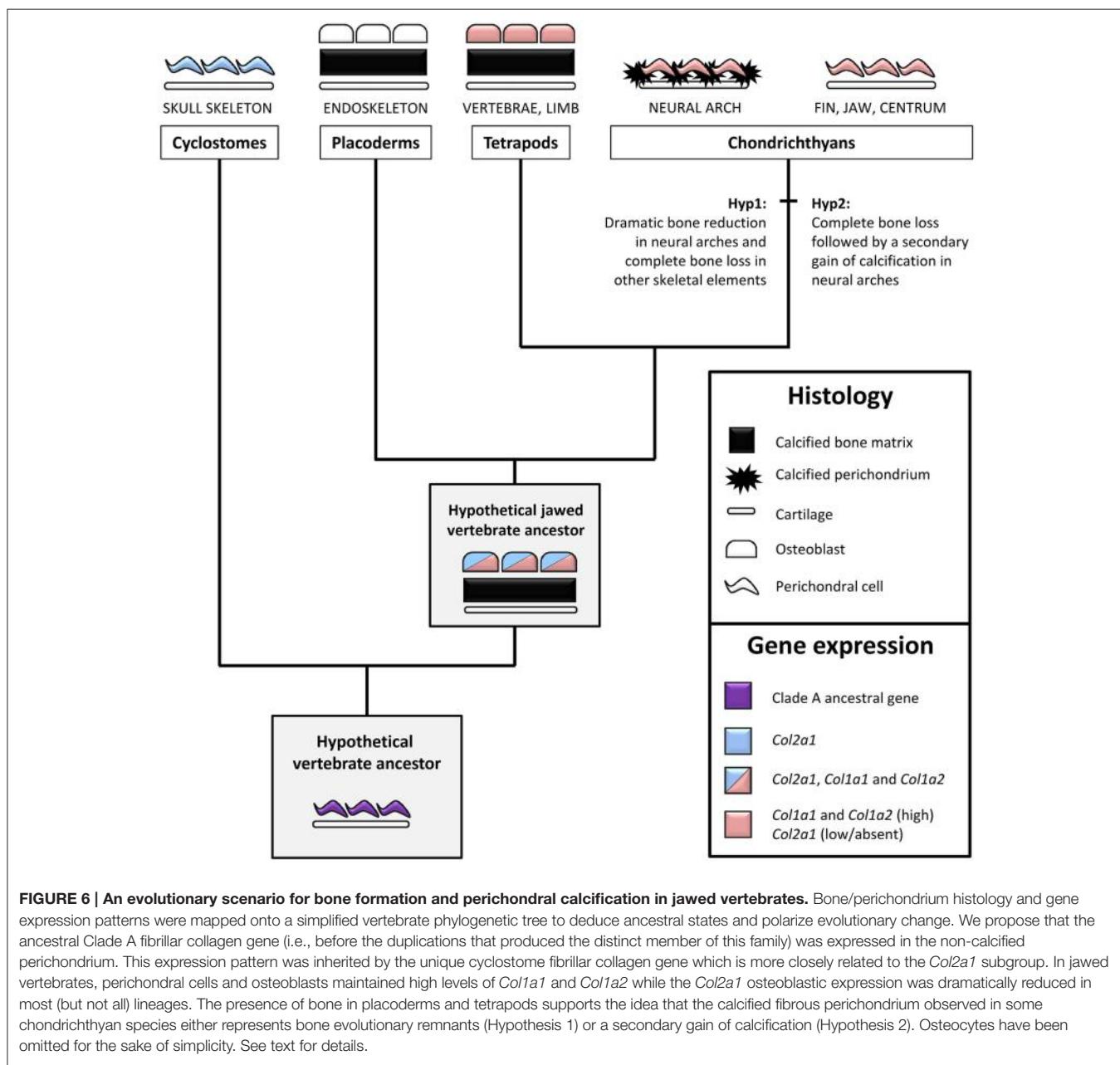
We detected *X.t.* *Col2a1* transcripts in osteoblasts of the vertebrae, albeit they displayed a weaker *in situ* hybridization

signal than hypertrophic chondrocytes present on the same section (Figures 5I,L), which is consistent with expression results obtained with primary cultures of *X.t.* osteoblasts (Bertin et al., 2015). While *Col2a1* is traditionally considered to be a chondrocyte-specific marker (Kobayashi and Kronenberg, 2005; Hartmann, 2009), its robust osteoblastic expression has been reported in embryos from several species of actinopterygian fishes (Benjamin and Ralphs, 1991; Albertson et al., 2010; Eames et al., 2012). The moderate *Col2a1* expression levels described in the clawed frog (this study), chick (Abzhanov et al., 2007) and mouse (Hilton et al., 2007) therefore support the idea that the osteogenic transcription of *Col2a1* was significantly reduced in the tetrapod lineage, and almost completely abolished in mammals (Figure 6).

***Scyliorhinus canicula* Neural Arches, Tesserae, and Centrum Calcification Occur in Distinct Molecular Contexts**

Our results reveal that at least three skeletal sites expressing different combinations of collagen genes are associated with robust *S.c.* calcification in: (i) the fibrous perichondrium of the neural arches, (ii) the tesserae developing in Meckel's cartilage, and (iii) the compact cartilage embedded within the vertebral bodies.

In neural arches, the cartilaginous scaffold is surrounded by a fibrous perichondrium whose matrix is highly calcified and devoid of Col2 protein, and whose cells express Sc-*Col1a1* and Sc-*Col1a2* and no detectable levels of Sc-*Col2a1* (Figures 2, 5). The evolutionary relationship between this calcified perichondrium



and the osteichthyan bone has remained enigmatic and controversial (Peignoux-Deville et al., 1982; Eames et al., 2007; Zhang et al., 2009; Ryll et al., 2014). In the light of fossil evidence demonstrating that extant chondrichthyans are quite derived, having lost the perichondral bone surrounding the cartilaginous elements (Coates et al., 1998; Donoghue and Sansom, 2002), two hypotheses might account for the unusual calcification pattern observed in neural arches (**Figure 6**). On the one hand, it is possible that the perichondral bone was dramatically reduced to some evolutionary remnants of calcified fibrous perichondrium located in the neural arches (hypothesis 1). In this case, the cells involved in matrix calcification would correspond to highly derived osteoblasts having lost many crucial cellular features typically observed in osteichthyans, such as the ability to organize as a polarized pseudoepithelium (Izu et al., 2011; Liu et al., 2011). On the other hand, the perichondral bone might have been completely lost, and secondarily compensated by an independent ability to calcify the perichondral extracellular matrix (hypothesis 2). Below, we discuss two complementary strategies that might help resolve this issue. Firstly, a broader phylogenetic sampling is required to precisely assess the occurrence of a calcified perichondrium in neural arches, which currently seems to be limited to some chondrichthyan species. For instance, the skeleton of holocephalans displays little or no calcified tissue (a ring-shaped calcification of the centrum is reported in some fossil holocephalan and in the extant genus *Chimaera*) while batoids (rays and skates) have a tesserae-based calcification at the surface of their vertebral units (Reynolds, 1897; Goodrich, 1930; Zangerl, 1981). Secondly, it will be important to investigate the nature of the *Col1a1* and *Col1a2* positive cells embedded within the mineralized matrix (**Figures 2E,F'**). Indeed, such cells have been proposed to be osteocytes (Peignoux-Deville et al., 1982), which is consistent with the fact that cellular bone evolved before the origin of the jawed vertebrates (Donoghue and Sansom, 2002; Donoghue et al., 2006; Sanchez et al., 2013). Extensive phenotypical and molecular similarities between the scattered cells embedded within the *S.c.* calcified perichondrium and osteichthyan osteocytes would support their homology, and, therefore, the aforementioned hypothesis 1.

Another site of calcification in *S.c.* corresponds to the developing tesserae embedded in Meckel's cartilage, a process classically described to occur at the surface of the cartilaginous skeletal piece (Kemp and Westrin, 1979; Dean et al., 2009). As we show here, the onset of this type of calcification takes place in a Col2-positive context, within the cartilaginous scaffold (**Figures 1, 6**). We failed to detect *Col1a1/Col1a2* expression in the chondrocytes neighboring the mineralized matrix, suggesting that the cellular processes involved in matrix calcification are very different from what has been described in osteichthyan bone or chondroid bone (Mizoguchi et al., 1997). This type of calcification is well developed in extant batoid and selachimorph species, and is also known in fossil holocephalan species (Grogan and Lund, 2000; Finarelli and Coates, 2014) and, therefore, is considered to be an early evolutionary innovation of the chondrichthyan lineage (**Figure 7**).

Below, we will discuss the third type of calcification mechanism, which occurs in the *Col1a1/Col1a2*

negative *S.c.* vertebral cartilage experiencing a drastic *Col2a1* downregulation, in the light of the striking similarities that it shares with the *X.t.* vertebrae.

An Ancient Type of Calcified Vertebral Cartilage Associated to the Down-regulation of *Col2a1*

The tetrapod hyaline cartilage calcifies its extracellular matrix, albeit to a much lesser extent than the bone tissue (Claassen et al., 1996; Khanarian et al., 2014) and, therefore, only weakly stains with Alizarin red (Kirsch et al., 1997). Here, we report an unusual type of calcified cartilage displaying remarkable similarities between *X.t.* and *S.c.* at three distinct biological levels: (i) anatomically, this cartilage is located in the vertebrae of both species, and, at least at the stages analyzed, in no other skeletal elements; (ii) from an histological perspective its robust calcification is reflected by intense Alizarin red and HES stainings; (iii) molecularly, both types of cartilages are *Col1a1/Col1a2* negative and probably experience a *Col2a1* downregulation, because in both species all cells of the vertebral cartilage express *Col2a1* during early, non-calcified, developmental stages (see **Figures 2B,N,N', 5C,E,I,L**). In this respect, both types of vertebral cartilages seem to recapitulate the initial phase of endochondral bone formation typically seen in tetrapod long bones, during which proliferative chondrocytes progressively downregulate the expression of *Col2a1*, undergo hypertrophy, and calcify their extracellular matrix (**Figure 7**). Our observations, combined to data from mouse (Chandraraj and Briggs, 1988), and lizards (Lozito and Tuan, 2015), suggest that an calcified form of vertebral cartilage was present in the last common ancestor of jawed vertebrates, at least as a transitory developmental process.

As vertebral developmental processes are highly variable, homology relationships between the calcified ring surrounding the *S.c.* notochord and the calcified cartilage of the *X.t.* vertebrae cannot be inferred (Fleming et al., 2015). Rather, we propose that the genetic programme involving a downregulation of the *Col2a1* gene predates the emergence of the last vertebrate common ancestor, and was subsequently co-opted and modified to produce a variety of novel non-calcified (Zhang and Cohn, 2006; Zhang et al., 2009) and calcified (Hogg, 1982; Claassen et al., 1996; Janvier and Arsenault, 2002; Porter et al., 2007) cartilaginous structures (**Figure 7**). One intriguing possibility is that the ancient, *Col2a1*-negative, calcified cartilage present in the last common ancestor of jawed vertebrates later came to play a key role in the subsequent elimination of cartilaginous matrix and its replacement by bone tissue. In this respect, it might have served as a crucial pre-patterning step contributing to the emergence of endochondral ossification commonly observed in tetrapods and whose precise origin still remains to be determined. In the future, a comprehensive comparison of gene expression signatures between cell types present in diverse skeletal tissues, anatomical locations, developmental stages, and species will provide a solid basis to unravel the complex and fascinating evolutionary history of the vertebrate skeleton.

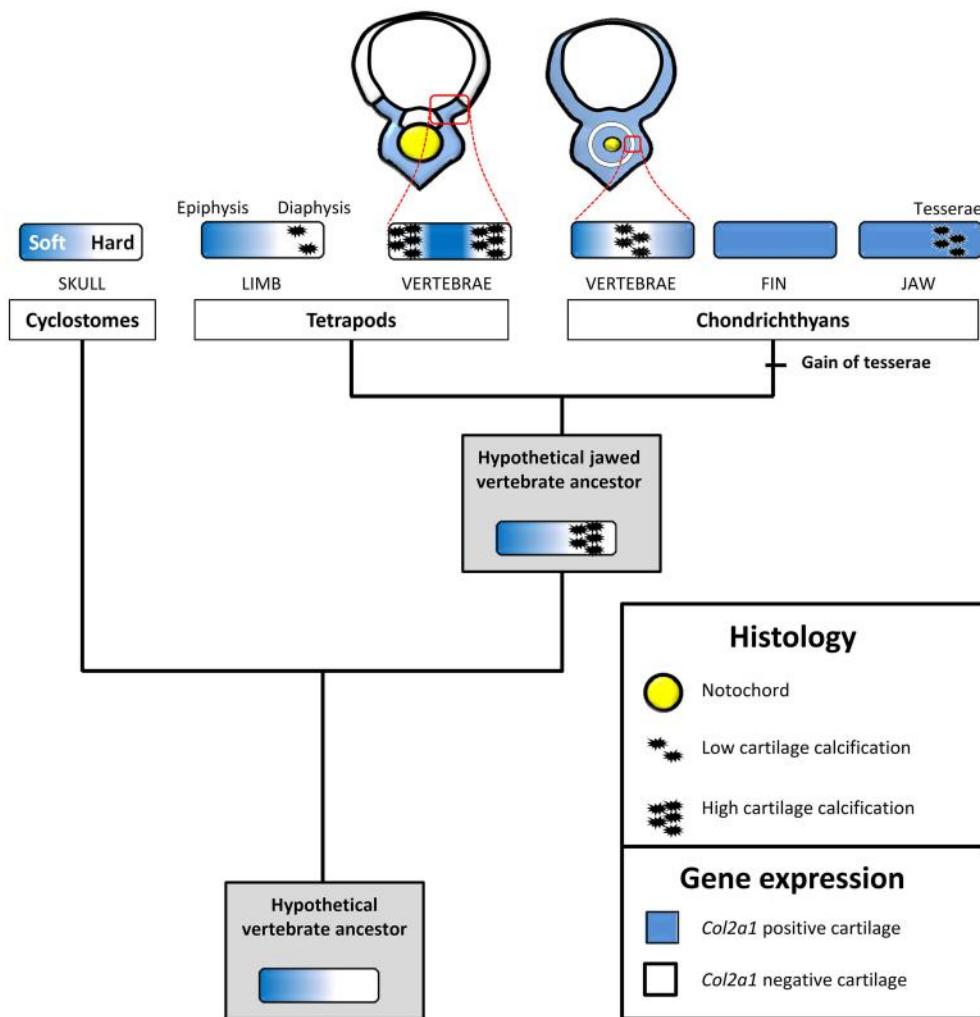


FIGURE 7 | An evolutionary scenario for cartilage calcification in jawed vertebrates. Expression patterns and cartilage matrix calcification were mapped onto a simplified vertebrate phylogenetic tree to deduce ancestral states and polarize evolutionary change. We propose that, in the last vertebrate common ancestor, the expression of *Col2a1* experienced a strong downregulation in maturing, non-calcified, cartilaginous regions. This downregulation was subsequently inherited by distinct vertebrate lineages, and is associated to hard cartilage in cyclostomes and to calcified cartilage in jawed vertebrates. The chondrichthyan and osteichthyan representatives analyzed in this study display a calcified *Col2a1*-negative vertebral cartilage, a likely jawed vertebrate synapomorphy. Tesserae calcification, a recent chondrichthyan innovation, occurs in the absence of *Col2a1* downregulation. Perichondrium and bone have been omitted for the sake of simplicity. See text for details.

Acknowledgments

We are grateful to both reviewers for their constructive comments that improved the quality of the present manuscript. We thank Isabelle Germon, Marie-Ka Tilak and Fabienne Justy for help with molecular biology. This research was funded by a FONDECYT grant 1151196 to SM and by PEPS ExoMOD to MDT. This is ISEM contribution # ISEM 2015-146.

Supplementary Material

The Supplementary Material for this article can be found online at: <http://journal.frontiersin.org/article/10.3389/fgene.2015.00283>

Supplementary Table 1 | list of *Scyliorhinus canicula* and *Xenopus tropicalis* specific primers.

Data Sheet 1 | Sequences of the S.c. clones.

Data Sheet 2 | Sense probe negative *in situ* hybridization results for *Scyliorhinus canicula* and *Xenopus tropicalis*.

Data Sheet 3 | X.t. *in situ* hybridization protocol.

Data Sheet 4 | Molecular Phylogenetic analysis of gnathostome Clade A fibrillar collagen genes. Phylogenetic relationships were inferred and the tree with the highest log likelihood (-7413.8950) is shown. The percentages of trees in which the associated taxa clustered together are indicated. The tree is drawn to scale, and branch lengths correspond to the number of substitutions per site. Orthology groups (OG) are identified as blue circles, and S.c. sequences are shown in red. The tree was rooted according to (Zhang and Cohn, 2008).

Data Sheet 5 | Fibrillar collagen sequences.

References

- Abzhanov, A., Rodda, S. J., McMahon, A. P., and Tabin, C. J. (2007). Regulation of skeletogenic differentiation in cranial dermal bone. *Development* 134, 3133–3144. doi: 10.1242/dev.002709
- Albertson, R. C., Yan, Y. L., Titus, T. A., Pisano, E., Vacchi, M., Yelick, P. C., et al. (2010). Molecular pedomorphism underlies craniofacial skeletal evolution in Antarctic notothenioid fishes. *BMC Evol. Biol.* 10:4. doi: 10.1186/1471-2148-10-4
- Aldea, D., Hanna, P., Munoz, D., Espinoza, J., Torrejon, M., Sachs, L., et al. (2013). Evolution of the vertebrate bone matrix: an expression analysis of the network forming collagen paralogues in amphibian osteoblasts. *J. Exp. Zool. B Mol. Dev. Evol.* 320, 375–384. doi: 10.1002/jez.b.22511
- Arendt, D. (2008). The evolution of cell types in animals: emerging principles from molecular studies. *Nat. Rev. Genet.* 9, 868–882. doi: 10.1038/nrg2416
- Benjamin, M., and Evans, E. J. (1990). Fibrocartilage. *J. Anat.* 171, 1–15.
- Benjamin, M., and Ralphs, J. R. (1991). Extracellular matrix of connective tissues in the heads of teleosts. *J. Anat.* 179, 137–148.
- Bertin, A., Hanna, P., Otarola, G., Fritz, A., Henriquez, J. P., and Marcellini, S. (2015). Cellular and molecular characterization of a novel primary osteoblast culture from the vertebrate model organism *Xenopus tropicalis*. *Histochem. Cell Biol.* 143, 431–442. doi: 10.1007/s00418-014-1289-8
- Cattell, M., Lai, S., Cerny, R., and Medeiros, D. M. (2011). A new mechanistic scenario for the origin and evolution of vertebrate cartilage. *PLoS ONE* 6:e22474. doi: 10.1371/journal.pone.0022474
- Chandraraj, S., and Briggs, C. A. (1988). Role of cartilage canals in osteogenesis and growth of the vertebral centra. *J. Anat.* 158, 121–136.
- Claassen, H., Kampen, W. U., and Kirsch, T. (1996). Localization of collagens and alkaline phosphatase activity during mineralization and ossification of human first rib cartilage. *Histochem. Cell Biol.* 105, 213–219. doi: 10.1007/BF01462294
- Close, B., Banister, K., Baumans, V., Bernoth, E. M., Bromage, N., Bunyan, J., et al. (1996). Recommendations for euthanasia of experimental animals: Part 1. DGXI of the European Commission. *Lab. Anim.* 30, 293–316. doi: 10.1258/002367796780739871
- Coates, M. I., Sequeira, S. E. K., Sansom, I. J., and Smith, M. M. (1998). Spines and tissues of ancient sharks. *Nature* 396, 729–730. doi: 10.1038/25467
- Dahdul, W. M., Balhoff, J. P., Blackburn, D. C., Diehl, A. D., Haendel, M. A., Hall, B. K., et al. (2012). A unified anatomy ontology of the vertebrate skeletal system. *PLoS ONE* 7:e51070. doi: 10.1371/journal.pone.0051070
- Dean, M. N., Chiou, W. A., and Summers, A. P. (2005). Morphology and ultrastructure of prismatic calcified cartilage. *Microsc. Microanal.* 11, 1196–1197. doi: 10.1017/S1431927605507888
- Dean, M. N., Mull, C. G., Gorb, S. N., and Summers, A. P. (2009). Ontogeny of the tessellated skeleton: insight from the skeletal growth of the round stingray *Urotrygon halleri*. *J. Anat.* 215, 227–239. doi: 10.1111/j.1469-7580.2009.01116.x
- Dean, M. N., and Summers, A. P. (2006). Mineralized cartilage in the skeleton of chondrichthyan fishes. *Zoology (Jena.)* 109, 164–168. doi: 10.1016/j.zool.2006.03.002
- Donoghue, P. C., 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., 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
- Eames, B. F., Allen, N., Young, J., Kaplan, A., Helms, J. A., and Schneider, R. A. (2007). Skeletogenesis in the swell shark *Cephaloscyllium ventriosum*. *J. Anat.* 210, 542–554. doi: 10.1111/j.1469-7580.2007.00723.x
- Eames, B. F., Amores, A., Yan, Y. L., and Postlethwait, J. H. (2012). Evolution of the osteoblast: skeletogenesis in gar and zebrafish. *BMC Evol. Biol.* 12:27. doi: 10.1186/1471-2148-12-27
- Eames, B. F., de la Fuente, L., and Helms, J. A. (2003). Molecular ontogeny of the skeleton. *Birth Defects Res. C Embryo Today* 69, 93–101. doi: 10.1002/bdrc.10016
- 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., Silva, N., Redruello, B., Costa, R., Gregório, S., Canário, A. V., 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
- Eyre, D. R., and Weis, M. A. (2013). Bone collagen: new clues to its mineralization mechanism from recessive osteogenesis imperfecta. *Calcif. Tissue Int.* 93, 338–347. doi: 10.1007/s00223-013-9723-9
- Finarelli, J. A., and Coates, M. I. (2014). *Chondrenchelys problematica* (Traquair, 1888) redescribed: a Lower Carboniferous, eel-like holocephalan from Scotland. *Earth Environ. Sci. Trans. R. Soc. Edinb.* 105, 35–59. doi: 10.1017/S1755691014000139
- Fleming, A., Kishida, M. G., Kimmel, C. B., and Keynes, R. J. (2015). Building the backbone: the development and evolution of vertebral patterning. *Development* 142, 1733–1744. doi: 10.1242/dev.118950
- 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.
- Goodrich, E. S. (ed.). (1930). *Studies on the Structure and Development of Vertebrates*. London: Macmillan.
- Grogan, E. D., and Lund, R. (2000). Debeerius ellefseni (Fam. Nov., Gen. Nov., Spec. Nov.), an autodiastylid chondrichthyan from the Mississippian bear gulch limestone of Montana (USA), the relationships of the chondrichthyes, and comments on gnathostome evolution. *J. Morphol.* 243, 219–245. doi: 10.1002/(SICI)1097-4687(200003)243:3<219::AID-JMOR1>3.0.CO;2-1
- Hartmann, C. (2009). Transcriptional networks controlling skeletal development. *Curr. Opin. Genet. Dev.* 19, 437–443. doi: 10.1016/j.gde.2009.09.001
- Hilton, M. J., Tu, X., and Long, F. (2007). Tamoxifen-inducible gene deletion reveals a distinct cell type associated with trabecular bone, and direct regulation of PTTHrP expression and chondrocyte morphology by Ihh in growth region cartilage. *Dev. Biol.* 308, 93–105. doi: 10.1016/j.ydbio.2007.05.011
- Hogg, D. A. (1982). Ossification of the laryngeal, tracheal and syringeal cartilages in the domestic fowl. *J. Anat.* 134, 57–71.
- Huysseune, A., and Sire, J. Y. (1990). Ultrastructural observations on chondroid bone in the teleost fish *Hemicromis bimaculatus*. *Tissue Cell* 22, 371–383. doi: 10.1016/0040-8166(90)90011-W
- Huysseune, A., and Verraes, W. (1986). Chondroid bone on the upper pharyngeal jaws and neurocranial base in the adult fish *Astatotilapia elegans*. *Am. J. Anat.* 177, 527–535. doi: 10.1002/aja.1001770411
- Izu, Y., Sun, M., Zwolanek, D., Veit, G., Williams, V., Cha, B., et al. (2011). Type XII collagen regulates osteoblast polarity and communication during bone formation. *J. Cell Biol.* 193, 1115–1130. doi: 10.1083/jcb.201010010
- Janvier, P. (2015). Facts and fancies about early fossil chordates and vertebrates. *Nature* 520, 483–489. doi: 10.1038/nature14437
- Janvier, P., and Arsenault, M. (2002). Palaeobiology: calcification of early vertebrate cartilage. *Nature* 417, 609. doi: 10.1038/417609a
- Kaneto, S., and Wada, H. (2011). Regeneration of amphioxus oral cirri and its skeletal rods: implications for the origin of the vertebrate skeleton. *J. Exp. Zool. B Mol. Dev. Evol.* 316, 409–417. doi: 10.1002/jez.b.21411
- Kemp, N. E., and Westrin, S. K. (1979). Ultrastructure of calcified cartilage in the endoskeletal tesserae of sharks. *J. Morphol.* 160, 75–109. doi: 10.1002/jmor.1051600106
- Kerney, R., and Hanken, J. (2008). Gene expression reveals unique skeletal patterning in the limb of the direct-developing frog, *Eleutherodactylus coqui*. *Evol. Dev.* 10, 439–448. doi: 10.1111/j.1525-142X.2008.00255.x
- Khanarian, N. T., Boushell, M. K., Spalazzi, J. P., Pleshko, N., Boskey, A. L., and Lu, H. H. (2014). FTIR-I compositional mapping of the cartilage-to-bone interface as a function of tissue region and age. *J. Bone Miner. Res.* 29, 2643–2652. doi: 10.1002/jbm.2284
- Kirsch, T., Nah, H. D., Shapiro, I. M., and Pacifici, M. (1997). Regulated production of mineralization-competent matrix vesicles in hypertrophic chondrocytes. *J. Cell Biol.* 137, 1149–1160. doi: 10.1083/jcb.137.5.1149
- Kobayashi, T., and Kronenberg, H. (2005). Minireview: transcriptional regulation in development of bone. *Endocrinology* 146, 1012–1017. doi: 10.1210/en.2004-1343

- Kranenborg, S., van Cleynenbreugel, T., Schipper, H., and Van Leeuwen, J. (2005). Adaptive bone formation in acellular vertebrae of sea bass (*Dicentrarchus labrax* L.). *J. Exp. Biol.* 208, 3493–3502. doi: 10.1242/jeb.01808
- Landis, W. J., and Silver, F. H. (2009). Mineral deposition in the extracellular matrices of vertebrate tissues: identification of possible apatite nucleation sites on type I collagen. *Cells Tissues Organs* 189, 20–24. doi: 10.1159/000151454
- Le, S. Q., and Gascuel, O. (2008). An improved general amino acid replacement matrix. *Mol. Biol. Evol.* 25, 1307–1320. doi: 10.1093/molbev/msn067
- 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
- Liu, L., Schlesinger, P. H., Slack, N. M., Friedman, P. A., and Blair, H. C. (2011). High capacity Na(+) /H(+) exchange activity in mineralizing osteoblasts. *J. Cell. Physiol.* 226, 1702–1712. doi: 10.1002/jcp.22501
- Long, F. (2011). Building strong bones: molecular regulation of the osteoblast lineage. *Nat. Rev. Mol. Cell Biol.* 13, 27–38. doi: 10.1038/nrm3254
- Lozito, T. P., and Tuan, R. S. (2015). Lizard tail regeneration: regulation of two distinct cartilage regions by Indian hedgehog. *Dev. Biol.* 399, 249–262. doi: 10.1016/j.ydbio.2014.12.036
- Mizoguchi, I., Takahashi, I., Sasano, Y., Kagayama, M., Kuboki, Y., and Mitani, H. (1997). Localization of types I, II and X collagen and osteocalcin in intramembranous, endochondral and chondroid bone of rats. *Anat. Embryol.* 196, 291–297. doi: 10.1007/s004290050098
- Nah, H. D., Swoboda, B., Birk, D. E., and Kirsch, T. (2001). Type IIA procollagen: expression in developing chicken limb cartilage and human osteoarthritic articular cartilage. *Dev. Dyn.* 220, 307–322. doi: 10.1002/dvdy.1109
- Nieuwkoop, P. D., and Faber, J. (1967). *Normal Table of Xenopus laevis (Daudin)*. Amsterdam: North Holland.
- Omelon, S., Georgiou, J., Variola, F., and Dean, M. N. (2014). Colocation and role of polyphosphates and alkaline phosphatase in apatite biominerization of elasmobranch tesserae. *Acta Biomater.* 10, 3899–3910. doi: 10.1016/j.actbio.2014.06.008
- Ota, K. G., and Kuratani, S. (2010). Expression pattern of two collagen type 2 alpha1 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, 157–165. doi: 10.1002/jez.b.21322
- Oulion, S., Debiais-Thibaud, M., D'Aubenton-Carafa, Y., Thermes, C., Da Silva, C., Bernard-Samain, S., et al. (2010). Evolution of Hox gene clusters in gnathostomes: insights from a survey of a shark (*Scyliorhinus canicula*) transcriptome. *Mol. Biol. Evol.* 27, 2829–2838. doi: 10.1093/molbev/msq172
- Peignoux-Deville, J., Lallier, F., and Vidal, B. (1982). Evidence for the presence of osseous tissue in dogfish vertebrae. *Cell Tissue Res.* 222, 605–614. doi: 10.1007/BF00213858
- Porter, M. E., Koob, T. J., and Summers, A. P. (2007). The contribution of mineral to the material properties of vertebral cartilage from the smooth-hound shark *Mustelus californicus*. *J. Exp. Biol.* 210, 3319–3327. doi: 10.1242/jeb.006189
- 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
- Reynolds, S. H. (ed.). (1897). *The Vertebrate Skeleton*. Cambridge Zoological Series. Cambridge: Cambridge University Press.
- Rychel, A. L., and Swalla, B. J. (2007). Development and evolution of chordate cartilage. *J. Exp. Zool. B Mol. Dev. Evol.* 308, 325–335. doi: 10.1002/jez.b.21157
- Ryll, B., Sanchez, S., Haitina, T., Tafforeau, P., and Ahlberg, P. E. (2014). The genome of *Callorhinchus* and the fossil record: a new perspective on SCPP gene evolution in gnathostomes. *Evol. Dev.* 16, 123–124. doi: 10.1111/ede.12071
- Sanchez, S., Dupret, V., Tafforeau, P., Trinajstic, K. M., Ryll, B., Gouttenoire, P. J., et al. (2013). 3D microstructural architecture of muscle attachments in extant and fossil vertebrates revealed by synchrotron microtomography. *PLoS ONE* 8:e56992. doi: 10.1371/journal.pone.0056992
- Tamura, K., Stecher, G., Peterson, D., Filipski, A., and Kumar, S. (2013). MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Mol. Biol. Evol.* 30, 2725–2729. doi: 10.1093/molbev/mst197
- Veis, A., and Dorvee, J. R. (2013). Biomineralization mechanisms: a new paradigm for crystal nucleation in organic matrices. *Calcif. Tissue Int.* 93, 307–315. doi: 10.1007/s00223-012-9678-2
- Wada, H., Okuyama, M., Satoh, N., and Zhang, S. (2006). Molecular evolution of fibrillar collagen in chordates, with implications for the evolution of vertebrate skeletons and chordate phylogeny. *Evol. Dev.* 8, 370–377. doi: 10.1111/j.1525-142X.2006.00109.x
- Yamamoto, M., Morita, R., Mizoguchi, T., Matsuo, H., Isoda, M., Ishitani, T., et al. (2010). MiR-Jag1-Notch signalling regulates patterning and structural roles of the notochord by controlling cell-fate decisions. *Development* 137, 2527–2537. doi: 10.1242/dev.051011
- Zangerl, R. (ed.). (1981). *Chondrichthyes I: Paleozoic Elasmobranchii (Handbook of Paleoichthyology)*. Stuttgart: G. Fisher.
- Zhang, G., and Cohn, M. J. (2006). Hagfish and lancelet fibrillar collagens reveal that type II collagen-based cartilage evolved in stem vertebrates. *Proc. Natl. Acad. Sci. U.S.A.* 103, 16829–16833. doi: 10.1073/pnas.0605630103
- Zhang, G., and Cohn, M. J. (2008). Genome duplication and the origin of the vertebrate skeleton. *Curr. Opin. Genet. Dev.* 18, 387–393. doi: 10.1016/j.gde.2008.07.009
- Zhang, G., Eames, B. F., and Cohn, M. J. (2009). Chapter 2. Evolution of vertebrate cartilage development. *Curr. Top. Dev. Biol.* 86, 15–42. doi: 10.1016/S0070-2153(09)01002-3
- Zhang, G., Miyamoto, M. M., and Cohn, M. J. (2006). Lamprey type II collagen and Sox9 reveal an ancient origin of the vertebrate collagenous skeleton. *Proc. Natl. Acad. Sci. U.S.A.* 103, 3180–3185. doi: 10.1073/pnas.0508313103

Conflict of Interest Statement: 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 © 2015 Enault, Muñoz, Silva, Borday-Birraux, Bonade, Oulion, Ventéo, Marcellini 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) or licensor 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.



On the evolutionary relationship between chondrocytes and osteoblasts

Patsy Gómez-Picos and B. Frank Eames *

Department of Anatomy and Cell Biology, University of Saskatchewan, Saskatoon, SK, Canada

OPEN ACCESS

Edited by:

Hector Escriva,
Centre National de la Recherche
Scientifique, France

Reviewed by:

Sylvain Marcellini,
University of Concepcion, Chile
Daniel Medeiros,
University of Colorado Boulder, USA

*Correspondence:

B. Frank Eames,
Department of Anatomy and Cell
Biology, University of Saskatchewan,
3D01-107 Wiggins Road, Saskatoon,
SK S7N 5E5, Canada
b.frank@usask.ca

Specialty section:

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

Received: 02 July 2015

Accepted: 07 September 2015

Published: 23 September 2015

Citation:

Gómez-Picos P and Eames BF (2015)
On the evolutionary relationship
between chondrocytes and
osteoblasts. *Front. Genet.* 6:297.
doi: 10.3389/fgene.2015.00297

Vertebrates are the only animals that produce bone, but the molecular genetic basis for this evolutionary novelty remains obscure. Here, we synthesize information from traditional evolutionary and modern molecular genetic studies in order to generate a working hypothesis on the evolution of the gene regulatory network (GRN) underlying bone formation. Since transcription factors are often core components of GRNs (i.e., kernels), we focus our analyses on Sox9 and Runx2. Our argument centers on three skeletal tissues that comprise the majority of the vertebrate skeleton: immature cartilage, mature cartilage, and bone. Immature cartilage is produced during early stages of cartilage differentiation and can persist into adulthood, whereas mature cartilage undergoes additional stages of differentiation, including hypertrophy and mineralization. Functionally, histologically, and embryologically, these three skeletal tissues are very similar, yet unique, suggesting that one might have evolved from another. Traditional studies of the fossil record, comparative anatomy and embryology demonstrate clearly that immature cartilage evolved before mature cartilage or bone. Modern molecular approaches show that the GRNs regulating differentiation of these three skeletal cell fates are similar, yet unique, just like the functional and histological features of the tissues themselves. Intriguingly, the Sox9 GRN driving cartilage formation appears to be dominant to the Runx2 GRN of bone. Emphasizing an embryological and evolutionary transcriptomic view, we hypothesize that the Runx2 GRN underlying bone formation was co-opted from mature cartilage. We discuss how modern molecular genetic experiments, such as comparative transcriptomics, can test this hypothesis directly, meanwhile permitting levels of constraint and adaptation to be evaluated quantitatively. Therefore, comparative transcriptomics may revolutionize understanding of not only the clade-specific evolution of skeletal cells, but also the generation of evolutionary novelties, providing a modern paradigm for the evolutionary process.

Keywords: **EvoDevo, comparative transcriptomics, Sox9, Runx2, bone, cartilage, GRN**

Introduction: Cartilage and Bone might Share an Evolutionary History

Most of evolutionary theory has focussed on studies of morphological change (morphogenesis) among taxa, but the formation of tissue types (histogenesis) also can evolve in clade-specific manners. Therefore, we focus our attentions on a relatively understudied subject of evolutionary research: the evolution of histogenesis. A classic problem in evolutionary theory is to explain

novelties, or traits with no clear ancestral antecedent (Shubin, 2002; Moczek, 2008; Wagner and Lynch, 2010). For example, vertebrates are the only animals that produce bone, but so far, the molecular genetic basis for this evolutionary novelty remains obscure. Here, we synthesize information from traditional evolutionary and modern molecular studies in order to generate a working hypothesis on the evolution of the genetic system underlying bone formation. Many studies argue that bone evolved from dentine (Kawasaki et al., 2004; Wagner and Aspenberg, 2011). However, using molecular genetic and embryological arguments that favor gradualism over saltationism (Gould, 2002), we hypothesize that bone (and perhaps all mineralizing tissues, such as dentine) appeared during evolution by co-opting a gene regulatory network (GRN) that was under prior natural selection to mineralize cartilage. In order to present an argument for skeletal tissue development and evolution over the past 500 million years, we make some generalizations that may trouble some readers, of whom we ask their indulgence, hoping that such generalizations help to reveal broader trends during the evolution of skeletal tissues.

An introductory look at the similarities and differences among cartilage and bone suggests that the underlying GRNs may be related. Cartilage and bone are specialized connective tissues that provide form and structural support to the body, protect vital organs, and play a crucial role in locomotion through muscle attachments (Gray and Williams, 1989). Despite these similarities, they also have distinct functions (**Figure 1**). Cartilage typically offers a flexible structure to support soft tissues and also to serve as a load-bearing surface between bones. On the other hand, bone is a hard, rigid structure that protects vital organs and acts as a storage site for minerals, such as calcium and phosphorus (Smith and Hall, 1990; Volkmann and Baluska, 2006). Also unlike cartilage, which has almost no capacity for regeneration, bone is a highly dynamic structure that undergoes constant remodeling, preserving bone strength and regulating calcium homeostasis (Datta et al., 2008). Perhaps related to regenerative capacity, these tissues differ in vascularity. Bone is highly vascularized, but cartilage typically is avascular. However, important exceptions to cartilage vascularization occur. Mature cartilage in tetrapods often is invaded by vasculature as it degrades, creating the marrow cavity (Johnson, 1980; Roach, 1997; Stricker et al., 2002; Ortega et al., 2004; Moriishi et al., 2005), and even immature cartilage is highly vascularized near articulating surfaces in some avian and mammalian species (Ytrehus et al., 2004; Blumer et al., 2005). When cartilage extracellular matrix (ECM) undergoes mineralization, its functions change. In some vertebrates, such as sharks, mineralized cartilage can serve as the major rigid structural support for the body, meanwhile providing a mineral reservoir (Daniel, 1934; Kemp and Westrin, 1979; Eames et al., 2007). In most extant vertebrates, however, mineralized cartilage mainly serves as a scaffold during endochondral ossification, outlined below.

During embryonic development, cartilage and bone formation share many features (**Figure 1**). Both cartilage and bone are differentiated from common mesenchymal (osteochondral) progenitor cells (Fang and Hall, 1997; Day et al., 2005; Hill et al., 2005). Both cartilage and bone initiate overt differentiation

by aggregating mesenchymal cells into condensations, which can go on directly to secrete cartilage- or bone-specific matrix (Hall and Miyake, 1995, 2000; Kronenbergs, 2003; Day et al., 2005). However, a unique feature of bone formation is that, in addition to differentiating directly from an osteogenic condensation (intramembranous ossification), bone also forms on a pre-existing cartilage template (endochondral ossification). Endochondral ossification actually involves the formation of the three skeletal tissues that comprise the majority of the extant vertebrate skeleton: immature cartilage, mature cartilage and bone (Eames et al., 2003, 2004; Eames and Helms, 2004). Some cartilage remains throughout development at the growth plates and throughout life at articular surfaces (we term this immature cartilage). Most of the cartilage produced during endochondral ossification, however, undergoes a series of changes, termed maturation (thus the terms immature vs. mature cartilage). In most vertebrates, cartilage maturation involves cell hypertrophy, matrix mineralization, cell death, and matrix degradation (Leboy et al., 1988; Hatori et al., 1995; Takeda et al., 2001; Miura et al., 2008). Although exceptions exist (Thorogood, 1988; Hirasawa and Kuratani, 2015), endochondral ossification typically gives rise to the bones of the endoskeleton, such as the chondrocranium or limb skeleton, whereas intramembranous ossification produces the exoskeleton, such as lateral plates in teleosts or the calvarium (Smith and Hall, 1990).

Histologically, immature cartilage, mature cartilage, and bone are very similar, yet each also has some unique features (**Figure 1**). All three skeletal tissues are comprised of cells embedded in an ECM that is rich in collagens and proteoglycans (Hardingham, 1981; Eames et al., 2003, 2004; Eames and Helms, 2004; Gentili and Cancedda, 2009). Immature cartilage is formed by chondrocytes that deposit a network of loose collagen fibers and a rich substance of proteoglycans, whereas chondrocytes of mature cartilage alter the immature cartilage ECM by decreasing its proteoglycan sulfation and mineralizing it (Lohmander and Hjerpe, 1975; Buckwalter et al., 1987; Bayliss et al., 1999). The requirement of proteoglycan degradation for mature cartilage ECM mineralization is debated (Hirschman and Dziewiatkowski, 1966; Granda and Posner, 1971; Poole et al., 1982; Campo and Romano, 1986). Bone is formed by osteoblasts that produce an ECM of tightly wound and highly cross-linked collagen fibers, and bone ECM has lower levels of proteoglycans than cartilage (Gentili and Cancedda, 2009). As a result of these collagen and proteoglycan concentrations, these three skeletal tissues have overlapping and unique histological staining patterns. High concentrations of sulfated proteoglycans cause immature cartilage to stain with Alcian blue and Safranin O (by comparison, mature cartilage and bone bind these dyes with decreasing intensity, respectively). The tightly wound collagen fibers of bone stain with Direct red and Aniline blue (by comparison, loose collagen fibers of cartilage matrix bind these dyes with lower intensity; Villanueva et al., 1983; Hall, 1986; Eames and Helms, 2004; Eames et al., 2004, 2007). Alizarin red can stain mineralized tissues of mature cartilage and bone (Hogg, 1982; Kirsch et al., 1997; Eames and Helms, 2004; Eames et al., 2007).

Immature cartilage, mature cartilage, and bone have overlapping, but distinct, gene and protein expression profiles

	Similarities	Differences	
Functional	Connective tissues that enable: 1) Form and structural support 2) Movement 3) Protection of vital organs	Flexible structure; often serves as a load-bearing surface between bones	immature cartilage
		Often serves a transient function; in chondrichthyans: rigid structure, mineral reservoir	mature cartilage
		Rigid structure, mineral reservoir	bone
Embryonic	Share mesenchymal cell lineage (osteochondroprogenitor); initiate formation with mesenchymal condensation, then abundant ECM secretion	Persists through development; forms independently and during endochondral ossification	immature cartilage
		Often cell death and matrix degradation; forms during endochondral ossification	mature cartilage
		Matrix remodelling; forms during intramembranous and endochondral ossification	bone
Histological	Specialized cells embedded in abundant extracellular matrix (ECM) consisting of fibrous proteins and sulfated proteoglycans	Chondrocyte embedded in avascular, Col2/Aggrecan ECM	immature cartilage
		Hypertrophic chondrocyte embedded in avascular, mineralized Col2/Col10/Aggrecan ECM	mature cartilage
		Osteoblast/osteocyte embedded in vascular, Col1 ECM	bone

FIGURE 1 | Similarities and differences among immature cartilage, mature cartilage, and bone suggest that these three skeletal tissues share an evolutionary history.

(Figure 1). All these skeletal tissues express Collagen 11 and the proteoglycans Biglycan and Decorin (Li et al., 1998; Knudson and Knudson, 2001; Rees et al., 2001; Roughley, 2006). Immature cartilage expresses high levels of Collagens 2 and 9, as well as the proteoglycans Aggrecan, Fibromodulin, and Epiphycan, which distribute growth factors and provide swelling pressure due to water attraction (Yanagishita, 1993; Lefebvre et al., 1997; Lefebvre and de Crombrugghe, 1998; Watanabe et al., 1998; Liu et al., 2000). Mature cartilage has reduced expression of these same collagens and proteoglycans, while also expressing high levels of Collagen 10 (Orth et al., 1996; Eames et al., 2004; Talwar et al., 2006). In contrast to both types of cartilage, bone expresses high levels of Collagen 1 (Yasui et al., 1984; Kream et al., 1995). Interestingly (and central to the argument of this review), both mature cartilage and bone share expression of genes not expressed in immature cartilage, including *Sp7* (formerly called *Osterix*), *Matrix metallopeptidase 13* and *Indian hedgehog* (Vortkamp et al., 1996; Inada et al., 1999; Neuhold et al., 2001; Zaragoza et al., 2006; Abzhanov et al., 2007; Mak et al., 2008; Huycke et al., 2012; Nishimura et al., 2012; Weng and Su, 2013). In fact, very few genes expressed in bone are not expressed in mature cartilage, and this list of genes decreases further when comparisons among mature cartilage and bone are carried out in

actinopterygians (Eames et al., 2012). Multiple genes associated with matrix mineralization are expressed in both mature cartilage and bone, such as *Alkaline phosphatase, liver/bone/kidney* (*Alpl*, formerly called *Tissue-nonspecific alkaline phosphatase*), *Secreted phosphoprotein 1* (*Spp1*, formerly called *Osteopontin* or *Bone sialoprotein*), *Secreted protein, acidic, cysteine-rich* (*Sparc*, formerly called *Osteonectin*), and *Bone gamma-carboxyglutamate protein* (*Bglap*, formerly called *Osteocalcin*; Termine et al., 1981; Pacifici et al., 1990; Chen et al., 1991; Bonucci et al., 1992; McKee et al., 1992; Mundlos et al., 1992; Nakase et al., 1994; Roach, 1999; Sasaki et al., 2000).

Currently, the evolutionary relationship among skeletal tissues is unclear, but the similarities highlighted above suggest that immature cartilage, mature cartilage, and bone share an evolutionary history. From a molecular genetic perspective, these observations lead to the hypothesis that the GRNs governing the formation of these three skeletal tissues (in particular, the differentiation of three skeletal cell types) also share an evolutionary history. Indeed, the many varieties of skeletal tissues intermediate between cartilage and bone observed in extant and fossil vertebrates may owe their existence to this shared history (Benjamin, 1990; Benjamin and Ralphs, 1991; Benjamin et al., 1992; Mizoguchi et al., 1997; Hall, 2005; Witten

et al., 2010). In this review, we explore this hypothesis using traditional evolutionary and modern molecular genetic studies. We are not focussing on the exact anatomical location of a tissue, given that once the GRN regulating formation of that skeletal tissue is established in the genome, any cell in the body can co-opt its expression. Traditional studies have provided insight into the evolutionary relationship among skeletal tissues, since they demonstrate that immature cartilage originated first during phylogeny (Mallatt and Chen, 2003; Rychel et al., 2006). Interestingly, modern molecular genetic studies reveal that two GRNs dictate the formation of these three skeletal tissues (Bi et al., 1999; Inada et al., 1999; Eames et al., 2004; Hattori et al., 2010; Leung et al., 2011), and also that the GRN underlying cartilage formation is dominant to that of bone (Eames et al., 2004; Zhou et al., 2006). We expand upon this finding using an argument based on the relative parsimony of gradualism versus saltationism to hypothesize that bone evolved from a cartilage maturation program. In closing, we discuss how comparative transcriptomics will enhance dramatically our ability to test hypotheses on the evolution of the GRNs underlying cartilage and bone formation.

GRN Underlying Immature Cartilage Formation Evolved First

Traditional studies, such as the fossil record, comparative anatomy, and embryology, demonstrate that the first skeletal tissue to evolve was immature cartilage (**Figure 2**). The fossil record reveals a great diversity of mineralized tissues about 500 million years ago (Mya; Janvier, 1996, 2015; Donoghue and Sansom, 2002; Donoghue et al., 2006), suggesting that GRNs of skeletal histogenesis were undergoing an adaptive radiation. So which skeletal tissue appeared first in the fossil record? This question is complicated by the facts that currently discovered fossils may represent a biased fraction of ancestral tissues, and that non-mineralized, lightly mineralized, or transiently mineralized tissues likely are not preserved well in the fossil record. Despite these limitations, however, the oldest skeletal tissue in the fossil record is unmineralized cartilage in the chordate fossil *Haikouella* from 530 Mya (**Figure 2A**; Mallatt and Chen, 2003). Many specimens preserving soft tissues of this incredibly important fossil have been found, but they appear to be represented only in a small region of the Yunnan province in China (Chen et al., 1999), reflecting potential bias in the fossil record.

Bone and mature cartilage appeared much later than immature cartilage in the fossil record (**Figure 2A**). Conodonts, a group of agnathans (jawless vertebrate fish), are the earliest (~515 Mya) known fossils with a mineralized skeleton, characterized by pharyngeal tooth-like elements comprised of tissues that were bone-like, enamel-like, and mineralized cartilage-like (Sansom et al., 1992). However, subsequent analyses of conodont fossils refuted the conclusion that bone or mineralized cartilage was present in these primitive jawless fish, instead attributing the first appearance of bone in the fossil record to the exoskeleton of pteraspidomorphi (~480 Mya), a group of armored agnathans (Janvier, 1996; Donoghue, 1998; Donoghue et al., 2006).

Interestingly, some pteraspidomorph species (e.g., eriptychiids and arandaspidids) and other, primitive fossil fish show traces of both mineralized cartilage and bone in their endoskeleton (Janvier, 1996, 1997; Zhang et al., 2009). Also, fossils of the ancestral vertebrate *Palaeospndylus gunni* (~385 Mya) reveal an entire adult skeleton comprised of hypertrophic, mineralized cartilage, while bone is completely absent (Johanson et al., 2010). Despite these findings, the current fossil record generally suggests that bone preceded mineralized cartilage (Smith and Hall, 1990; Janvier, 1997; Donoghue et al., 2006), although the molecular genetic and embryological arguments of this review call into question the accuracy of this conclusion. What is clear from the fossil record is that unmineralized cartilage was the first skeletal tissue to appear leading to the evolution of vertebrates (Northcutt and Gans, 1983; Smith and Hall, 1990).

Comparative anatomy also supports the notion that immature cartilage was the first skeletal tissue to evolve, because immature cartilage is distributed in a broader range of taxonomic lineages than mature cartilage or bone (**Figure 2B**). Immature cartilage appears in both vertebrate and non-vertebrate species, whereas mature cartilage and bone are shared, derived traits of vertebrates only (Cole and Hall, 2004, 2009; Rychel et al., 2006). In a seminal study by Cole and Hall (2004), cartilage was demonstrated in a variety of taxonomically distinct invertebrates, such as polychaetes, arthropods, and molluscs. Reflecting the different evolutionary histories of immature and mature cartilage, cartilage in any invertebrate lineage, and also in extant agnathans, is unmineralized (Cole and Hall, 2004; Hall, 2005). The finding that lamprey cartilage can mineralize *in vitro* suggests that early agnathans may have possessed mineralized cartilage and these mineralization programs were repressed in cyclostomes (Langille and Hall, 1993).

The taxonomic distribution of cartilage suggests that the ancestor of vertebrates, cephalochordates, and hemichordates had an ability to make immature cartilage (**Figure 2B**). In fact, the deuterostome ancestor was proposed to be a benthic worm with cartilaginous gill slits (Rychel et al., 2006). Homology between invertebrate and vertebrate cartilages is supported by biochemical and histological analyses, which demonstrate high amounts of fibrous proteins and mucopolysaccharides (Cole and Hall, 2004; Cole, 2011). In fact, recent studies have shown that the cirri in amphioxus share many histological and molecular features with vertebrate immature cartilage (Kaneto and Wada, 2011; Jandzik et al., 2015). However, homology between deuterostome and protostome cartilage is still uncertain and must be confirmed by modern molecular analyses, including examination of gene expression patterns, GRN architectures, and GRN regulation. The ECM of hemichordate skeletal tissues may show features of both cartilage and bone (Cole and Hall, 2004), supporting the notion that these two tissues share an evolutionary history. Mineralized cartilage and bone, however, are only found in extant gnathostomes (**Figure 2**). These comparative anatomy analyses suggest that immature cartilage evolved before mature cartilage and bone.

Final support for the idea that cartilage arose earlier in evolution than mature cartilage and bone comes from comparative embryology. While the Biogenetic Law of Ernst

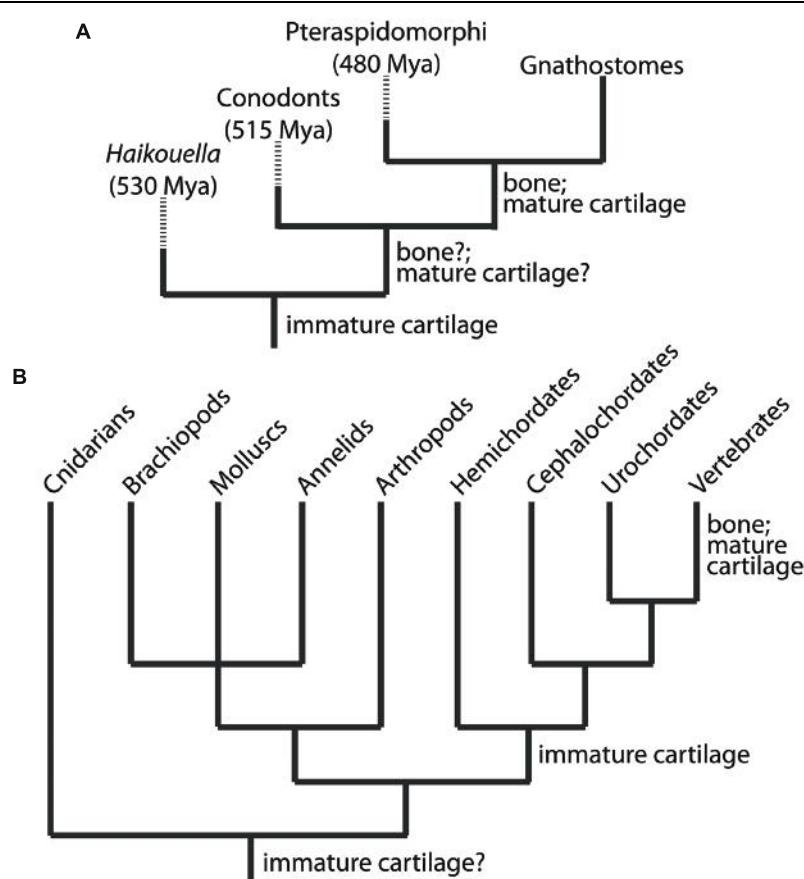


FIGURE 2 | Clues to the evolutionary relationship between the chondrocyte and osteoblast emerge from analyses of the fossil record and comparative anatomy. **(A)** Appearance of immature cartilage, mature cartilage, and bone from available fossil record. These data indicate clearly that immature cartilage appeared first, then mature cartilage and bone. By extension, the chondrocyte preceded the osteoblast during evolution. **(B)** Extant taxa with at least one species containing cartilage or cartilage-like tissues, which are non-mineralized outside of vertebrates. These data suggest that a GRN driving differentiation of an immature chondrocyte evolved first, and then became established in the genome of chordates (along with the notochord, a cartilage-like tissue). Subsequently, this GRN was modified by another GRN that drove differentiation of a mature chondrocyte (and osteoblast) within vertebrates. Branch lengths in trees are arbitrary; dashed lines indicate extinct taxa.

Haeckel definitely has its theoretical problems (Haeckel, 1866), a general correlation (recapitulation) between the timing of events during ontogeny with events during phylogeny is undeniable. Indeed, many early evolutionary biologists assumed this to be true (Gould, 2002). In this context, it is interesting to note that immature cartilage is the first skeletal tissue to undergo histogenesis during embryonic development, while cartilage maturation and bone formation are later events. The relative timing of cartilage maturation to bone formation, on the other hand, appears to vary among vertebrate taxa (Mori-Akiyama et al., 2003; Eames et al., 2004, 2012; Moriishi et al., 2005). While such relationships between the timing of developmental events have been argued to reflect simply the increasing complexity of ontogeny during phylogeny (Wallace, 1997), we believe that this issue, which has been debated for 100s of years, remains unresolved.

To sum up traditional studies of the fossil record, comparative anatomy, and embryology, the ability to make immature cartilage predates the ability to make mature cartilage or bone during

evolution. Therefore, from a molecular genetic perspective, the GRN governing chondrocyte differentiation clearly appeared prior to that of the osteoblast. However, traditional approaches are still unclear whether mature cartilage or bone appeared next during evolution. With hopes that modern molecular and embryological analyses can shed light into the evolutionary origins of the vertebrate skeleton, we next discuss how the GRNs underlying the formation of immature cartilage, mature cartilage, and bone are organized.

Sox9 GRN is Dominant to the Runx2 GRN

Skeletal histogenesis is governed by complex sets of genes, largely controlled by central transcription factors that are responsible for determining cell fate decisions (Eames et al., 2003, 2004; Kronenberg, 2003; Karsenty et al., 2009). Molecular genetic experiments demonstrate that the transcription factors Sox9 and Runx2 are the “master regulatory genes” of skeletal histogenesis.

Sox9 and Runx2 expression patterns during mesenchymal condensation predict whether osteochondroprogenitor cells differentiate into immature cartilage, mature cartilage, or bone (Eames and Helms, 2004; Eames et al., 2004). Loss of Sox9 function abrogated immature and mature cartilage formation (Bi et al., 1999; Mori-Akiyama et al., 2003), whereas Runx2 loss of function blocked mature cartilage and bone formation (Hoshi et al., 1999; Inada et al., 1999; Kim et al., 1999; Enomoto et al., 2000). In gain-of-function experiments, Sox9 mis-expression induced ectopic cartilage formation, whereas Runx2 mis-expression induced ectopic mature cartilage and bone formation (Eames et al., 2004). These and other experiments show clearly that a Sox9 GRN regulates immature cartilage formation, a Runx2 GRN drives bone formation, and a combination of Sox9 and Runx2 GRNs produce mature cartilage (**Figure 3**). We emphasize the relevance of these transcription factors to the evolution of GRNs underlying skeletal histogenesis, since conserved, core components of GRNs (i.e., kernels) are often transcription factors (Levine and Davidson, 2005; Davidson and Erwin, 2006).

Expression studies of skeletal tissues in a range of organisms suggest an ancestral interaction between Sox and Runx GRNs. *Runx2*, along with its related family members, *Runx1* and *3*, derive from gnathostome duplications of an ancestral *Runx*, while agnathan *Runx* genes may have undergone an independent duplication (Meulemans and Bronner-Fraser, 2007; Hecht et al., 2008; Cattell et al., 2011; Kaneto and Wada, 2011; Nah et al., 2014). *Sox9*, along with its related family members, *Sox8* and *10*,

derive from duplications to the ancestral *SoxE*, while agnathan *SoxE* genes may have undergone an independent duplication (Meulemans and Bronner-Fraser, 2007; Ohtani et al., 2008; Yu et al., 2008; Cattell et al., 2011; Uy et al., 2012; Jandzik et al., 2015). *Runx* and *SoxE* orthologs are expressed in cartilage of amphioxus, lamprey, and hagfish, suggesting that the gene ancestral to Runx2 primitively functioned with the gene ancestral to Sox9 in early cartilage formation (Hecht et al., 2008; Wada, 2010; Kaneto and Wada, 2011). Notably, these animals do not have bone, and they do not mineralize their skeletons. Interestingly, the amphioxus cirral skeleton shows features of both cartilage and bone, suggesting that this ancient skeleton might have diverged to form cellular cartilage and bone of vertebrates (Kaneto and Wada, 2011). We argue that evaluating the interactions between Sox9 and Runx2 GRNs leads to a novel hypothesis for the evolution of bone.

Many studies in mammals and chick demonstrate that the Sox9 GRN is at least partially dominant to the Runx2 GRN. First, co-expression of Sox9 and Runx2 typically causes cartilage formation, not bone (Eames and Helms, 2004; Eames et al., 2004). Second, ectopic expression of Sox9 in Runx2-expressing cells of developing bone (achieved either normally during secondary cartilage formation or experimentally using Sox9 mis-expression) diverts the cells to make cartilage, whereas ectopic Runx2 expression in Sox9-expressing cells of developing cartilage does not divert them to make bone (Eames et al., 2004). Third, Sox9 expression needs to be down-regulated in order for the full

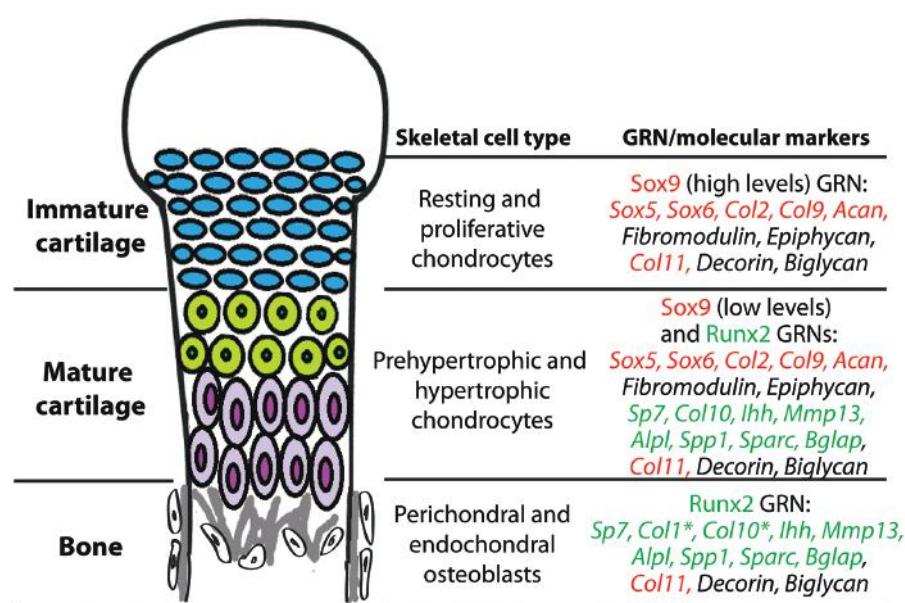


FIGURE 3 | During endochondral ossification, immature cartilage, mature cartilage, and bone differentiate under the control of Sox9 and Runx2 GRNs.

Chondrocytes of immature cartilage, termed resting and proliferative chondrocytes during endochondral ossification, express high levels of genes in the Sox9 GRN. Genes known to be under direct transcriptional control of Sox9 or Runx2 are highlighted in red or green text, respectively. Chondrocytes of mature cartilage, termed prehypertrophic and hypertrophic chondrocytes during endochondral ossification, express low levels of genes in the Sox9 GRN and also genes in the Runx2 GRN. Osteoblasts in perichondral and endochondral bone during endochondral ossification express genes in the Runx2 GRN. **Col1* is one of the only genes expressed in osteoblasts that is not expressed in mature chondrocytes; *Col10* expression in osteoblasts is high only in some vertebrates. *Col11*, *Decorin*, and *Biglycan* are expressed in all three of these skeletal cell types. Similar gene expression patterns are seen in immature cartilage, mature cartilage, and bone developing in the articular surface (not shown).

Runx2-dependent cartilage maturation program to be expressed (Akiyama et al., 2002; Eames et al., 2004). Fourth, Sox9 over-expression can inhibit Runx2 expression (Eames et al., 2004). Finally, and most conclusively, Sox9 directly binds to Runx2, inhibits its transcriptional activity, and increases ubiquitin-mediated degradation of Runx2 (Zhou et al., 2006; Cheng and Genever, 2010).

Given evidence that the Sox9 GRN can dominate the Runx2 GRN, the formation of mature cartilage during endochondral ossification, which requires both Sox9 and Runx2, must be regulated exquisitely (**Figure 3**). During early stages, both Sox9 and Runx2 are co-expressed in mesenchymal condensations (Akiyama et al., 2002; Eames and Helms, 2004; Eames et al., 2004; Zhou et al., 2006), so Sox9 must exert a dominant inhibitory effect over Runx2 in order to produce immature cartilage. Later, Sox9 is down-regulated and Runx2 activity increases, triggering cartilage maturation (Eames et al., 2004; Yoshida et al., 2004; Hattori et al., 2010). In fact, Sox9 down-regulation is a crucial step for mature cartilage formation (Hattori et al., 2010). Despite this down-regulation, a role for Sox9 in very late stages of cartilage maturation also has been revealed (Ikegami et al., 2011; Dy et al., 2012). One study even suggests that Runx2 can inhibit Sox9 activity (Cheng and Genever, 2010), illustrating that complex feedback mechanisms are in place to achieve the appropriate relative levels of Sox9 and Runx2 activity. In summary, the preponderance of published literature on molecular genetics demonstrates that Sox9 has dominant effects over Runx2, and we extend this conclusion to generate a new hypothesis on the evolution of bone.

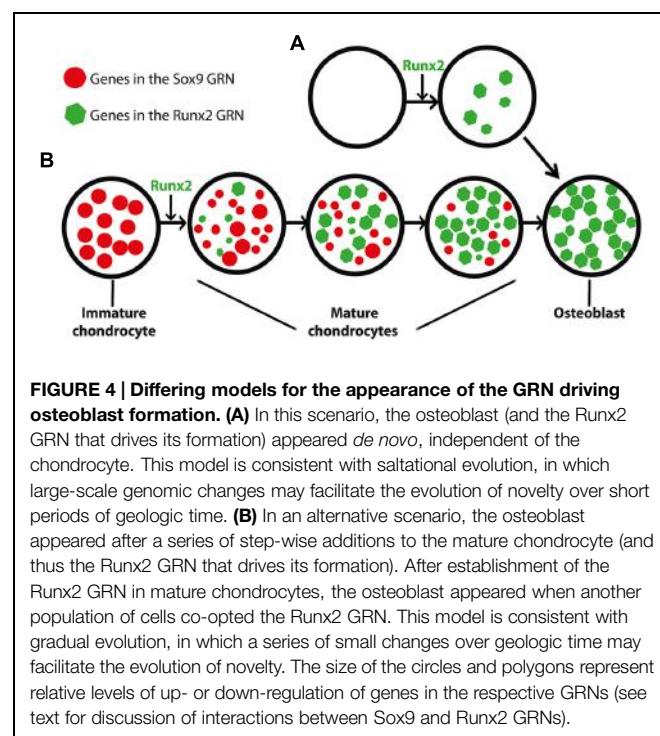
Bone Evolved from Mature Cartilage

Combining evidence from traditional and modern studies, we hypothesize that the GRN underlying bone formation evolved from a GRN underlying mature cartilage formation (**Figure 4**). Functional, histological, embryological, and molecular similarities among immature cartilage, mature cartilage, and bone suggest that these tissues may share an evolutionary history (**Figure 1**). The fossil record, comparative anatomy, and embryology demonstrate that immature cartilage evolved first (**Figure 2**). When combined with molecular genetic data (**Figure 3**), this means that the first evolved skeletal GRN was dominated by the gene ancestral to Sox9, driving immature cartilage formation. This GRN likely involved genes ancestral to Runx2 in early phylogenetic (and ontogenetic) stages. In gnathostomes, a Runx2 GRN drives formation of both mature cartilage and bone (**Figure 3**), but how did this novel GRN evolve to produce these novel skeletal tissues?

We propose that immature cartilage provided a structural and molecular “buffer” for the gradual development of this novel, Runx2 GRN. The structural buffering effect refers to the fact that immature cartilage already had a functional role as a skeletal tissue, allowing more freedom for the evolving Runx2 GRN to develop new functions that simply modify a pre-existing skeletal tissue in a gradual, step-wise fashion. The molecular buffering effect refers to the partial dominance of the Sox9 GRN, which

might have shielded to some extent the evolving Runx2 GRN from natural selection. This concept recalls the principle of “weak linkage,” which contributes to evolvability by reducing the cost of generating variation (Kirschner and Gerhart, 1998; Gerhart and Kirschner, 2007).

We argue that these putative buffering effects provide a more parsimonious account for the gradual evolution of bone from mature cartilage than the alternative, which depends upon *de novo* establishment of bone in a more saltationist fashion (**Figure 4**). If bone had evolved before mature cartilage, then the Runx2 GRN would have been under much stronger natural selection than if it had been buffered by immature cartilage. Arguments that bone evolved from dentine suffer from the same limitations: how did dentine and its GRN appear? A new GRN appearing simultaneously with a completely new skeletal tissue, while possible, seems a less likely evolutionary scenario than the gradual establishment of the Runx2 GRN during evolution of mature cartilage. Assembling a GRN driving bone formation *de novo* appears to depend upon saltationist genetic mechanisms, such as large-scale genomic changes or small genetic effects acting early in development. Regarding the latter possibility, chondrocytes and osteoblasts are known to share a relatively late embryonic progenitor (Day et al., 2005). Therefore, the former, “macromutational” saltationist mechanism, favored by Goldschmidt (Goldschmidt, 1940), would have to have operated in the *de novo* appearance of the osteoblast. Even saltationists granted that gradualism is the more common evolutionary mechanism (Gould, 2002). Therefore, based on the relative parsimony and abundance of gradualism versus saltationism, we favor a model in which the Runx2 GRN evolved within immature cartilage to produce mature cartilage, and then a



different mesenchymal (non-chondrogenic) cell population co-opted this GRN, producing the world's first example of bone formation (**Figure 4B**).

The hypothesis that bone evolved from mature cartilage also is consistent with a variety of other observations on skeletal tissues (Fisher and Franz-Odendaal, 2012). During evolution, the features of mature cartilage seen in various vertebrate taxa did not appear at the same time (Hall, 1975; Smith and Hall, 1990). Hypertrophy and mineralization occurred first, followed by cartilage matrix degradation, replacement by fat and endochondral bone deposition, and finally, invasion by the vasculature (in tetrapods). These findings suggest that cartilage maturation is a highly evolvable process. Also, the progression from immature cartilage to mature cartilage to bone during evolution is mimicked during endochondral ossification. Recently, cell lineage analyses suggest that some cells that express immature cartilage genes go on to express mature cartilage genes, and finally they express bone genes, effectively transitioning from an immature chondrocyte to a mature chondrocyte to an osteoblast (Hammond and Schulte-Merker, 2009; Zhou et al., 2014; Park et al., 2015). Finally, gene expression patterns appear to overlap much more when comparing mature cartilage to bone in actinopterygians, such as teleosts, than in sarcopterygians, such as tetrapods (Eames et al., 2012). This may reflect differential retention of molecular signatures of the evolutionary history between mature cartilage and bone in earlier diverging versus later diverging vertebrates.

Comparative Transcriptomics: A Novel Approach to Solve Evo-Devo Issues

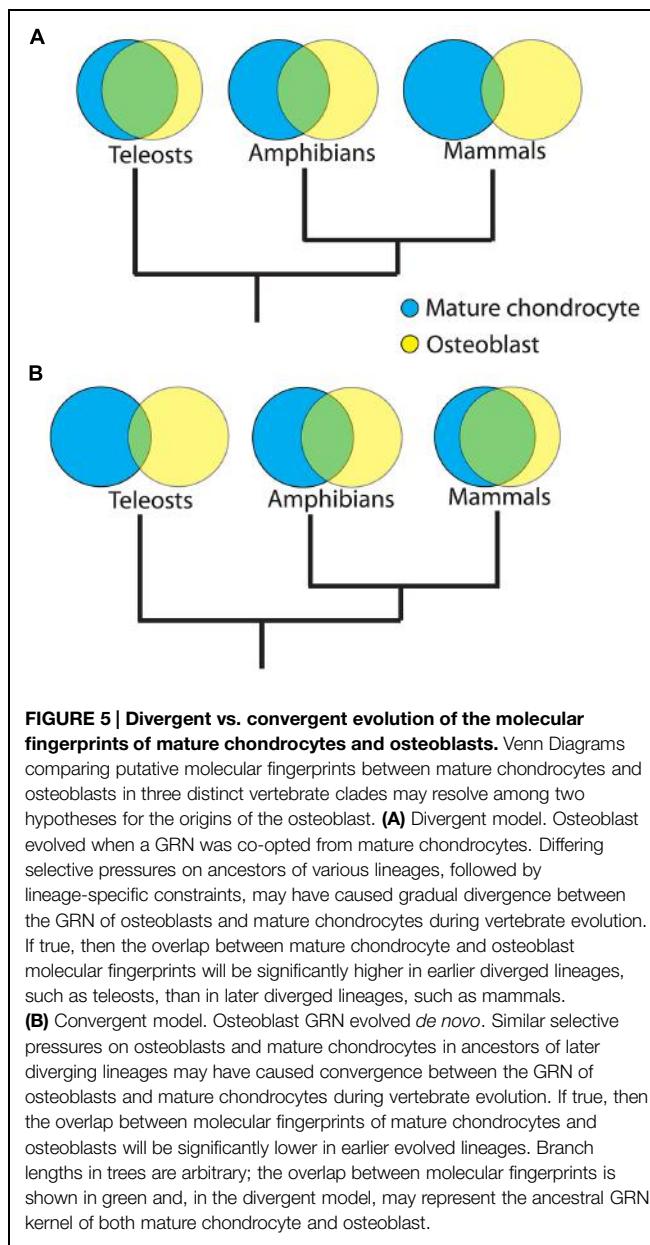
Identification of homologous tissue types among different taxonomic lineages using histology and cell morphology has enabled evolutionary studies of histogenesis, but modern molecular techniques will expand dramatically this field. Traditionally, comparative anatomy established homologies at the levels of organs, tissues, and cells. Homology among cartilage-like tissues can be relatively clear for closely related species, but can prove more difficult when comparing distant clades, where clade-specific differences can obscure homology. For example, histological features, such as cellularity of a tissue, may confuse homology designation; cartilage is cellular in vertebrates, but is acellular in hemichordates (Smith et al., 2003; Cole and Hall, 2004; Rychel et al., 2006). In addition, three types of agnathan cartilage have been distinguished by histology: hard cartilage, soft cartilage, and mucocartilage (Zhang and Cohn, 2006; Zhang et al., 2009; Cattell et al., 2011). Which of these would be homologous to hyaline cartilage of gnathostomes, or are they all? Modern evolutionary thinking overlooks such superficial histological differences, emphasizing instead the importance of tracking changes to the underlying molecular genetic factors during trait evolution.

Evolutionary studies of skeletal cells will benefit from transcriptomic techniques, such as RNAseq, that enable characterization of their molecular fingerprints, which are the sets of genes expressed in a homogenous population of

cells (Arendt, 2003). Comparing the molecular fingerprint of distinct cell types has yielded insight into evolutionary relationships among remote animal clades (Arendt, 2005, 2008; Eames et al., 2012). A few technologies can generate molecular fingerprints, but of these, RNAseq currently produces the most robust, unbiased results (Necsulea and Kaessmann, 2014). Some advantages of RNA-seq include a higher dynamic range, allowing the detection of transcripts that are expressed at very high or low levels, and the ability to detect novel genes and alternative splice variants in samples from any animal (Wang et al., 2009). Important for evolutionary studies, then, RNAseq allows for an accurate comparison of molecular fingerprints in both closely and distantly related species (Necsulea and Kaessmann, 2014; Pantalacci and Semon, 2015).

Tracking gene expression patterns that underlie a homologous trait through phylogeny provides unparalleled insight into molecular mechanisms of evolution. In fact, comparative transcriptomics might reveal that two tissues are homologous (so-called "deep homology"; Shubin et al., 2009), despite superficial histological or cellular differences. For example, the presence of immature cartilage in a variety of invertebrate taxa raises the possibility of a tissue with deep homology to cartilage present in the ancestor to all metazoans (**Figure 2B**). Also, identifying invertebrate tissues that express "bone genes" may reveal deep homology of these cells to osteoblasts, potentially facilitating the *de novo* appearance of the Runx2 GRN underlying bone formation. Genes in the vertebrate *Sparc* family play a role in skeletal matrix mineralization *in vitro* (Termine et al., 1981; Pataquiva-Mateus et al., 2012). Although similar *in vivo* roles for *Sparc* genes have not been demonstrated clearly (Roach, 1994; Gilmour et al., 1998; Rotllant et al., 2008), comparative genomics reveal a clear correlation between some *Sparc* genes and bone formation (Kawasaki and Weiss, 2006; Martinek et al., 2007; Koehler et al., 2009; Bertrand et al., 2013; Venkatesh et al., 2014). Interestingly, *Sparc* genes are expressed in amphioxus, which do not have bone nor mineralize their tissues (Bertrand et al., 2013). If Runx2 co-opted regulation of these genes during the *de novo* appearance of the osteoblast, then *Sparc*-expressing cells in amphioxus may have deep homology to osteoblasts.

Comparative transcriptomics can be used to evaluate quantitatively important features of GRN evolution, including constraint and adaptation. Although Gould recently revived the formalist pleas of Galton, Whitman, and others for constraint to have a positive role during evolution (Gould, 2002), constraint commonly is considered a restriction or limitation on the evolutionary process (Arnold, 1992). Evidence of constraint can be seen when transcriptomes are highly conserved among various tissues or clades, presumably due to genomic, developmental, or structural limitations. In addition to these constraints, a GRN under stabilizing selection would not vary much with respect to the genes expressed and their levels of expression, thus giving a transcriptomic signal of constraint. In fact, the architecture of GRN kernels, which usually consist of transcription factors and other regulatory genes, can remain highly conserved for a long period of time (Levine and Davidson, 2005; Davidson and Erwin, 2006). In contrast, adaptation commonly is considered positive for change during evolution (Gould, 2002; Stayton, 2008; Losos,



2011). Evidence of adaptation can be seen when transcriptomes differ widely among various tissues or clades, presumably in response to tissue- or clade-specific selective pressures. A GRN under negative or positive selection would vary a lot in the genes expressed and their levels of expression.

Comparative transcriptomics has unraveled the complexity of several important developmental and evolutionary processes in both invertebrate (Levin et al., 2012; McKenzie et al., 2014) and vertebrate organisms (Chan et al., 2009; Brawand et al., 2011). A major challenge in evolutionary biology is to explain the appearance of novel traits and the GRNs underlying their formation. Two different models have been proposed, with only one currently receiving much experimental support. In the first model, a GRN driving a novel trait also evolved *de novo*

(Figure 4A). For example, orphan genes, or genes without clear family members, might be important drivers of evolutionary novelty. First described in the yeast genome (Dujon, 1996), they occur also in many taxa, including rodents, primates, and humans (Heinen et al., 2009; Toll-Riera et al., 2009a,b; Li et al., 2010). Orphan genes might have appeared *de novo* from non-coding sequences rather than from existing genes (Tautz and Domazet-Loso, 2011). Subsequent interactions that these orphan genes establish among other genes would create a novel GRN with the capability of driving formation of a novel trait. This “*de novo*” model has received little experimental support in metazoans, but currently serves as the basis for the hypothesis that bone (or dentine, if dentine appeared before bone during evolution) evolved before mature cartilage (Figure 4A). In molecular terms, the GRN driving formation of the osteoblast would have appeared *de novo*, presumably in a short evolutionary timeframe.

In the second model for appearance of evolutionary novelties, which is increasingly supported by the literature, a novel trait appears by co-opting a pre-existing GRN (Figure 4B; Fisher and Franz-Odendaal, 2012; Achim and Arendt, 2014). For example, comparative genomic studies on muscle cells, immune cells, and neurons suggested that these cell types evolved by co-opting pre-existing genetic systems (Achim and Arendt, 2014). In addition, the appearance of a novel embryonic cell lineage in vertebrates, the neural crest cell, has been argued to result from the co-option of pre-existing GRNs that were employed by cells in the neural tube, notochord, and pharynx in ancestral chordates (Baker and Bronner-Fraser, 1997; Donoghue and Sansom, 2002; Meulemans and Bronner-Fraser, 2005, 2007; McCauley and Bronner-Fraser, 2006; Zhang and Cohn, 2006). In fact, the neural crest-derived vertebrate cartilaginous head skeleton might have arisen after neural crest cells co-opted an ancestral chordate GRN that was used for cartilage formation in other parts of the body (Jandzik et al., 2015). Here, we use the same argument to support our idea that the osteoblast appeared when a non-chondrogenic mesenchymal cell co-opted expression of the mature cartilage Runx2 GRN.

Comparative Transcriptomics and Skeletal Tissue Evolution

How extensive is our understanding of the GRNs driving cartilage and bone formation? As outlined above, Sox9 and Runx2 GRNs are critical in a variety of vertebrates, but is this the whole story? Few studies have analyzed the molecular fingerprint of the chondrocyte and osteoblast using unbiased transcriptomics, but such experiments may identify unknown GRN's driving formation of these cell types. The chondrocyte molecular fingerprint was estimated by compiling data from the literature and summarizing their interactions into a GRN (Cole, 2011). Recently, transcriptomics on Sox9 and Runx2 loss-of-function skeletal cells *in vitro* have shed light on Sox9 and Runx2 GRNs that are relevant to chondrocyte and osteoblast differentiation (Oh et al., 2014; Wu et al., 2014). A promising future direction is to use transcriptomics to define these GRNs *in vivo* using Sox9 and Runx2 loss-of-function animals. Comparative

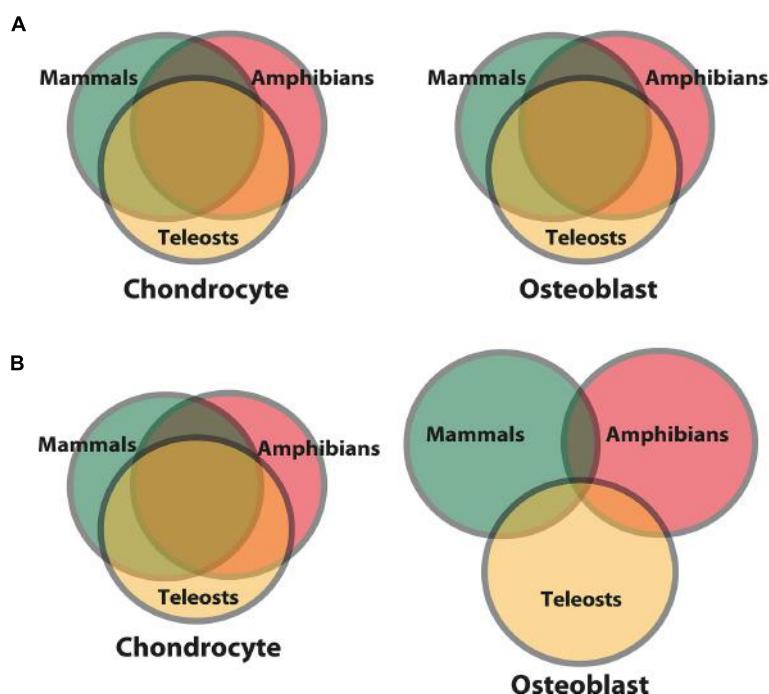


FIGURE 6 | Differing models for levels of constraint and adaptation among skeletal cells of different vertebrate lineages. Venn diagrams comparing putative molecular fingerprints of chondrocytes and osteoblasts from three vertebrate clades. The ancestral chondrocyte and osteoblast GRN kernels are represented in the overlap of the circles. More overlap represents more constraint/less adaptation among clades. **(A)** The first scenario predicts that the molecular fingerprints of the chondrocyte and osteoblast (and thus the GRNs governing their formation) are constrained to equal extents among vertebrates (Fisher and Franz-Odendaal, 2012; Vieira et al., 2013). **(B)** The second scenario predicts that the chondrocyte molecular fingerprint is more constrained among vertebrate clades, while the osteoblast molecular fingerprint shows more signs of clade-specific adaptations (Eames et al., 2012). In general, this latter scenario posits that a cell type appearing later during animal phylogeny is more free to vary than a cell type appearing earlier, whose molecular fingerprint was fixed via stabilizing selection.

transcriptomics between vertebrae and gill arch skeletal elements of a teleost demonstrated a high degree of overlap in gene expression between these two tissues (Vieira et al., 2013), but the presence of multiple cell types, including chondrocytes and osteoblasts, in both samples confounds attribution of these data to a particular cell type. Therefore, more specific techniques should be used to isolate a pure population of cells *in vivo* in order to accurately reveal and compare the molecular fingerprints of different skeletal cell types (Figure 3).

Two related, fascinating questions remain for future research: how did the GRNs directing skeletal cell differentiation appear, and how did they evolve afterward? In this review, we argue that gradual establishment of the Runx2 GRN during evolution of the mature chondrocyte (subsequently co-opted by a non-chondrogenic mesenchymal cell to form bone) is more parsimonious than the *de novo* appearance of the Runx2 GRN in osteoblasts (Figure 4). Given the latter possibility, however, the tremendous gene expression similarities between mature cartilage and bone in tetrapods also may reflect co-option of the Runx2 GRN by the mature chondrocyte after it was established in the osteoblast. These possibilities predict divergent vs. convergent evolution, respectively, of the Runx2 GRN in mature chondrocytes after the appearance of the osteoblast. Therefore, we propose an examination of skeletal cell molecular fingerprints in a variety of vertebrates to resolve this issue.

Our divergent model predicts that the overlap between mature chondrocyte and osteoblast molecular fingerprints will decrease in more recently evolved organisms (Figure 5A). For example, molecular fingerprints of mature chondrocytes and osteoblasts would overlap more in earlier diverged lineages of vertebrates, such as teleosts, than in later evolved lineages, such as amphibians or mammals. On the other hand, the convergent model predicts the opposite result (Figure 5B).

But do skeletal cell molecular fingerprints evolve in clade-specific manners? A limited number of studies trying to answer this question suggest two competing ideas. On the one hand, molecular fingerprints of the chondrocyte and the osteoblast have been proposed to be highly constrained among various vertebrate clades (Figure 6A; Fisher and Franz-Odendaal, 2012; Vieira et al., 2013). On the other hand, gene expression comparisons between gar, zebrafish, chick, and mouse suggest that the chondrocyte molecular fingerprint is constrained among vertebrates, while the osteoblast molecular fingerprint varied, perhaps in response to clade-specific selective pressures (Figure 6B; Eames et al., 2012). Interestingly, generalizing these results puts forward the hypothesis that earlier-evolved cell types, in this case chondrocytes, might be more constrained in their gene expression than cell types that appeared later, such as osteoblasts, perhaps due to stabilizing selection over geologic timescales. Comparative transcriptomics can quantitate

constraint and adaptation, by measuring how transcript levels vary among samples from different taxonomic lineages.

In the future, comparative transcriptomics will elucidate the dynamics of skeletal cell type evolution, identifying lineage-specific changes in gene expression, providing quantitative measures of constraint and adaptation, and potentially establishing deep homology of skeletal cells with previously unappreciated cell types. Indeed, appropriate application of comparative transcriptomics has the potential to revolutionize understanding of the molecular mechanisms of trait evolution.

Summary

Given the role that fossilized bones played in devising early evolutionary theory, skeletal tissue evolution has fascinated scientists for centuries. In particular, the appearance of bone as an evolutionary novelty demands explanation, which modern molecular and embryological techniques address in ways never imagined by studies of the fossil record alone. Here, we focus on the three main skeletal tissues present in vertebrates (immature cartilage, mature cartilage, and bone), and use findings from both traditional and modern studies to argue that bone evolved from mature cartilage. Standing in contrast to the available fossil record, which suggests that bone appeared prior to mature cartilage, this hypothesis posits that a GRN driving traits such as matrix mineralization in mature cartilage was co-opted by non-chondrogenic mesenchymal cells to produce bone. Alternatively, the GRN driving bone formation may

have evolved first and subsequently was co-opted by mature cartilage, but we use an argument based on parsimony that this scenario would be more complicated to achieve. Comparing the molecular fingerprints of skeletal tissues in agnathans and sister chordate species with those in vertebrates might resolve among these possibilities. In addition to comparative transcriptomics revealing the origins of evolutionary novelties, tracking molecular fingerprints of skeletal cells in various vertebrate lineages can identify quantitative measures of constraint and adaptation within the GRNs that govern the formation of skeletal tissues. Therefore, we strongly believe that this novel approach may revolutionize understanding of the evolution of cartilage and bone and more generally provide a modern paradigm for molecular genetic changes during the evolutionary process.

Acknowledgments

PGP was supported by scholarships from the Consejo Nacional de Ciencia y Tecnología (CONACYT Mexico), Saskatchewan Innovation and Opportunity Graduate Scholarships (SIOGS), and College of Graduate Studies and Research at the University of Saskatchewan. BFE was supported on this study by a Saskatchewan Health Research Foundation (SHRF) Establishment Grant, a Natural Sciences and Engineering Research Council (NSERC) Discovery Grant, and a Canadian Institutes of Health Research (CIHR) New Investigator Salary Award.

References

- Abzhanov, A., Rodda, S. J., McMahon, A. P., and Tabin, C. J. (2007). Regulation of skeletogenic differentiation in cranial dermal bone. *Development* 134, 3133–3144. doi: 10.1242/dev.002709
- Achim, K., and Arendt, D. (2014). Structural evolution of cell types by step-wise assembly of cellular modules. *Curr. Opin. Genet. Dev.* 27, 102–108. doi: 10.1016/j.gde.2014.05.001
- Akiyama, H., Chaboissier, M. C., Martin, J. F., Schedl, A., and de Crombrugghe, B. (2002). The transcription factor Sox9 has essential roles in successive steps of the chondrocyte differentiation pathway and is required for expression of Sox5 and Sox6. *Genes Dev.* 16, 2813–2828. doi: 10.1101/gad.1017802
- Arendt, D. (2003). Evolution of eyes and photoreceptor cell types. *Int. J. Dev. Biol.* 47, 563–571.
- Arendt, D. (2005). Genes and homology in nervous system evolution: comparing gene functions, expression patterns, and cell type molecular fingerprints. *Theory Biosci.* 124, 185–197. doi: 10.1007/BF02814483
- Arendt, D. (2008). The evolution of cell types in animals: emerging principles from molecular studies. *Nat. Rev. Genet.* 9, 868–882. doi: 10.1038/nrg2416
- Arnold, S. J. (1992). Constraints on phenotypic evolution. *Am. Nat.* 140(Suppl. 1), S85–S107. doi: 10.1086/285398
- Baker, C. V., and Bronner-Fraser, M. (1997). The origins of the neural crest. Part II: an evolutionary perspective. *Mech. Dev.* 69, 13–29. doi: 10.1016/S0925-4773(97)00129-9
- Bayliss, M. T., Osborne, D., Woodhouse, S., and Davidson, C. (1999). Sulfation of chondroitin sulfate in human articular cartilage. The effect of age, topographical position, and zone of cartilage on tissue composition. *J. Biol. Chem.* 274, 15892–15900. doi: 10.1074/jbc.274.22.15892
- Benjamin, M. (1990). The cranial cartilages of teleosts and their classification. *J. Anat.* 169, 153–172.
- Benjamin, M., and Ralphs, J. R. (1991). Extracellular matrix of connective tissues in the heads of teleosts. *J. Anat.* 179, 137–148.
- Benjamin, M., Ralphs, J. R., and Eberewariye, O. S. (1992). Cartilage and related tissues in the trunk and fins of teleosts. *J. Anat.* 181(Pt 1), 113–118.
- 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. Biol. Sci.* 280:20122963. doi: 10.1098/rspb.2012.2963
- Bi, W., Deng, J. M., Zhang, Z., Behringer, R. R., and de Crombrugghe, B. (1999). Sox9 is required for cartilage formation. *Nat. Genet.* 22, 85–89. doi: 10.1038/8792
- Blumer, M. J., Longato, S., Richter, E., Perez, M. T., Konakci, K. Z., and Fritsch, H. (2005). The role of cartilage canals in endochondral and perichondral bone formation: are there similarities between these two processes? *J. Anatomy* 206, 359–372. doi: 10.1111/j.1469-7580.2005.00404.x
- Bonucci, E., Silvestrini, G., and Bianco, P. (1992). Extracellular alkaline phosphatase activity in mineralizing matrices of cartilage and bone: ultrastructural localization using a cerium-based method. *Histochemistry* 97, 323–327. doi: 10.1007/BF00270033
- Brawand, D., Soumilon, M., Necsulea, A., Julien, P., Csardi, G., Harrigan, P., et al. (2011). The evolution of gene expression levels in mammalian organs. *Nature* 478, 343–348. doi: 10.1038/nature10532
- Buckwalter, J. A., Rosenberg, L. C., and Ungar, R. (1987). Changes in proteoglycan aggregates during cartilage mineralization. *Calcif. Tissue Int.* 41, 228–236. doi: 10.1007/BF02555243
- Campo, R. D., and Romano, J. E. (1986). Changes in cartilage proteoglycans associated with calcification. *Calcif. Tissue Int.* 39, 175–184. doi: 10.1007/BF02555115
- Cattell, M., Lai, S., Cerny, R., and Medeiros, D. M. (2011). A new mechanistic scenario for the origin and evolution of vertebrate cartilage. *PLoS ONE* 6:e22474. doi: 10.1371/journal.pone.0022474

- Chan, E. T., Quon, G. T., Chua, G., Babak, T., Trochesset, M., Zirngibl, R. A., et al. (2009). Conservation of core gene expression in vertebrate tissues. *J. Biol.* 8:33. doi: 10.1186/jbiol130
- Chen, J., Zhang, Q., McCulloch, C. A., and Sodek, J. (1991). Immunohistochemical localization of bone sialoprotein in foetal porcine bone tissues: comparisons with secreted phosphoprotein 1 (SPP-1, osteopontin) and SPARC (osteonectin). *Histochem. J.* 23, 281–289. doi: 10.1007/BF01045047
- Chen, J.-Y., Huang, D.-Y., and Li, C.-W. (1999). An early Cambrian craniate-like chordate. *Nature* 402, 518–522. doi: 10.1038/990080
- Cheng, A., and Genever, P. G. (2010). SOX9 determines RUNX2 transactivity by directing intracellular degradation. *J. Bone. Miner. Res.* 25, 2680–2689. doi: 10.1002/jbmr.174
- Cole, A. G. (2011). A review of diversity in the evolution and development of cartilage: the search for the origin of the chondrocyte. *Eur. Cell Mater.* 21, 122–129.
- Cole, A. G., and Hall, B. K. (2004). The nature and significance of invertebrate cartilages revisited: distribution and histology of cartilage and cartilage-like tissues within the Metazoa. *Zoology (Jena)* 107, 261–273. doi: 10.1016/j.zool.2004.05.001
- Cole, A. G., and Hall, B. K. (2009). Cartilage differentiation in cephalopod molluscs. *Zoology (Jena)* 112, 2–15. doi: 10.1016/j.zool.2008.01.003
- Daniel, J. F. (1934). *The Elasmobranch Fishes*. Berkeley, CA: University of California Press. doi: 10.5962/bhl.title.10180
- Datta, H. K., Ng, W. F., Walker, J. A., Tuck, S. P., and Varanasi, S. S. (2008). The cell biology of bone metabolism. *J. Clin. Pathol.* 61, 577–587. doi: 10.1136/jcp.2007.048868
- Davidson, E. H., and Erwin, D. H. (2006). Gene regulatory networks and the evolution of animal body plans. *Science* 311, 796–800. doi: 10.1126/science.1113832
- Day, T. F., Guo, X., Garrett-Beal, L., and Yang, Y. (2005). Wnt/beta-catenin signaling in mesenchymal progenitors controls osteoblast and chondrocyte differentiation during vertebrate skeletogenesis. *Dev. Cell* 8, 739–750. doi: 10.1016/j.devcel.2005.03.016
- Donoghue, P. C. (1998). Growth and patterning in the conodont skeleton. *Philos. Trans. R. Soc. Lond.* 363, 633–666. doi: 10.1098/rstb.1998.0231
- Donoghue, P. C., 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., 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
- Dujon, B. (1996). The yeast genome project: what did we learn? *Trends Gen.* 12, 263–270. doi: 10.1016/0168-9525(96)10027-5
- Dy, P., Wang, W., Bhattacharay, P., Wang, Q., Wang, L., Ballock, R. T., et al. (2012). Sox9 directs hypertrophic maturation and blocks osteoblast differentiation of growth plate chondrocytes. *Dev. Cell* 22, 597–609. doi: 10.1016/j.devcel.2011.12.024
- Eames, B. F., Allen, N., Young, J., Kaplan, A., Helms, J. A., and Schneider, R. A. (2007). Skeletogenesis in the swell shark *Cephaloscyllium ventriosum*. *J. Anat.* 210, 542–554. doi: 10.1111/j.1469-7580.2007.00723.x
- Eames, B. F., Amores, A., Yan, Y. L., and Postlethwait, J. H. (2012). Evolution of the osteoblast: skeletogenesis in gar and zebrafish. *BMC Evol. Biol.* 12:27. doi: 10.1186/1471-2148-12-27
- Eames, B. F., de la Fuente, L., and Helms, J. A. (2003). Molecular ontogeny of the skeleton. *Birth Defects Res. C Embryo Today* 69, 93–101. doi: 10.1002/bdrc.10016
- Eames, B. F., and Helms, J. A. (2004). Conserved molecular program regulating cranial and appendicular skeletogenesis. *Dev. Dyn.* 231, 4–13. doi: 10.1002/dvdy.20134
- Eames, B. F., Sharpe, P. T., and Helms, J. A. (2004). Hierarchy revealed in the specification of three skeletal fates by Sox9 and Runx2. *Dev. Biol.* 274, 188–200. doi: 10.1016/j.ydbio.2004.07.006
- Enomoto, H., Enomoto-Iwamoto, M., Iwamoto, M., Nomura, S., Himeno, M., Kitamura, Y., et al. (2000). Cbfa1 is a positive regulatory factor in chondrocyte maturation. *J. Biol. Chem.* 275, 8695–8702. doi: 10.1074/jbc.275.12.8695
- Fang, J., and Hall, B. K. (1997). Chondrogenic cell differentiation from membrane bone periosteum. *Anat. Embryol.* 196, 349–362. doi: 10.1007/s00429-0050104
- Fisher, S., and Franz-Ondelaaal, T. (2012). Evolution of the bone gene regulatory network. *Curr. Opin. Genet. Dev.* 22, 390–397. doi: 10.1016/j.gde.2012.04.007
- Gentili, C., and Cancedda, R. (2009). Cartilage and bone extracellular matrix. *Curr. Pharm. Des.* 15, 1334–1348. doi: 10.2174/1381612097846739
- Gerhart, J., and Kirschner, M. (2007). The theory of facilitated variation. *Proc. Natl. Acad. Sci. U.S.A.* 104(Suppl. 1), 8582–8589. doi: 10.1073/pnas.0701035104
- Gilmour, D. T., Lyon, G. J., Carlton, M. B., Sanes, J. R., Cunningham, J. M., Anderson, J. R., et al. (1998). Mice deficient for the secreted glycoprotein SPARC/osteonectin/BM40 develop normally but show severe age-onset cataract formation and disruption of the lens. *EMBO J.* 17, 1860–1870. doi: 10.1093/emboj/17.7.1860
- Goldschmidt, R. (1940). *The Material Basis for Evolution*. New Haven, CT: Yale University Press.
- Gould, S. J. (2002). *The Structure of Evolutionary Theory*. Cambridge, MA: Belknap of Harvard University Press.
- Granda, J. L., and Posner, A. S. (1971). Distribution of four hydrolases in the epiphyseal plate. *Clin. Orthop. Relat. Res.* 74, 269–272. doi: 10.1097/00003086-197101000-00032
- Gray, H., and Williams, P. L. (1989). *Gray's Anatomy*, 37th Edn. Edinburgh: C. Livingstone.
- Haeckel, E. (1866). *Generelle Morphologie der Organismen*. Berlin: G. Reimer.
- Hall, B. K. (1975). Evolutionary consequences of skeletal Differentiation. *Am. Zool.* 15, 329–350. doi: 10.1093/icb/15.2.329
- Hall, B. K. (1986). The role of movement and tissue interactions in the development and growth of bone and secondary cartilage in the clavicle of the embryonic chick. *J. Embryol. Exp. Morphol.* 93, 133–152.
- Hall, B. K. (2005). *Bones and Cartilage: Development and Evolutionary Skeletal Biology*. San Diego, CA: Elsevier Academic Press.
- Hall, B. K., and Miyake, T. (1995). Divide, accumulate, differentiate: cell condensation in skeletal development revisited. *Int. J. Dev. Biol.* 39, 881–893.
- Hall, B. K., and Miyake, T. (2000). All for one and one for all: condensations and the initiation of skeletal development. *Bioessays* 22, 138–147. doi: 10.1002/(SICI)1521-1878(200002)22:2<138::AID-BIES5>3.0.CO;2-4
- Hammond, C. L., and Schulte-Merker, S. (2009). Two populations of endochondral osteoblasts with differential sensitivity to Hedgehog signalling. *Development* 136, 3991–4000. doi: 10.1242/dev.042150
- Hardingham, T. (1981). Proteoglycans: their structure, interactions and molecular organization in cartilage. *Biochem. Soc. Trans.* 9, 489–497. doi: 10.1042/bst0090489
- Hatori, M., Klatte, K. J., Teixeira, C. C., and Shapiro, I. M. (1995). End labeling studies of fragmented DNA in the avian growth plate: evidence of apoptosis in terminally differentiated chondrocytes. *J. Bone. Miner. Res.* 10, 1960–1968. doi: 10.1002/jbm.5650101216
- Hattori, T., Muller, C., Gebhard, S., Bauer, E., Pausch, F., Schlund, B., et al. (2010). SOX9 is a major negative regulator of cartilage vascularization, bone marrow formation and endochondral ossification. *Development* 137, 901–911. doi: 10.1242/dev.045203
- Hecht, J., Stricker, S., Wiecha, U., Stiege, A., Panopoulou, G., Podsiadlowski, L., et al. (2008). Evolution of a core gene network for skeletogenesis in chordates. *PLoS Genet.* 4:e1000025. doi: 10.1371/journal.pgen.1000025
- Heinen, T. J., Staubach, F., Haming, D., and Tautz, D. (2009). Emergence of a new gene from an intergenic region. *Curr. Biol.* 19, 1527–1531. doi: 10.1016/j.cub.2009.07.049
- Hill, T. P., Spater, D., Taketo, M. M., Birchmeier, W., and Hartmann, C. (2005). Canonical Wnt/beta-catenin signaling prevents osteoblasts from differentiating into chondrocytes. *Dev. Cell* 8, 727–738. doi: 10.1016/j.devcel.2005.02.013
- Hirasawa, T., and Kuratani, S. (2015). Evolution of the vertebrate skeleton: morphology, embryology, and development. *Zool. Lett.* 1:2. doi: 10.1186/s40851-014-0007-7
- Hirschman, A., and Dziewiatkowski, D. D. (1966). Protein-polysaccharide loss during endochondral ossification: immunochemical evidence. *Science* 154, 393–395. doi: 10.1126/science.154.3747.393
- Hogg, D. A. (1982). Ossification of the laryngeal, tracheal and syringeal cartilages in the domestic fowl. *J. Anat.* 134, 57–71.
- Hoshi, K., Komori, T., and Ozawa, H. (1999). Morphological characterization of skeletal cells in Cbfa1-deficient mice. *Bone* 25, 639–651. doi: 10.1016/S8756-3282(99)00223-9

- Huycke, T. R., Eames, B. F., and Kimmel, C. B. (2012). Hedgehog-dependent proliferation drives modular growth during morphogenesis of a dermal bone. *Development* 139, 2371–2380. doi: 10.1242/dev.079806
- Ikegami, D., Akiyama, H., Suzuki, A., Nakamura, T., Nakano, T., Yoshikawa, H., et al. (2011). Sox9 sustains chondrocyte survival and hypertrophy in part through Ptk3ca-Akt pathways. *Development* 138, 1507–1519. doi: 10.1242/dev.057802
- Inada, M., Yasui, T., Nomura, S., Miyake, S., Deguchi, K., Himeno, M., et al. (1999). Maturational disturbance of chondrocytes in Cbfα1-deficient mice. *Dev. Dyn.* 214, 279–290. doi: 10.1002/(SICI)1097-0177(199904)214:4<279::AID-AJA1>3.0.CO;2-W
- Jandzik, D., Garnett, A. T., Square, T. A., Cattell, M. V., Yu, J. K., and Medeiros, D. M. (2015). Evolution of the new vertebrate head by co-option of an ancient chordate skeletal tissue. *Nature* 518, 534–537. doi: 10.1038/nature14000
- Janvier, P. (1996). *Early Vertebrates*. Oxford: Clarendon Press.
- Janvier, P. (1997). *Pteraspidomorphi-The Tree of Life Web Project*. Available at: <http://tolweb.org/Pteraspidomorphi/14834>
- Janvier, P. (2015). Facts and fancies about early fossil chordates and vertebrates. *Nature* 520, 483–489. doi: 10.1038/nature14437
- 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. Cell Dev. Biol.* 21, 414–423. doi: 10.1016/j.semcdb.2009.10.011
- Johnson, D. R. (1980). Formation of marrow cavity and ossification in mouse limb buds grown in vitro. *J. Embryol. Exp. Morphol.* 56, 301–307.
- Kaneto, S., and Wada, H. (2011). Regeneration of amphioxus oral cirri and its skeletal rods: implications for the origin of the vertebrate skeleton. *J. Exp. Zool. B Mol. Dev. Evol.* 316, 409–417. doi: 10.1002/jez.b.21411
- Karsenty, G., Kronenberg, H. M., and Settembre, C. (2009). Genetic control of bone formation. *Annu. Rev. Cell Dev. Biol.* 25, 629–648. doi: 10.1146/annurev.cellbio.042308.113308
- Kawasaki, K., Suzuki, T., and Weiss, K. M. (2004). Genetic basis for the evolution of vertebrate mineralized tissue. *Proc. Natl. Acad. Sci. U.S.A.* 101, 11356–11361. doi: 10.1073/pnas.0404279101
- Kawasaki, K., and Weiss, K. M. (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. doi: 10.1002/jez.b.21088
- Kemp, N. E., and Westrin, S. K. (1979). Ultrastructure of calcified cartilage in the endoskeletal tesserae of sharks. *J. Morphol.* 160, 75–109. doi: 10.1002/jmor.1051600106
- Kim, I. S., Otto, F., Zabel, B., and Mundlos, S. (1999). Regulation of chondrocyte differentiation by Cbfα1. *Mech. Dev.* 80, 159–170. doi: 10.1016/S0925-4773(98)00210-X
- Kirschner, T., Nah, H. D., Shapiro, I. M., and Pacifici, M. (1997). Regulated production of mineralization-competent matrix vesicles in hypertrophic chondrocytes. *J. Cell Biol.* 137, 1149–1160. doi: 10.1083/jcb.137.5.1149
- Kirschner, M., and Gerhart, J. (1998). Evolvability. *Proc. Natl. Acad. Sci. U.S.A.* 95, 8420–8427. doi: 10.1073/pnas.95.15.8420
- Knudson, C. B., and Knudson, W. (2001). Cartilage proteoglycans. *Semin. Cell Dev. Biol.* 12, 69–78. doi: 10.1006/scdb.2000.0243
- Koehler, A., Desser, S., Chang, B., MacDonald, J., Tepass, U., and Ringuette, M. (2009). Molecular evolution of SPARC: absence of the acidic module and expression in the endoderm of the starlet sea anemone, *Nematostella vectensis*. *Dev. Genes Evol.* 219, 509–521. doi: 10.1007/s00427-009-0313-9
- Kream, B. E., Harrison, J. R., Krebsbach, P. H., Bogdanovic, Z., Bedalov, A., Pavlin, D., et al. (1995). Regulation of type I collagen gene expression in bone. *Connect. Tissue Res.* 31, 261–264. doi: 10.3109/03008209509010819
- Kronenberg, H. M. (2003). Developmental regulation of the growth plate. *Nature* 423, 332–336. doi: 10.1038/nature01657
- Langille, R. M., and Hall, B. K. (1993). Calcification of Cartilage from the Lamprey *Petromyzon marinus* (L.) *in vitro*. *Acta Zool.* 74, 31–41. doi: 10.1111/j.1463-6395.1993.tb01218.x
- Leboy, P. S., Shapiro, I. M., Uschmann, B. D., Oshima, O., and Lin, D. (1988). Gene expression in mineralizing chick epiphyseal cartilage. *J. Biol. Chem.* 263, 8515–8520.
- Lefebvre, V., and de Crombrugghe, B. (1998). Toward understanding SOX9 function in chondrocyte differentiation. *Matrix Biol.* 16, 529–540. doi: 10.1016/S0945-053X(98)90065-8
- Lefebvre, V., Huang, W., Harley, V. R., Goodfellow, P. N., and de Crombrugghe, B. (1997). SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro α1(I) collagen gene. *Mol. Cell. Biol.* 17, 2336–2346.
- Leung, V. Y., Gao, B., Leung, K. K., Melhado, I. G., Wynn, S. L., Au, T. Y., et al. (2011). SOX9 governs differentiation stage-specific gene expression in growth plate chondrocytes via direct concomitant transactivation and repression. *PLoS Genet.* 7:e1002356. doi: 10.1371/journal.pgen.1002356
- Levin, M., Hashimshony, T., Wagner, F., and Yanai, I. (2012). Developmental milestones punctuate gene expression in the *Caenorhabditis* embryo. *Dev. Cell* 22, 1101–1108. doi: 10.1016/j.devcel.2012.04.004
- Levine, M., and Davidson, E. H. (2005). Gene regulatory networks for development. *Proc. Natl. Acad. Sci. U.S.A.* 102, 4936–4942. doi: 10.1073/pnas.0408031102
- Li, C. Y., Zhang, Y., Wang, Z., Zhang, Y., Cao, C., Zhang, P. W., et al. (2010). A human-specific de novo protein-coding gene associated with human brain functions. *PLoS Comput. Biol.* 6:e100734. doi: 10.1371/journal.pcbi.100734
- Li, S. W., Arita, M., Kopen, G. C., Phinney, D. G., and Prockop, D. J. (1998). A 1,064 bp fragment from the promoter region of the Col1a2 gene drives lacZ expression not only in cartilage but also in osteoblasts adjacent to regions undergoing both endochondral and intramembranous ossification in mouse embryos. *Matrix Biol.* 17, 213–221. doi: 10.1016/S0945-053X(98)90060-9
- Liu, Y., Li, H., Tanaka, K., Tsumaki, N., and Yamada, Y. (2000). Identification of an enhancer sequence within the first intron required for cartilage-specific transcription of the alpha2(XI) collagen gene. *J. Biol. Chem.* 275, 12712–12718. doi: 10.1074/jbc.275.17.12712
- Lohmander, S., and Hjerpe, A. (1975). Proteoglycans of mineralizing rib and epiphyseal cartilage. *Biochim. Biophys. Acta* 404, 93–109. doi: 10.1016/0304-4165(75)90151-8
- Losos, J. B. (2011). Convergence, adaptation, and constraint. *Evolution* 65, 1827–1840. doi: 10.1111/j.1558-5646.2011.01289.x
- Mak, K. K., Kronenberg, H. M., Chuang, P. T., Mackem, S., and Yang, Y. (2008). Indian hedgehog signals independently of PTHrP to promote chondrocyte hypertrophy. *Development* 135, 1947–1956. doi: 10.1242/dev.018044
- Mallatt, J., and Chen, J. Y. (2003). Fossil sister group of craniates: predicted and found. *J. Morphol.* 258, 1–31. doi: 10.1002/jmor.10081
- Martinek, N., Shahab, J., Sodek, J., and Ringuette, M. (2007). Is SPARC an evolutionarily conserved collagen chaperone? *J. Dent. Res.* 86, 296–305. doi: 10.1177/154405910708600402
- McCauley, D. W., and Bronner-Fraser, M. (2006). Importance of SoxE in neural crest development and the evolution of the pharynx. *Nature* 441, 750–752. doi: 10.1038/nature04691
- McKee, M. D., Glimcher, M. J., and Nanci, A. (1992). High-resolution immunolocalization of osteopontin and osteocalcin in bone and cartilage during endochondral ossification in the chicken tibia. *Anat. Rec.* 234, 479–492. doi: 10.1002/ar.1092340404
- McKenzie, S. K., Oxley, P. R., and Kronauer, D. J. (2014). Comparative genomics and transcriptomics in ants provide new insights into the evolution and function of odorant binding and chemosensory proteins. *BMC Genomics* 15:718. doi: 10.1186/1471-2164-15-718
- Meulemans, D., and Bronner-Fraser, M. (2005). Central role of gene cooption in neural crest evolution. *J. Exp. Zool. B Mol. Dev. Evol.* 304, 298–303. doi: 10.1002/jez.b.21047
- Meulemans, D., and Bronner-Fraser, M. (2007). Insights from amphioxus into the evolution of vertebrate cartilage. *PLoS ONE* 2:e787. doi: 10.1371/journal.pone.0000787
- Miura, S., Hanaoka, K., and Togashi, S. (2008). Skeletogenesis in *Xenopus tropicalis*: characteristic bone development in an anuran amphibian. *Bone* 43, 901–909. doi: 10.1016/j.bone.2008.07.005
- Mizoguchi, I., Takahashi, I., Sasano, Y., Kagayama, M., Kuboki, Y., and Mitani, H. (1997). Localization of types I, II and X collagen and osteocalcin in intramembranous, endochondral and chondroid bone of rats. *Anat. Embryol. (Berl.)* 196, 291–297. doi: 10.1007/s004290050098
- Moczek, A. P. (2008). On the origins of novelty in development and evolution. *Bioessays* 30, 432–447. doi: 10.1002/bies.20754

- Mori-Akiyama, Y., Akiyama, H., Rowitch, D. H., and de Crombrugge, B. (2003). Sox9 is required for determination of the chondrogenic cell lineage in the cranial neural crest. *Proc. Natl. Acad. Sci. U.S.A.* 100, 9360–9365. doi: 10.1073/pnas.1631288100
- Moriishi, T., Shibata, Y., Tsukazaki, T., and Yamaguchi, A. (2005). Expression profile of *Xenopus* banded hedgehog, a homolog of mouse Indian hedgehog, is related to the late development of endochondral ossification in *Xenopus laevis*. *Biochem. Biophys. Res. Commun.* 328, 867–873. doi: 10.1016/j.bbrc.2005.01.032
- Mundlos, S., Schwahn, B., Reichert, T., and Zabel, B. (1992). Distribution of osteonectin mRNA and protein during human embryonic and fetal development. *J. Histochem. Cytochem.* 40, 283–291. doi: 10.1177/40.2.1552170
- Nah, G. S., Tay, B. H., Brenner, S., Osato, M., and Venkatesh, B. (2014). Characterization of the Runx gene family in a jawless vertebrate, the Japanese lamprey (*Lethenteron japonicum*). *PLoS ONE* 9:e113445. doi: 10.1371/journal.pone.0113445
- Nakase, T., Takaoka, K., Hirakawa, K., Hirota, S., Takemura, T., Onoue, H., et al. (1994). Alterations in the expression of osteonectin, osteopontin and osteocalcin mRNAs during the development of skeletal tissues in vivo. *Bone Miner.* 26, 109–122. doi: 10.1016/S0169-6009(08)80056-6
- Necsulea, A., and Kaessmann, H. (2014). Evolutionary dynamics of coding and non-coding transcriptomes. *Nat. Rev. Genet.* 15, 734–748. doi: 10.1038/nrg3802
- Neuhold, L. A., Killar, L., Zhao, W., Sung, M. L., Warner, L., Kulik, J., et al. (2001). Postnatal expression in hyaline cartilage of constitutively active human collagenase-3 (MMP-13) induces osteoarthritis in mice. *J. Clin. Invest.* 107, 35–44. doi: 10.1172/JCI10564
- Nishimura, R., Wakabayashi, M., Hata, K., Matsubara, T., Honma, S., Wakisaka, S., et al. (2012). Osterix regulates calcification and degradation of chondrogenic matrices through matrix metalloproteinase 13 (MMP13) expression in association with transcription factor Runx2 during endochondral ossification. *J. Biol. Chem.* 287, 33179–33190. doi: 10.1074/jbc.M111.337063
- Northcutt, R. G., and Gans, C. (1983). The genesis of neural crest and epidermal placodes: a reinterpretation of vertebrate origins. *Q. Rev. Biol.* 58, 1–28. doi: 10.1086/413055
- Oh, C. D., Lu, Y., Liang, S., Mori-Akiyama, Y., Chen, D., de Crombrugge, B., et al. (2014). SOX9 regulates multiple genes in chondrocytes, including genes encoding ECM proteins, ECM modification enzymes, receptors, and transporters. *PLoS ONE* 9:e107577. doi: 10.1371/journal.pone.0107577
- Ohtani, K., Yao, T., Kobayashi, M., Kusakabe, R., Kuratani, S., and Wada, H. (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, 596–607. doi: 10.1002/jez.b.21231
- Ortega, N., Behonick, D. J., and Werb, Z. (2004). Matrix remodeling during endochondral ossification. *Trends Cell Biol.* 14, 86–93. doi: 10.1016/j.tcb.2003.12.003
- Orth, M. W., Luchene, L. J., and Schmid, T. M. (1996). Type X collagen isolated from the hypertrophic cartilage of embryonic chick tibiae contains both hydroxylysyl- and lysylpyridinoline cross-links. *Biochem. Biophys. Res. Commun.* 219, 301–305. doi: 10.1006/bbrc.1996.0227
- Pacifci, M., Oshima, O., Fisher, L. W., Young, M. F., Shapiro, I. M., and Leboy, P. S. (1990). Changes in osteonectin distribution and levels are associated with mineralization of the chicken tibial growth cartilage. *Calcif. Tissue Int.* 47, 51–61. doi: 10.1007/BF02555866
- Pantalacci, S., and Semon, M. (2015). Transcriptomics of developing embryos and organs: a raising tool for evo-devo. *J. Exp. Zool. B Mol. Dev. Evol.* 324, 363–371. doi: 10.1002/jez.b.22595
- Park, J., Gebhardt, M., Golovchenko, S., Perez-Branguli, F., Hattori, T., Hartmann, C., et al. (2015). Dual pathways to endochondral osteoblasts: a novel chondrocyte-derived osteoprogenitor cell identified in hypertrophic cartilage. *Biol. Open* 4, 608–621. doi: 10.1242/bio.201411031
- Pataquiva-Mateus, A. Y., Wu, H. C., Lucchesi, C., Ferraz, M. P., Monteiro, F. J., and Spector, M. (2012). Supplementation of collagen scaffolds with SPARC to facilitate mineralization. *J. Biomed. Mater. Res. B Appl. Biomater.* 100, 862–870. doi: 10.1002/jbm.b.32650
- Poole, A. R., Pidoux, I., and Rosenberg, L. (1982). Role of proteoglycans in endochondral ossification: immunofluorescent localization of link protein and proteoglycan monomer in bovine fetal epiphyseal growth plate. *J. Cell Biol.* 92, 249–260. doi: 10.1083/jcb.92.2.249
- Rees, S. G., Hughes Wassell, D. T., Waddington, R. J., and Embrey, G. (2001). Interaction of bone proteoglycans and proteoglycan components with hydroxyapatite. *Biochim. Biophys. Acta* 1568, 118–128. doi: 10.1016/S0304-4165(01)00209-4
- Roach, H. I. (1994). Why does bone matrix contain non-collagenous proteins? The possible roles of osteocalcin, osteonectin, osteopontin and bone sialoprotein in bone mineralisation and resorption. *Cell Biol. Int.* 18, 617–628. doi: 10.1006/cbir.1994.1088
- Roach, H. I. (1997). New aspects of endochondral ossification in the chick: chondrocyte apoptosis, bone formation by former chondrocytes, and acid phosphatase activity in the endochondral bone matrix. *J. Bone. Miner. Res.* 12, 795–805. doi: 10.1359/jbmr.1997.12.5.795
- Roach, H. I. (1999). Association of matrix acid and alkaline phosphatases with mineralization of cartilage and endochondral bone. *Histochem. J.* 31, 53–61. doi: 10.1023/A:1003519104980
- 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
- Roughley, P. J. (2006). The structure and function of cartilage proteoglycans. *Eur. Cell Mater.* 12, 92–101.
- Rychel, A. L., Smith, S. E., Shimamoto, H. T., and Swalla, B. J. (2006). Evolution and development of the chordates: collagen and pharyngeal cartilage. *Mol. Biol. Evol.* 23, 541–549. doi: 10.1093/molbev/msj055
- Sansom, I. J., Smith, M. P., Armstrong, H. A., and Smith, M. M. (1992). Presence of the earliest vertebrate hard tissue in conodonts. *Science* 256, 1308–1311. doi: 10.1126/science.1598573
- Sasaki, T., Amizuka, N., Irie, K., Ejiri, S., and Ozawa, H. (2000). Localization of alkaline phosphatase and osteopontin during matrix mineralization in the developing cartilage of coccygeal vertebrae. *Arch. Histol. Cytol.* 63, 271–284. doi: 10.1679/aohc.63.271
- Shubin, N. H. (2002). Origin of evolutionary novelty: examples from limbs. *J. Morphol.* 252, 15–28. doi: 10.1002/jmor.10017
- Shubin, N., Tabin, C., and Carroll, S. (2009). Deep homology and the origins of evolutionary novelty. *Nature* 457, 818–823. doi: 10.1038/nature07891
- Smith, M. M., and Hall, B. K. (1990). Development and evolutionary origins of vertebrate skeletogenic and odontogenetic tissues. *Biol. Rev. Camb. Philos. Soc.* 65, 277–373. doi: 10.1111/j.1469-185X.1990.tb01427.x
- Smith, S. E., Douglas, R., da Silva, K. B., and Swalla, B. J. (2003). Morphological and molecular identification of *Saccoglossus* species (Hemichordata: Harrimaniidae) in the Pacific Northwest. *Can. J. Zool.* 81, 133–141. doi: 10.1139/z02-228
- Stayton, C. T. (2008). Is convergence surprising? An examination of the frequency of convergence in simulated datasets. *J. Theor. Biol.* 252, 1–14. doi: 10.1016/j.jtbi.2008.01.008
- Stricker, S., Fundele, R., Vortkamp, A., and Mundlos, S. (2002). Role of Runx genes in chondrocyte differentiation. *Dev. Biol.* 245, 95–108. doi: 10.1006/dbio.2002.0640
- Takeda, S., Bonnamy, J. P., Owen, M. J., Ducy, P., and Karsenty, G. (2001). Continuous expression of Cbfa1 in nonhypertrophic chondrocytes uncovers its ability to induce hypertrophic chondrocyte differentiation and partially rescues Cbfa1-deficient mice. *Genes Dev.* 15, 467–481. doi: 10.1101/gad.845101
- Talwar, R. M., Wong, B. S., Svoboda, K., and Harper, R. P. (2006). Effects of estrogen on chondrocyte proliferation and collagen synthesis in skeletally mature articular cartilage. *J. Oral Maxillofac. Surg.* 64, 600–609. doi: 10.1016/j.joms.2005.12.006
- Tautz, D., and Domazet-Loso, T. (2011). The evolutionary origin of orphan genes. *Nat. Rev. Genet.* 12, 692–702. doi: 10.1038/nrg3053
- 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
- Thorogood, P. (1988). The developmental specification of the vertebrate skull. *Development* 103(Suppl.), 141–153.
- Toll-Riera, M., Bosch, N., Bellora, N., Castello, R., Armengol, L., Estivill, X., et al. (2009a). Origin of primate orphan genes: a comparative genomics approach. *Mol. Biol. Evol.* 26, 603–612. doi: 10.1093/molbev/msn281

- Toll-Riera, M., Castelo, R., Bellora, N., and Alba, M. M. (2009b). Evolution of primate orphan proteins. *Biochem. Soc. Trans.* 37, 778–782. doi: 10.1042/BST0370778
- Uy, B. R., Simoes-Costa, M., Sauka-Spengler, T., and Bronner, M. E. (2012). Expression of Sox family genes in early lamprey development. *Int. J. Dev. Biol.* 56, 377–383. doi: 10.1387/ijdb.113416bu
- 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
- Vieira, F. A., Thorne, M. A., Stueber, K., Darias, M., Reinhardt, R., Clark, M. S., et al. (2013). Comparative analysis of a teleost skeleton transcriptome provides insight into its regulation. *Gen. Comp. Endocrinol.* 191, 45–58. doi: 10.1016/j.ygenc.2013.05.025
- Villanueva, A. R., Kujawa, M., Mathews, C. H., and Parfitt, A. M. (1983). Identification of the mineralization front: comparison of a modified toluidine blue stain with tetracycline fluorescence. *Metab. Bone Dis. Relat. Res.* 5, 41–45. doi: 10.1016/0221-8747(83)90049-8
- Volkmann, D., and Baluska, F. (2006). Gravity: one of the driving forces for evolution. *Protoplasma* 229, 143–148. doi: 10.1007/s00709-006-0200-4
- Vortkamp, A., Lee, K., Lanske, B., Segre, G. V., Kronenberg, H. M., and Tabin, C. J. (1996). Regulation of rate of cartilage differentiation by Indian hedgehog and PTH-related protein. *Science* 273, 613–622. doi: 10.1126/science.273.5275.613
- Wada, H. (2010). Origin and genetic evolution of the vertebrate skeleton. *Zool. Sci.* 27, 119–123. doi: 10.2108/zsj.27.119
- Wagner, D. O., and Aspenberg, P. (2011). Where did bone come from? *Acta Orthop.* 82, 393–398. doi: 10.3109/17453674.2011.588861
- Wagner, G. P., and Lynch, V. J. (2010). Evolutionary novelties. *Curr. Biol.* 20, R48–R52. doi: 10.1016/j.cub.2009.11.010
- Wallace, A. (1997). *The Origin of Animal Body Plans: A Study in Evolutionary Developmental Biology*. Cambridge: Cambridge University Press, 256–284.
- Wang, Z., Gerstein, M., and Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nat. Rev. Genet.* 10, 57–63. doi: 10.1038/nrg2484
- Watanabe, H., Yamada, Y., and Kimata, K. (1998). Roles of aggrecan, a large chondroitin sulfate proteoglycan, in cartilage structure and function. *J. Biochem.* 124, 687–693. doi: 10.1093/oxfordjournals.jbchem.a022166
- Weng, J. J., and Su, Y. (2013). Nuclear matrix-targeting of the osteogenic factor Runx2 is essential for its recognition and activation of the alkaline phosphatase gene. *Biochim. Biophys. Acta* 1830, 2839–2852. doi: 10.1016/j.bbagen.2012.12.021
- Witten, P. E., Huysseune, A., and Hall, B. K. (2010). A practical approach for the identification of the many cartilaginous tissues in teleost fish. *J. Appl. Ichthyol.* 26, 257–262. doi: 10.1111/j.1439-0426.2010.01416.x
- Wu, H., Whitfield, T. W., Gordon, J. A., Dobson, J. R., Tai, P. W., van Wijnen, A. J., et al. (2014). Genomic occupancy of Runx2 with global expression profiling identifies a novel dimension to control of osteoblastogenesis. *Genome Biol.* 15:R52. doi: 10.1186/gb-2014-15-3-r52
- Yanagisita, M. (1993). Function of proteoglycans in the extracellular matrix. *Acta Pathol. Jpn.* 43, 283–293. doi: 10.1111/j.1440-1827.1993.tb02569.x
- Yasui, N., Ono, K., Konomi, H., and Nagai, Y. (1984). Transitions in collagen types during endochondral ossification in human growth cartilage. *Clin. Orthop. Relat. Res.* 215–218.
- Yoshida, C. A., Yamamoto, H., Fujita, T., Furuchi, T., Ito, K., Inoue, K., et al. (2004). Runx2 and Runx3 are essential for chondrocyte maturation, and Runx2 regulates limb growth through induction of Indian hedgehog. *Genes Dev.* 18, 952–963. doi: 10.1101/gad.1174704
- Ytrehus, B., Carlson, C. S., Lundeheim, N., Mathisen, L., Reinholt, F. P., Teige, J., et al. (2004). Vascularisation and osteochondrosis of the epiphyseal growth cartilage of the distal femur in pigs—development with age, growth rate, weight and joint shape. *Bone* 34, 454–465. doi: 10.1016/j.bone.2003.07.011
- Yu, J. K., Meulemans, D., McKeown, S. J., and Bronner-Fraser, M. (2008). Insights from the amphioxus genome on the origin of vertebrate neural crest. *Genome Res.* 18, 1127–1132. doi: 10.1101/gr.076208.108
- Zaragoza, C., Lopez-Rivera, E., Garcia-Rama, C., Saura, M., Martinez-Ruiz, A., Lizarbe, T. R., et al. (2006). Cbfa-1 mediates nitric oxide regulation of MMP-13 in osteoblasts. *J. Cell Sci.* 119, 1896–1902. doi: 10.1242/jcs.02895
- Zhang, G., and Cohn, M. J. (2006). Hagfish and lancelet fibrillar collagens reveal that type II collagen-based cartilage evolved in stem vertebrates. *Proc. Natl. Acad. Sci. U.S.A.* 103, 16829–16833. doi: 10.1073/pnas.0605630103
- Zhang, G., Eames, B. F., and Cohn, M. J. (2009). Chapter 2. Evolution of vertebrate cartilage development. *Curr. Top. Dev. Biol.* 86, 15–42. doi: 10.1016/S0070-2153(09)01002-3
- Zhou, G., Zheng, Q., Engin, F., Munivez, E., Chen, Y., Sebald, E., et al. (2006). Dominance of SOX9 function over RUNX2 during skeletogenesis. *Proc. Natl. Acad. Sci. U.S.A.* 103, 19004–19009. doi: 10.1073/pnas.0605170103
- Zhou, X., von der Mark, K., Henry, S., Norton, W., Adams, H., and de Crombrugghe, B. (2014). Chondrocytes transdifferentiate into osteoblasts in endochondral bone during development, postnatal growth and fracture healing in mice. *PLoS Genet.* 10:e1004820. doi: 10.1371/journal.pgen.1004820

Conflict of Interest Statement: 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 © 2015 Gómez-Picos and Eames. 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) or licensor 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.

The TALE face of Hox proteins in animal evolution

Samir Merabet^{1,2*} and Brigitte Galliot³

¹ Centre National de Recherche Scientifique, Institut de Génomique Fonctionnelle de Lyon, Lyon, France, ² Institut de Génomique Fonctionnelle de Lyon, Ecole Normale Supérieure de Lyon, Lyon, France, ³ Department of Genetics and Evolution, Faculty of Science, Institute of Genetics and Genomics in Geneva, University of Geneva, Geneva, Switzerland

OPEN ACCESS

Edited by:

Sylvain Marcellini,
 University of Concepcion, Chile

Reviewed by:

Ingo Braasch,
 University of Oregon, USA
 Pedro Martinez,
 Universitat de Barcelona, Spain

***Correspondence:**

Samir Merabet,

Centre National de Recherche Scientifique, Ecole Normale Supérieure de Lyon, UMR5242, 46 Allée d'Italie, Lyon 69007, France
 samir.merabet@ens-lyon.fr

Specialty section:

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

Received: 14 June 2015

Accepted: 31 July 2015

Published: 18 August 2015

Citation:

Merabet S and Galliot B (2015) The TALE face of Hox proteins in animal evolution. *Front. Genet.* 6:267.
 doi: 10.3389/fgene.2015.00267

Hox genes are major regulators of embryonic development. One of their most conserved functions is to coordinate the formation of specific body structures along the anterior-posterior (AP) axis in Bilateria. This architectural role was at the basis of several morphological innovations across bilaterian evolution. In this review, we traced the origin of the Hox patterning system by considering the partnership with PBC and Meis proteins. PBC and Meis belong to the TALE-class of homeodomain-containing transcription factors and act as generic cofactors of Hox proteins for AP axis patterning in Bilateria. Recent data indicate that Hox proteins acquired the ability to interact with their TALE partners in the last common ancestor of Bilateria and Cnidaria. These interactions relied initially on a short peptide motif called hexapeptide (HX), which is present in Hox and non-Hox protein families. Remarkably, Hox proteins can also recruit the TALE cofactors by using specific PBC Interaction Motifs (SPIMs). We describe how a functional Hox/TALE patterning system emerged in eumetazoans through the acquisition of SPIMs. We anticipate that interaction flexibility could be found in other patterning systems, being at the heart of the astonishing morphological diversity observed in the animal kingdom.

Keywords: Hox, PBC, Meis, Metazoa, patterning, early-branching phyla, HX, SPIMs

Introduction

The phenotypic diversity observed in the animal kingdom arose from genetic innovations that modulate developmental processes, a step in evolution that often precedes speciation events (Gould, 1992; Arthur, 2002). A major challenge in biology is to characterize these genetic innovations and to understand how they impact developmental processes. Remarkably, the specification of body plans and body parts in species as different as humans or flies is controlled by a relatively small and highly conserved genetic repertoire called the “genetic toolkit” (True and Carroll, 2002; Erwin, 2009). This genetic toolkit, which acts at restricted stages of embryonic development, encodes for molecules involved in cell-cell communication, and gene regulation (Mann and Carroll, 2002). Components of the genetic toolkit are described in several bilaterian species to form character identification networks (Wagner, 2007), or kernels (Davidson and Erwin, 2006), which are part of large developmental networks that underlie body plan development (Davidson and Erwin, 2006). Several members of the genetic toolkit are also expressed in choanoflagellates, indicating that they originated prior to the emergence of the first metazoans (King et al., 2003; King, 2004; Wenger and Galliot, 2013).

Abbreviations: ANTP, Antennapedia; AP, anterior posterior; HD, Homeodomain; HX, Hexapeptide; PG, Paralog Group; SPIM, Specific PBC Interaction Motif; TALE, Three Amino acid Loop Extension; TF, Transcription factor.

The large majority of contemporary animals belong to Bilateria, which are characterized by three embryonic germ layers (ectoderm, mesoderm, endoderm) and a bilateral symmetry that results from the orthogonal intersection of two longitudinal axes, the anterior-posterior (AP) axis (also referred to as the primary axis), and the dorso-ventral (DV) axis (also referred to as the secondary axis). Bilaterians radiated during the Cambrian period some 500–550 million years ago. Other extant non-bilaterian species belong to Porifera (sponges), Ctenophora, Placozoa (Trichoplax), and Cnidaria, whose ancestors predate the Cambrian explosion, thus often named early-branched phyla (**Figure 1**). With the exception of Placozoa, species from these early-branched phyla display different types of symmetry, either radial (as seen in sponge larvae, some adult sponges, and in most cnidarians), or biradial (as seen in ctenophores), or partly bilateral (as seen in sea anemone species that belong to the anthozoan class of cnidarians). These various symmetries are especially evident during embryogenesis and larval stages and depend on the formation of a primary body axis (Ryan and Baxevanis, 2007).

Cnidaria, a sister group to Bilateria, share with them typical features of eumetazoans, i.e., an ectodermal layer that differentiates as an epidermis, an endodermal layer that differentiates as a gut, and a nervous system, which, at the oral pole/extremity, allows an active feeding behavior. Also, Cnidaria includes a large variety of taxa with a wide spectrum of morphological diversity. All together, these characteristics place Cnidaria at a key phylogenetic position for tracing the emergence of molecular innovations that underlie developmental changes and diversification in animal evolution (Steele et al., 2011). Representative(s) of the main gene families involved in the specification of eumetazoan features are also found in Cnidaria (Martindale, 2005). Their study is however more challenging, due to the lack of advanced genetic tools that could allow establishing transgenic animals for stable gene expression or extinction in a tissue- and/or stage-specific manner.

Among the different conserved developmental gene families are the Hox genes, which are considered as the “Rosetta Stone” of the genetic toolkit. Hox genes were initially discovered in *Drosophila*, then rapidly investigated in vertebrate species, showing striking conserved features throughout bilaterian lineages (Lewis, 1978; McGinnis and Krumlauf, 1992; Kmíta and Duboule, 2003). These conserved properties have been discussed in several reviews and relate to their clustered genomic organization that constrains embryonic expression (Duboule, 2007), but also to the presence of several typical protein signatures (Ogishima and Tanaka, 2007; Merabet et al., 2009). Modifications in Hox gene expression or in Hox protein function have been linked to several morphological innovations during the evolution of bilaterians (Pearson et al., 2005; Heffer et al., 2013). The presence of Hox genes in Cnidaria therefore raised the question of their role in the emergence of innovations shared by cnidarians and bilaterians, as well as in the emergence of innovations responsible for the morphological diversity observed among cnidarian species.

The most spectacular observation came from the embryo of the cnidarian sea anemone *Nematostella vectensis*, where

several Hox-related genes show a staggered-like expression pattern along the oral-aboral (OA) axis. This expression profile led to the proposition that the cnidarian OA axis could be homologous to the bilaterian AP axis (Finnerty et al., 2004; Matus et al., 2006). The OA expression profile of *Nematostella* Hox genes is however neither conserved in other cnidarian lineages nor strictly following the collinear rules normally observed in Bilateria (Gauchat et al., 2000; Finnerty et al., 2004; Kamm et al., 2006; Ryan et al., 2007; Chiori et al., 2009). These additional observations led to the opposite conclusion that a Hox patterning system is likely not existing in Cnidaria (Kamm et al., 2006).

Surprisingly, the question of the evolution of Hox patterning mechanisms is rarely approached at the protein level, in particular by considering members of the PBC and Meis families. PBC and Meis are crucial patterning cofactors of Hox proteins along the AP axis, a partnership that is evolutionarily-conserved throughout Bilateria (Moens and Selleri, 2006; Mann et al., 2009). PBC and Meis belong to the TALE (Three Amino acids Loop Extension) class of homeodomain (HD)-containing transcription factors (Bürglin, 1997), are widely conserved across metazoans and can therefore be used as a molecular hallmark of the Hox patterning system. In this review, we report how the intricate interaction properties between Hox and TALE proteins were progressively acquired in pre-bilaterian animal evolution to eventually constitute a major patterning system.

Origin and Early Evolution of the Hox/ParaHox and PBC/Meis Gene Families

Hox proteins belong to the ANTP (Antennapedia) class of HD-containing transcription factors. This class contains two large groups of sister gene families: (i) the non-Hox ANTP-class group, which includes the NK and Extended (Ext)-Hox gene families, and (ii) the Hox/ParaHox genes (Garcia-Fernández, 2005). The Hox gene family is usually found organized in clusters and contains several paralog groups (PGs) that are themselves classified into anterior (PG1-3), central (PG4-8), and posterior (PG9-14) (Duboule, 2007). The ParaHox family contains three clustered genes initially discovered in the cephalochordate amphioxus (Brooke et al., 1998), and named Gsx, Pdx/Xlox, and Cdx. ParaHox genes share common ancestors with specific Hox gene families, Gsx, and Pdx/Xlox with the anterior PG2/PG3, Cdx with the posterior PG9 (Quiquand et al., 2009).

Two different scenarios are proposed to explain the evolutionary history of the Hox/ParaHox gene family with regard to the other ANTP-class members. In the first one, the Hox/ParaHox family is specific to eumetazoans (which regroup Bilateria, Cnidaria, and Placozoa) and would have originated from duplications of a ProtoHox gene derived from NK genes and related to Evx/Mox (Ext-Hox family) (Gauchat et al., 2000; Minguijón and Garcia-Fernández, 2003; Larroux et al., 2007; Quiquand et al., 2009; Ryan et al., 2010) (**Figure 1A**). This scenario is supported by the presence of a Gsx ParaHox gene in the genome of the placozoan *Trichoplax adhaerens* (Schierwater and Kuhn, 1998; Schierwater et al., 2008b), the presence of several NK representatives and the absence of Hox/ParaHox

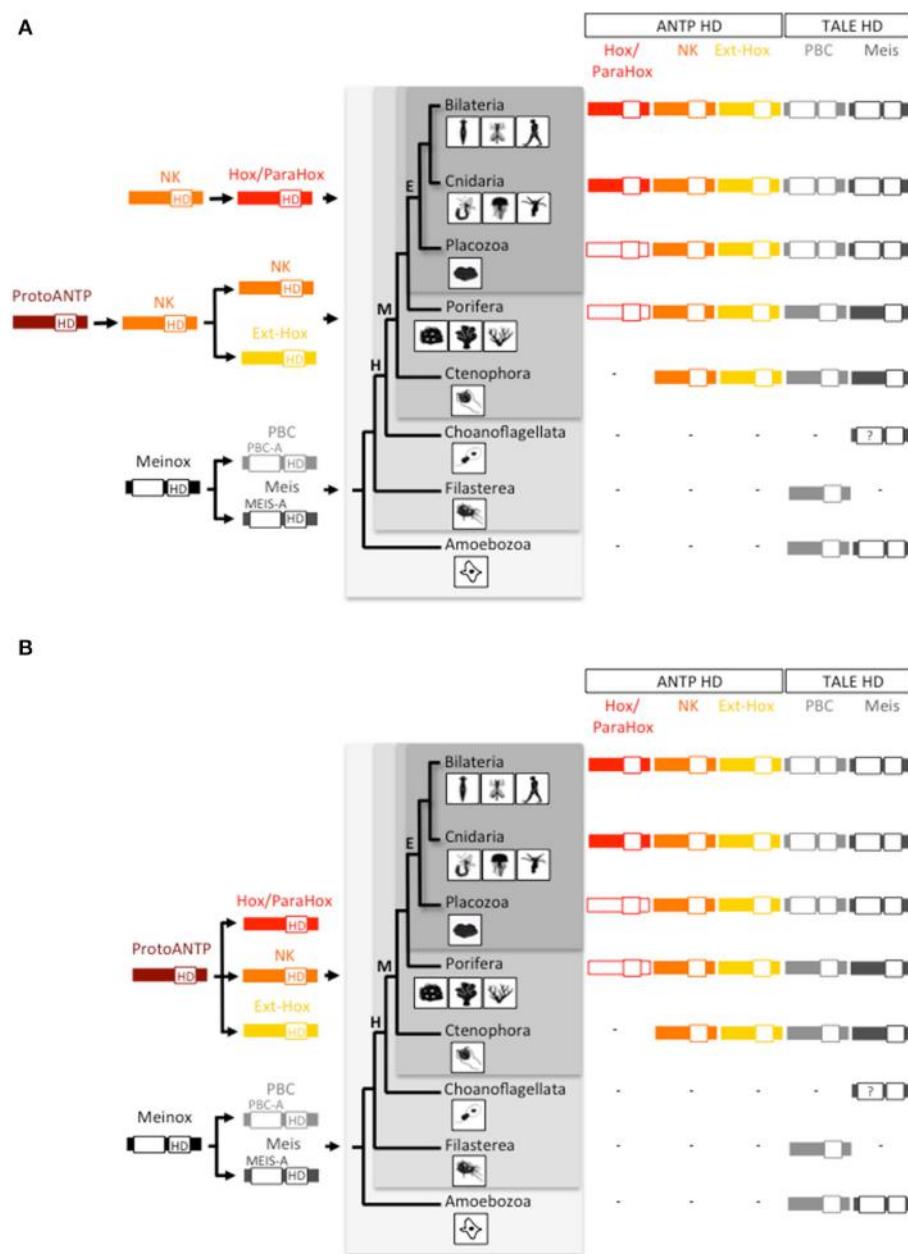


FIGURE 1 | Origin and early evolution of ANTP- and TALE-class gene families. **(A)** First evolutionary scenario whereby the Hox/ParaHox family would have derived from a NK member in the Eumetazoan ancestor. **(B)** Second evolutionary scenario, also named the “ghost loci hypothesis,” whereby the main homeobox gene families (Hox/ParaHox, NK, and Ext-Hox) would have derived from a ProtoANTP cluster of homeobox genes already present in the Last Common Ancestor (LCA) of metazoans. The recent finding of a ParaHox-like gene in Porifera (Fortunato et al., 2014) actually supports the second scenario. Note that only one ParaHox member (symbolized by the absence of red filling) is found in Placozoa [annotated as

a Gsx-like: (Schierwater et al., 2008a)] and Porifera [annotated as a Cdx-like: (Fortunato et al., 2014)] and that no Hox or ParaHox gene has been annotated in Ctenophora so far. In comparison, the PBC and Meis families originated earlier in the life tree, with representatives already present in unicellular phyla (Amoebozoa and Filasterea). Graded gray backgrounds highlight Eumetazoa (E), Metazoa (M), Holozoa (H), and Unikonta (U) super phyla. The homeodomain (HD) is indicated in each protein. PBC-A and Meis-A are domains required for the PBC/Meis partnership. Question mark in Choanoflagellata is for incomplete protein sequence of Meis. Animal drawings were taken from Ryan and Baxevanis (2007).

genes in the genome of the ctenophores *Mnemiopsis leidyi* (Ryan et al., 2010) and *Pleurobrachia bachei* (Moroz et al., 2014) and the demosponge *Amphimedon queenslandica* (Srivastava

et al., 2010). However, this scenario is challenged by the “ghost loci hypothesis”, which postulates that Hox/ParaHox genes were already present in the Last Common Ancestor (LCA) of

metazoans and secondarily lost in Porifera over evolutionary times (Mendivil Ramos et al., 2012). In this second scenario, the Hox/ParaHox and NK families emerged independently from a common ProtoANTP ancestor gene (**Figure 1B**). The recent finding of a Cdx-like ParaHox gene in the genome of two calcareous sponges (Fortunato et al., 2014), now argues in favor of the ghost loci hypothesis.

In addition to ANTP, other classes of HD-containing transcription factors are also present in early branch phyla. These include the Paired-like, Pax, Pou, Lim, Six, and TALE classes (Galliot and de Vargas, 1999; Larroux et al., 2008; Srivastava et al., 2010; Holland, 2013; Fortunato et al., 2014). Members of the TALE class contain an atypical 63-residues long HD, due to the presence of three extra residues in between the helices 1 and 2 of the HD (Mukherjee and Bürglin, 2007). TALE class members are among the most ancient transcription factors in eukaryotes, with several of them present in unicellular organisms, plants and fungi (Bürglin, 1997, 1998), therefore predating the origin of animals. Interestingly, TALE-class members can interact with different types of HD-containing proteins in plants (Bellaoui et al., 2001; Hackbusch et al., 2005; Kanrar et al., 2006; Hay and Tsiantis, 2010), fungi (Keleher et al., 1989; Stark and Johnson, 1994; Carr et al., 2004), and animals (Bürglin, 1998).

The TALE class comprises five families (PBC, Meis, Iro, TGIF, and MKX), among which two, PBC and Meis are known to interact with ANTP members (Mukherjee and Bürglin, 2007). PBC and Meis were already present before multicellular organisms appeared (King et al., 2008; Clarke et al., 2013; Suga et al., 2013) and (**Figure 1**). Animal representatives of PBC and Meis include the Pbx1-4 or Extradenticle (Exd) and Meis1-3 or Homothorax (Hth) proteins, as named in mammals and in the fruit fly *Drosophila melanogaster*, respectively. PBC and Meis families originated from the duplication of a common ancestor gene named MEINOX, and this duplication was proposed to coincide with the apparition of the first Hox cluster in metazoans (Bürglin, 1998). Genome comparisons between early-branched metazoan species and unicellular organisms now establish that the PBC/Meis duplication predated the ANTP class and therefore the Hox/ParaHox family. Thomas Bürglin was however the first one to consider the partnership between Hox and TALE proteins as an informative molecular hallmark to trace the origin of the Hox patterning system in Metazoa (Bürglin, 1998).

The Ground State of Hox/TALE Interaction Networks in Bilateria: Role of the Hexapeptide (HX) Motif

The formation of Hox/PBC/Meis complexes in Bilateria is described to rely on Hox-PBC and PBC-Meis interactions (**Figure 2A**). Interaction between PBC and Meis involves the N-terminal PBC-A and Meis-A domains, respectively (Mann and Affolter, 1998). In the absence of Meis, the PBC-A domain is masking two nuclear localization signals located in the HD of PBC. The interaction with Meis relieves the masking activity of the PBC-A domain, allowing the nuclear translocation of PBC (Saleh et al., 2000; Stevens and Mann, 2007).

Interactions between Hox and PBC have been extensively studied at the biochemical and structural levels. All these analyses

converge to show a preponderant role for a short conserved motif present in Hox proteins, named hexapeptide (HX) (Mann et al., 2009). The HX motif lies upstream to the HD and contains a core Y/FPWM sequence in all but Abdominal B-group Hox proteins, which have a more divergent sequence (Merabet et al., 2009). More generally, the HX motif is defined as a PBC interaction motif (PIM) that contains an invariant Tryptophan residue located in a hydrophobic environment, followed by basic residues from +2 to +5 (In der Rieden et al., 2003). Crystal structures of vertebrate and invertebrate Hox/PBC complexes solved with anterior, central or posterior Hox proteins point to the critical role of the Tryptophan residue in maintaining strong interactions within the hydrophobic pocket formed in part by the three extra residues of the PBC HD (Passner et al., 1999; Piper et al., 1999; LaRonde-LeBlanc and Wolberger, 2003; Joshi et al., 2007). A recent structural analysis of the Hox/PBC complex bound on a physiological DNA-binding site further underlined that Hox paralog specific residues located in the N-terminal arm of the HD and in the linker region connecting the HX motif to the HD are important for recognizing a specific shape of the DNA minor groove in the presence of PBC (Joshi et al., 2007). SELEX-seq based approaches confirmed that *Drosophila* Hox/PBC complexes preferentially recognize different nucleotide sequences characterized by distinct minor groove topographies (Slattery et al., 2011). These results open new avenues for apprehending the molecular mechanisms underlying Hox and Hox/PBC DNA-binding specificity (Abe et al., 2015). Nevertheless, the systematic involvement of a unique Hox protein motif in the interaction with PBC does not easily explain the broad variety of functions that Hox/TALE complexes have *in vivo* (Hueber and Lohmann, 2008; Mann et al., 2009).

Specific PBC Interactions Motifs (SPIMs) as Versatile Complements to Diversify Hox/TALE Interaction Properties in Bilateria and Cnidaria

Our knowledge of Hox-TALE interaction properties results mostly from *in vitro* approaches. Along the same line, the duplication of Pbx and Meis genes in vertebrates could provide a supplementary layer of complexity. For example, direct Hox-Meis interactions are described with mouse proteins but their functional significance remains to be elucidated (Shen et al., 1997; Williams et al., 2005). The existence of alternative modes in Hox-PBC interaction came from the observation that the HX mutation does not obligatorily affect PBC-dependent functions of Hox proteins in the *Drosophila* embryo (Galant et al., 2002; Merabet et al., 2003). Additionally, several central and posterior Hox proteins from vertebrates and invertebrates interact with the TALE cofactors independently of the HX motif *in vitro* and *in vivo* (Hudry et al., 2012). Interestingly, HX-independent interactions between Hox and PBC are most often observed in the presence of Meis, and the involvement of Meis in such HX-independent interactions actually depends on its DNA-binding near the Hox/PBC binding site (Hudry et al., 2012). In other words, in acting at the level of target *cis*-regulatory sequences, Meis contributes to diversify the mode of Hox-PBC interactions and thus Hox functions (Merabet and Hudry, 2013).

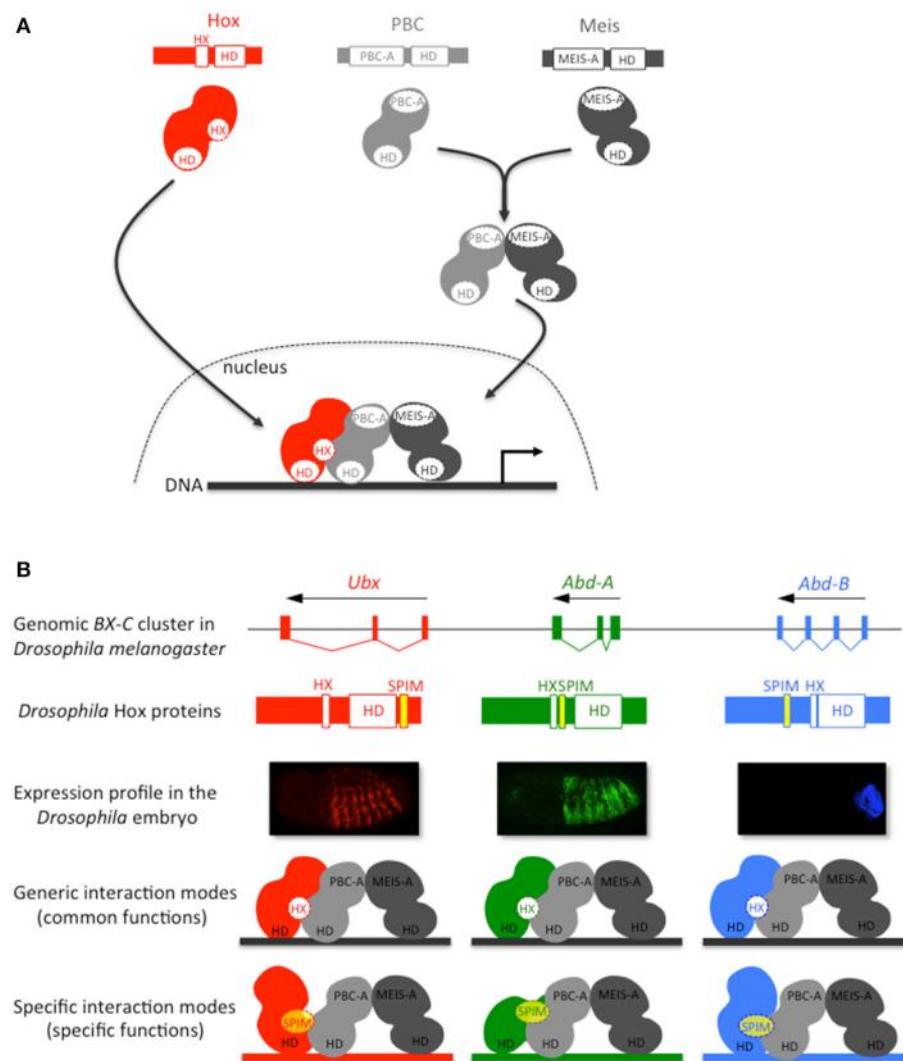


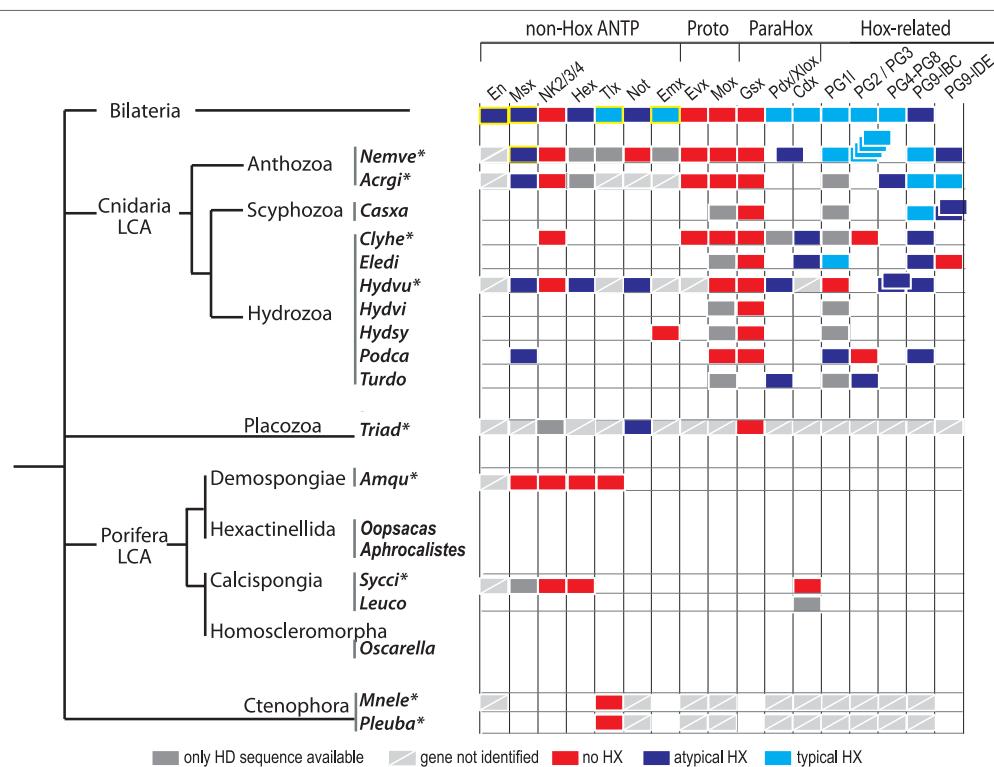
FIGURE 2 | The Hox-TALE interaction network: role of generic (HX motif) and specific PBC interaction motifs (SPIMs). **(A)** Generic association mode between Hox and TALE proteins. The interaction between Meis and PBC allows the nuclear translocation of PBC. The hexapeptide (HX) motif, present in Hox proteins of all bilaterian lineages, is necessary and sufficient for the generic association mode of the Hox/TALE complex on DNA. **(B)** Model for the role of SPIMs in specifying patterning functions

among *Drosophila* Hox proteins. The usage of SPIMs allows each Hox protein of the Bithorax complex (BX-C) to adopt different conformation modes with the TALE cofactors and regulate different target genes *in vivo* (as illustrated by the color code). The placement of SPIMs (highlighted in yellow) in Ultrabithorax (*Ubx*) and Abdominal-A (*Abd-A*) reflects the position of the UbdA and TDWM motifs, respectively (Hudry et al., 2012). The placement of the SPIM in Abdominal-B (*Abd-B*) is speculative.

The flexibility of Hox-TALE interaction properties is predicted to rely on Hox protein motifs that are more gene-specific than the generic HX motif. These motifs are named SPIMs [Specific PBC Interaction Motifs, see also Merabet and Hudry, 2013]. Like the HX motif, SPIMs belong to the so-called short linear motifs, which are classically 5–10 residues long and most often located within intrinsically disordered protein regions (Tompa et al., 2014). Two such motifs have been identified in the *Drosophila* Ultrabithorax (*Ubx*) and AbdominalA (*AbdA*) proteins (Merabet et al., 2007, 2011; Hudry et al., 2012). One of them is conserved in insect *AbdA* proteins, with a core TDWM sequence reminiscent of the HX motif. The other motif, named

UbdA, is conserved between the protostome *Ubx* and *AbdA* proteins (Balavoine et al., 2002). Recent structural analyses showed that the UbdA motif constitutes a flexible extension of the HD that can establish direct contacts with the PBC partner (Foos et al., 2015). Altogether, studies with *Ubx* and *AbdA* confirm that Hox-TALE interactions and functions can rely on species- and/or paralog-specific motifs.

SPIMs remain to be identified in the majority of Hox proteins exerting HX-independent interactions with the TALE cofactors. Still, the usage of different SPIMs in Hox proteins constitutes an appealing molecular strategy for supporting the specific patterning functions of Hox/TALE complexes during

TABLE 1 | Presence or absence of the HX motif among the Hox/ParaHox and non-Hox/ParaHox families across Metazoa.

The color code denotes for presence or absence of the HX motif, and for incomplete or non-annotated gene, as indicated. Boxes surrounded in yellow in non-Hox/ParaHox proteins highlight a demonstrated role of the HX motif for interaction with TALE partners. See main text for details. Protein sequences were retrieved from Uniprot/Swissprot. Stars denote species with sequenced genome.

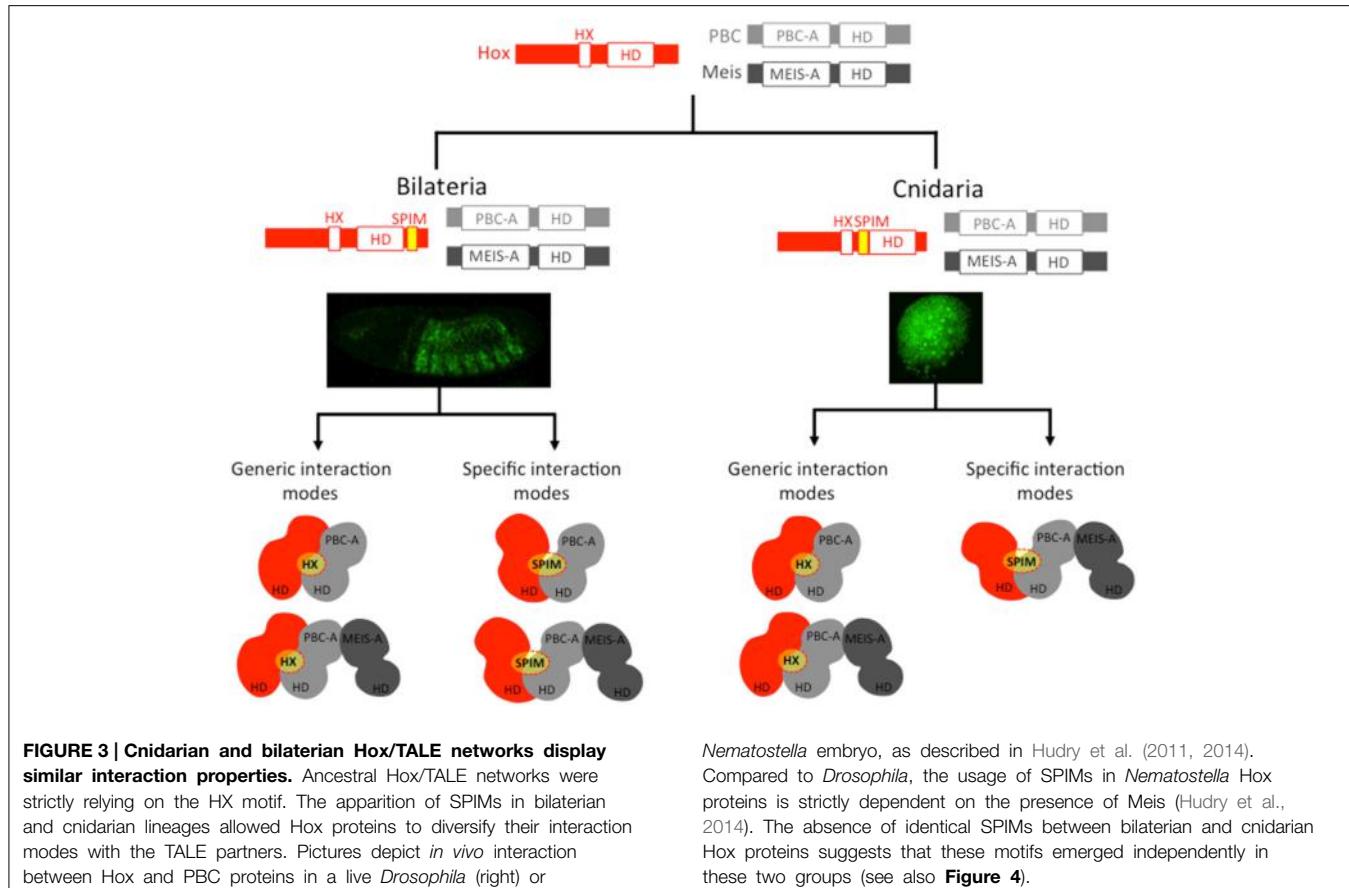
development (Figure 2B). Moreover, the conservation of this property in vertebrate and invertebrate species (Hudry et al., 2012) strongly suggests that interaction flexibility is ancient in Bilateria. As a consequence, it is of upmost interest to trace its origin beyond Bilateria and assess its role in developmental and/or patterning functions.

Besides Bilateria, Cnidaria is the only other phylum that contains a bona fide Hox repertoire (Chourrout et al., 2006; Kamm et al., 2006). As mentioned previously, the role of cnidarian Hox genes in axis patterning is unclear. Furthermore, not all cnidarian Hox proteins contain an intact HX motif [(Hudry et al., 2014) and Table 1]. Nevertheless, as cnidarians express PBC and Meis genes (Matus et al., 2006; Hudry et al., 2014), a Hox/PBC/Meis network could potentially exist. The interaction properties of Hox, PBC and Meis proteins of the sea anemone *Nematostella vectensis* were recently tested, and as expected, these proteins form dimeric and trimeric complexes *in vitro* (Hudry et al., 2014). In addition, mutating the HX motif leads to the loss of the cnidarian Hox/PBC complex, but this loss is rescued in the presence of Meis. Hence, as observed in bilaterians, the *Nematostella* Meis allows *Nematostella* Hox proteins to use alternative modes of interaction with PBC. Thus, bilaterian and cnidarian Hox proteins share the property of using different interfaces for recruiting the TALE cofactors. We propose that these additional interfaces could correspond to SPIMs that remain to be identified in several instances (Figure 3).

Moreover, with the exception of the HX motif, bilaterian and cnidarian Hox proteins do not share strong sequence similarities outside the HD, suggesting that those putative SPIMs could have evolved independently during eumetazoan evolution (see also below).

Genesis of ANTP-TALE Networks during Early Metazoan Evolution

Molecular analyses underline that the HX motif is a generic interaction platform for recruiting the TALE partners. We therefore analyzed a large number of available protein sequences for assessing the presence of a putative HX motif in ANTP class members. A peptide sequence was considered as a putative HX motif when containing the consensus Y/FPWM (typical HX motif) or a single W (atypical/divergent HX motif) residue followed by a basic residue (R or K) from +2 to +6 and not localized more than 30 residues away from the HD (Table 1). In Bilateria, the HX motif is found in almost all Hox/ParaHox members, and in several individual representatives of non-Hox/ParaHox protein families, including Engrailed (En), Msx, Hex, Tlx, Not, and Emx proteins (Table 1). The HX motif is found in cnidarian Hox/ParaHox members among early-branched animal phyla. It is however less conserved when compared to Bilateria, being lost or divergent in several cnidarian lineages (Table 1). Atypical HX motifs are also found in Msx and Hex members of Cnidaria, and in Not members of Cnidaria



and Placozoa (Table 1). Interestingly, the Evx, Mox, and Gsx proteins, which likely represent the most ancestral ProtoHox and Hox/ParaHox family members (Minguillón and García-Fernández, 2003; Quiquand et al., 2009) all lack the HX motif (Table 1).

PBC-recruiting functions have been assigned to few non-Hox proteins among the ANTP class so far. Among them are the mammalian Tlx, *Drosophila* En and *Nematostella* Msx proteins, which do interact in a fully HX-dependent manner with the TALE cofactors (Rhee et al., 2004; Brendolan et al., 2005; Fujioka et al., 2012; Hudry et al., 2014). Still, these proteins display subtle differences in their TALE interaction properties. For example, the *Drosophila* En protein interacts with PBC or PBC/Meis in a HX-dependent manner (Hudry et al., 2014). By comparison, the Msx protein from *Nematostella* interacts in a HX-dependent manner with PBC, but only in the presence of Meis (Hudry et al., 2014). These observations highlight that the role of the PBC/Meis partnership in HX-dependent interactions can be different depending on the protein family and animal lineage considered.

We propose two different evolutionary scenarios to explain the presence of the HX motif in several ANTP family members among metazoan lineages: (i) either the HX motif was already present in the ProtoANTP ancestor, constituting the first molecular interface for recruiting the TALE cofactors

(Figure 4A), or (ii) it emerged multiple times independently in the different ANTP families across animal evolution (Figure 4B). The position of the HX motif systematically located in the upstream vicinity of the HD supports the first scenario. As a corollary, the absence of any HX-like motif in all but one (Not) ANTP members of Placozoa, Porifera and Ctenophora would be attributed to repeated secondary losses. Although more sequences are needed in these three early-branched animal phyla, this apparently global and systematic loss of HX motif sequences is intriguing. This could argue in favor of the second scenario, whereby the HX motif would have appeared sporadically by convergent evolution in the different protein families. This second scenario does not exclude additional secondary losses, as observed in cnidarian Hox/ParaHox proteins (Table 1). Moreover, evolution by convergence is not atypical for short motifs in general (Van Roey et al., 2013), and has for example already been proposed for another motif widely found in ANTP-class members (including Gsx, En, Emx, and several NK) and other non-homeoproteins (Williams and Holland, 2000). In the case of the HX motif, it seemingly appeared later during evolution in bilaterian Tlx, Emx and En proteins (Table 1), suggesting a mechanism of convergent evolution. Of note, these bilaterian proteins are known to interact and/or participate with TALE cofactors in the context of tissue-specific functions (Brendolan et al., 2005; Capellini et al., 2010). Along the same line, an HX

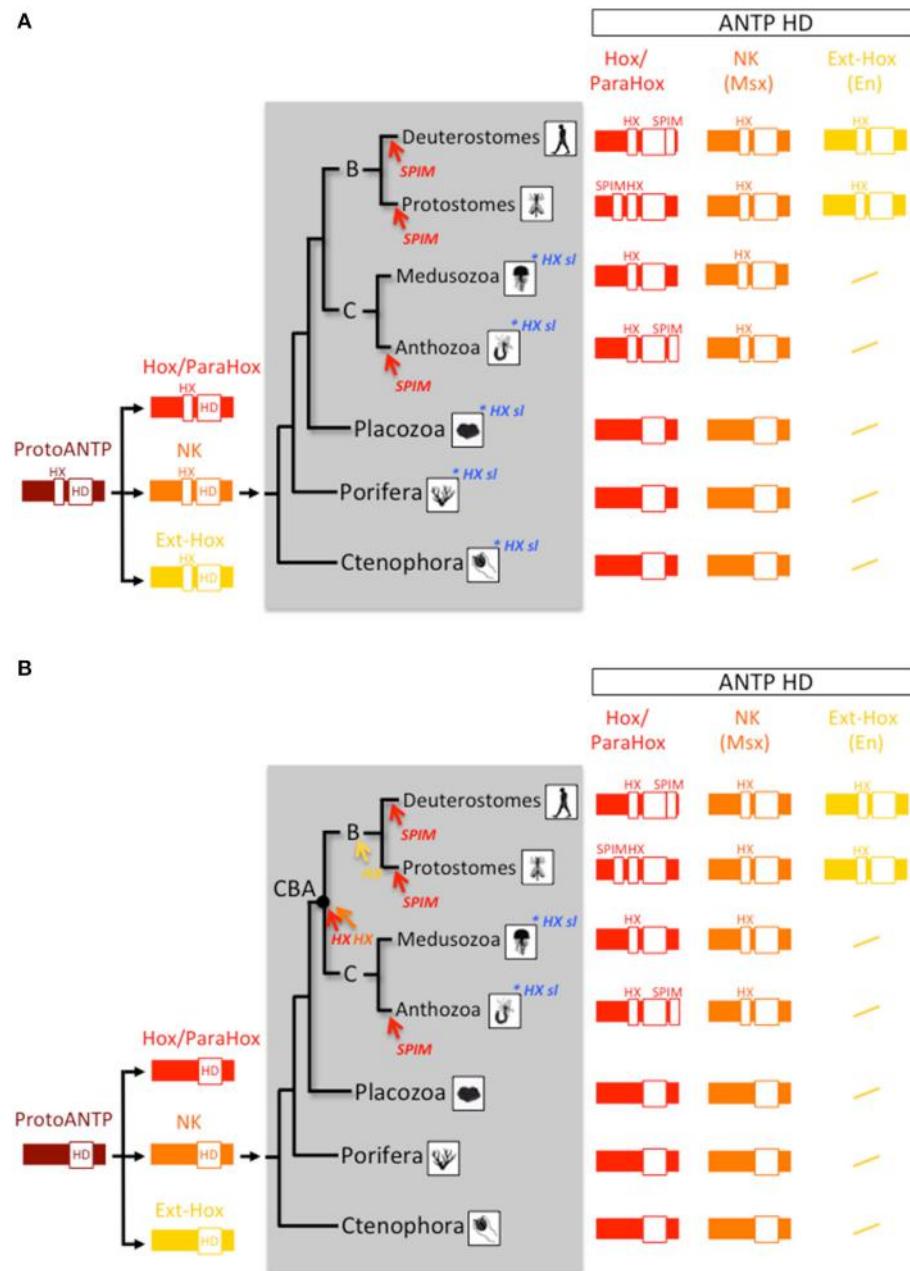


FIGURE 4 | Two evolutionary scenarios for the origin and early evolution of Hox-TALE interaction properties. **(A)** In the first scenario, the HX motif arose in the ProtoANTP gene of the metazoan LCA. There were multiple secondary losses in the Hox/ParaHox (sl; highlighted in blue) and other families (not indicated) in ctenophores, porifers, placozoans and cnidarians. **(B)** In the second scenario, the HX motif appeared independently several times during evolution, acquired in the Hox/ParaHox (red arrow), and NK (Msx, orange arrow) families of the last common ancestor of Cnidaria and

Bilateria (CBA), or in Ext-Hox members (as exemplified with En, yellow arrow) of the bilaterian ancestor. The absence of the HX motif in Hox/ParaHox members of several cnidarian species indicates secondary lost events (highlighted in blue; see also Table 1). In both scenarios, the HX motif served as a molecular template for diversifying TALE interaction properties only in the Hox/ParaHox family. This was achieved by the emergence of SPIMs. These motifs were independently acquired (highlighted in red) in Bilateria and Cnidaria, coinciding with strong morphological radiation in these two phyla.

motif is also present in non-ANTP class proteins, including LIM and several myogenic bHLH proteins (see In der Rieden et al., 2003, for a more complete list of HX-containing proteins). In the case of bHLH proteins the HX motif was further shown to be involved in the interaction and function with TALE cofactors

during skeletal muscle differentiation in vertebrates (Knoepfle et al., 1999; Maves et al., 2007, 2009; Yao et al., 2013). Together these observations highlight the strong evolutionary plasticity of the HX motif for providing a TALE-recruiting activity to highly divergent protein families.

Genesis of the Hox-TALE Patterning System during Metazoan Evolution

The evolutionarily conserved PBC-A and Meis-A domains in PBC and Meis proteins are restricted to Bilateria, Cnidaria and Placozoa, suggesting that a Hox/TALE network exists only in these three phyla (**Figure 1**). Like all cnidarian and bilaterian Gsx proteins, the unique ParaHox Gsx representative of *Trichoplax adhaerens* has no HX motif (**Table 1**) and cannot interact with PBC and Meis (Hudry et al., 2014). By contrast, the two other ParaHox and the Hox-related proteins have retained an HX motif in most cnidarians and bilaterians (**Table 1**). Thus, Cnidaria and Bilateria are the only phyla where a Hox-TALE interaction network is effective.

Since interaction with TALE proteins is not a specific feature of Hox proteins, the next question is “When did Hox proteins acquire their patterning functions linked to the interaction with the TALE cofactors?” We postulate here that the acquisition of differential patterning functions was tightly linked to the emergence of diversified interaction properties between Hox and TALE proteins. Then the question could be reformulated as: “When did alternative TALE interaction motifs appear in addition to the HX motif in the Hox/ParaHox family?”

Recent work with *Nematostella* Hox and TALE proteins (Hudry et al., 2014) suggests that SPIMs co-evolved with the specification of embryonic axes. As SPIMs are specific to a given Hox family or to a given species, they likely emerged independently several times during evolution (**Figure 4**). We propose that the original HX-dependent interaction mode served as an initial molecular template for experiencing these novel HX-independent interaction properties with the TALE partners. It is tempting to speculate that SPIMs were a molecular prerequisite for allowing Hox proteins to acquire patterning functions during early eumetazoan evolution. In this model, the acquisition of SPIMs in Hox proteins likely happened in parallel to mechanisms regulating their expression, allocating Hox genes to specific spatio-temporal domains along the longitudinal axis (**Figure 5**).

Finally, SPIMs do not necessarily correspond to related peptide sequences, as already noticed for the TDWM and UbdA motifs in *Drosophila* (Merabet and Hudry, 2013), making their identification difficult. Additional SPIMs need however to be identified to validate our model. Several tools are now available for predicting the presence of short interaction motifs in protein sequences, based on the analysis of amino acid chemical properties and the classification of hundreds of characterized short motifs in databases (Tompa et al., 2014). Interestingly, these tools predict a number of short motifs in several regions of bilaterian (Merabet and Dard, 2014) and cnidarian (Baëza et al., 2015) Hox proteins. These regions are often involved in the interaction with different TFs (Baëza et al., 2015), and could therefore contain good candidate SPIMs to test in the future.

Perspective: the HX Motif and SPIMs as Molecular Markers of Patterning Functions in the ParaHox Family?

ParaHox genes share several common features with the Hox genes. For example, they are organized in clusters and

display spatial-temporal constraints for their expression during embryogenesis of several bilaterian species (Garstang and Ferrier, 2013). The expression profile of ParaHox genes in Cnidaria is also reminiscent of important functions during embryogenesis, regeneration or budding, as seen in the solitary polyp *Hydra* (Schummer et al., 1992; Miljkovic-Licina et al., 2007), the coral *Acropora* (Hayward et al., 2001), the jellyfish *Podocoryne* (Yanze et al., 2001), the sea anemone *Nematostella* (Finnerty et al., 2003), or the colonial polyp *Hydractinia* (Cartwright et al., 2006). Moreover, ParaHox genes, and more particularly Gsx, could be more representative of the ProtoHox ancestor gene than any other Hox gene (Quiquand et al., 2009). Although Gsx does not contain any HX motif, it has a conserved role for the specification of neuroblast lineages in bilaterians (Weiss et al., 1998; Waclaw et al., 2009; Winterbottom et al., 2010; López-Juárez et al., 2013) and cnidarians, with a fine regulation along the body axis (Hayward et al., 2001; Miljkovic-Licina et al., 2007). Along the same line, Pdx-1 plays a crucial role in pancreatic beta-cell differentiation (Kaneto et al., 2007). These observations suggest that a primordial ParaHox (and Hox) function was dedicated to the emergence of novel cell types along the body axis, possibly in a TALE-independent manner (**Figure 5**) (De Jong et al., 2006; Miljkovic-Licina et al., 2007; Quiquand et al., 2009). This role could then have been deployed in several Hox/ParaHox members and in different tissues, requiring the acquisition of additional molecular features such as the HX motif and SLIMs for diversifying the novel patterning functions. In agreement with this hypothesis, in Bilateria Pdx/Xlox and Cdx transcription factors are required for the patterning of endodermal derivatives (Cole et al., 2009; Beck and Stringer, 2010; Annunziata et al., 2013; Ikuta et al., 2013) or during axis elongation with the Hox genes (Moreno and Morata, 1999; Van den Akker et al., 2002; Shinmyo et al., 2005; Young et al., 2009). The impact of TALE cofactors in those patterning functions remains to be investigated. The role of Pdx/Xlox and Cdx is also largely unknown in cnidarians. Testing their interaction properties with TALE cofactors could undoubtedly provide new insightful information into the origin and evolvability of the Hox/TALE patterning system in Metazoa. Ultimately, such studies should tell us whether the combination of one HX motif plus several SPIMs in the ParaHox proteins was necessary and sufficient to promote a spatial organization of cell differentiation along the body axis and thus the emergence of patterning functions in different tissues.

Conclusion

Hox proteins are TFs displaying highly similar DNA binding properties *in vitro*. Still, each Hox protein will dictate a specific developmental program with the same set of TALE cofactors. We proposed here that the apparition of a functional Hox/TALE patterning system during metazoan evolution was tightly linked to the acquisition of different short motifs named SPIMs. The usage of different SPIMs in Hox proteins constitutes an appealing molecular strategy for explaining the specific and various developmental functions of Hox/TALE complexes. Due to their small size, SPIMs present the advantage of being highly dynamic during evolution, allowing diversifying the molecular

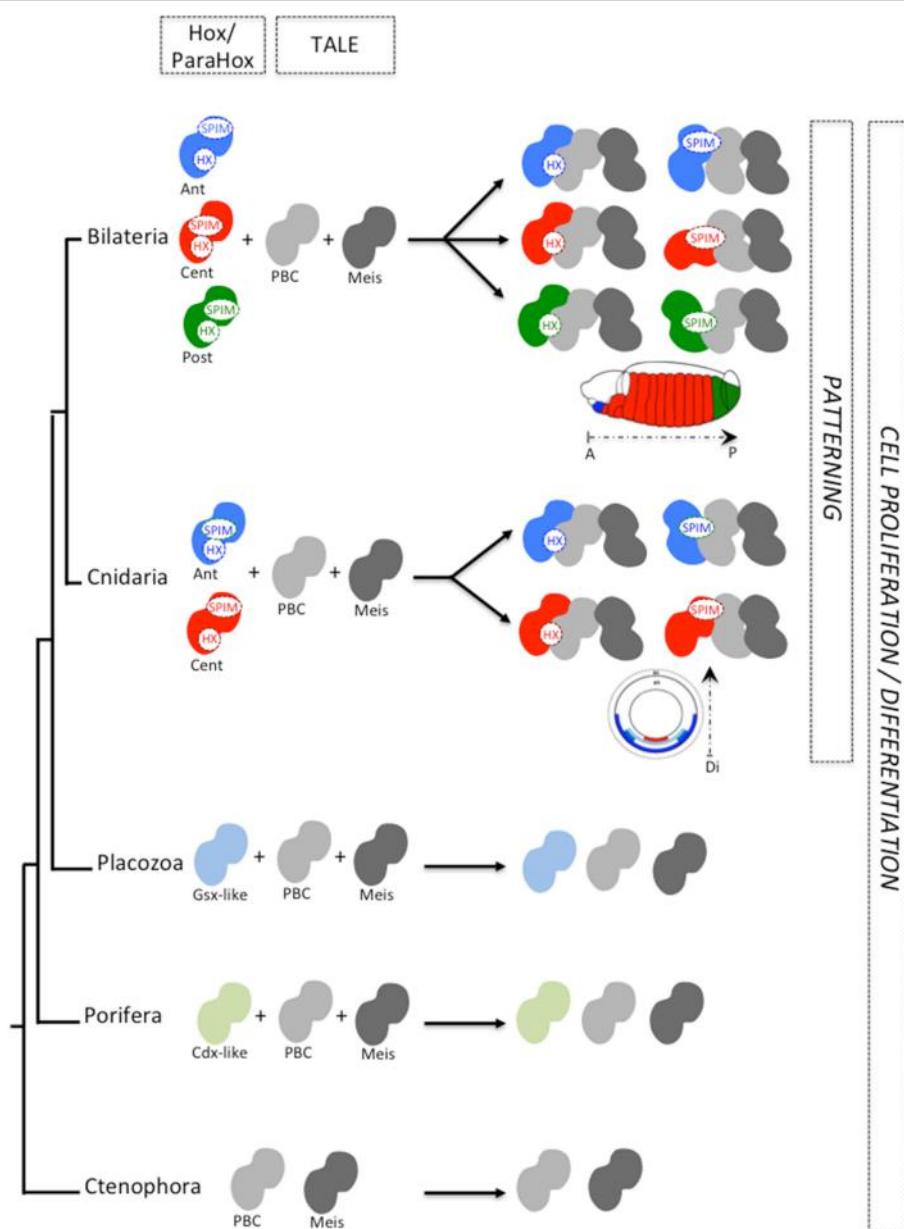


FIGURE 5 | SPIMs as molecular markers of a Hox/TALE patterning system during animal evolution. The acquisition of SPIMs in Cnidaria (*Nematostella*) and Bilateria (*Drosophila*, mammals) allowed Hox proteins to diversify their interaction modes with the TALE partners. This molecular diversification was essential for providing differential activities to Hox proteins along the longitudinal axis. Illustrative examples are provided along the anterior-posterior axis of the *Drosophila* embryo or along the directive axis of the *Nematostella* embryo. Anterior (ant), central (cent), and posterior (post) Hox proteins are depicted by a

different color. In Placozoa and Porifera, ParaHox-like members are present but these proteins do not contain any HX motif. Along the same line, placozoan and poriferan PBC and Meis representatives lack the PBC-A and MEIS-A domains (see Figure 1) and thus cannot interact together. As a consequence, TALE interaction networks do likely not exist in those two phyla. We postulate that Hox/ParaHox transcription factors were initially dedicated to cell proliferation/cell differentiation with no patterning function, whereas Hox/TALE interactions co-evolved with patterning functions.

code between Hox and TALE proteins. This model supposes that interaction flexibility is an important feature of the Hox/TALE patterning system. Whether this molecular strategy could more widely apply to other key patterning networks constitutes a major issue to investigate in the future.

Acknowledgments

We warmly thank Daniel Chourrout and Maja Adamska for helpful comments and criticisms on early versions of this review. Work in the laboratory of SM is supported by Association

Française contre les Myopathies (AFM), Fondation pour la recherché Médicale (FRM), Association pour la Recherche contre le Cancer (ARC), Ligue Régionale contre le Cancer, Centre National de la Recherche Scientifique (CNRS), and Ecole

Normale Supérieure (ENS) de Lyon. Work in the Galliot lab is supported by the State of Geneva, the Swiss National Fonds for Research (SNF-31003A-149630). We apologize to colleagues whose work was not cited due to space constraints.

References

- Abe, N., Dror, I., Yang, L., Slattery, M., Zhou, T., Bussemaker, H. J., et al. (2015). Deconvolving the recognition of DNA shape from sequence. *Cell* 161, 307–318. doi: 10.1016/j.cell.2015.02.008
- Annunziata, R., Martinez, P., and Arnone, M. I. (2013). Intact cluster and chordate-like expression of ParaHox genes in a sea star. *BMC Biol.* 11:68. doi: 10.1186/1741-7007-11-68
- Arthur, W. (2002). The emerging conceptual framework of evolutionary developmental biology. *Nature* 415, 757–764. doi: 10.1038/415757a
- Baëza, M., Viala, S., Heim, M., Dard, A., Hudry, B., Duffraisse, M., et al. (2015). Inhibitory activities of short linear motifs underlie Hox interactome specificity *in vivo*. *Elife* 4, 1–28. doi: 10.7554/elife.06034
- Balavoine, G., de Rosa, R., and Adoutte, A. (2002). Hox clusters and bilaterian phylogeny. *Mol. Phylogenet. Evol.* 24, 366–373. doi: 10.1016/S1055-7903(02)00237-3
- Beck, F., and Stringer, E. J. (2010). The role of *Cdx* genes in the gut and in axial development. *Biochem. Soc. Trans.* 38, 353–357. doi: 10.1042/BST0380353
- Bellaouï, M., Pidkowich, M. S., Samach, A., Kushalappa, K., Kohalmi, S. E., Modrusan, Z., et al. (2001). The *Arabidopsis* *BELL1* and *KNOX* TALE homeodomain proteins interact through a domain conserved between plants and animals. *Plant Cell* 13, 2455–2470. doi: 10.1105/tpc.13.11.2455
- Brendolan, A., Ferretti, E., Salsi, V., Moses, K., Quaggin, S., Blasi, F., et al. (2005). A *Pbx1*-dependent genetic and transcriptional network regulates spleen ontogeny. *Development* 132, 3113–3126. doi: 10.1242/dev.01884
- Brooke, N. M., García-Fernández, J., and Holland, P. W. (1998). The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. *Nature* 392, 920–922. doi: 10.1038/31933
- Bürglin, T. R. (1997). Analysis of TALE superclass homeobox genes (MEIS, PBC, KNOX, Iroquois, TGIF) reveals a novel domain conserved between plants and animals. *Nucleic Acids Res.* 25, 4173–4180. doi: 10.1093/nar/25.21.4173
- Bürglin, T. R. (1998). The PBC domain contains a MEINOX domain: coevolution of Hox and TALE homeobox genes? *Dev. Genes Evol.* 208, 113–116. doi: 10.1007/s004270050161
- Capellini, T. D., Vaccari, G., Ferretti, E., Fantini, S., He, M., Pellegrini, M., et al. (2010). Scapula development is governed by genetic interactions of *Pbx1* with its family members and with *Emx2* via their cooperative control of *Alx1*. *Development* 137, 2559–2569. doi: 10.1242/dev.048819
- Carr, E. A., Mead, J., and Vershon, A. K. (2004). Alpha1-induced DNA bending is required for transcriptional activation by the Mcm1-alpha1 complex. *Nucleic Acids Res.* 32, 2298–2305. doi: 10.1093/nar/gkh560
- Cartwright, P., Schierwater, B., and Buss, L. W. (2006). Expression of a *Gsx* parahox gene, *Cnox-2*, in colony ontogeny in *Hydractinia* (Cnidaria: Hydrozoa). *J. Exp. Zool. Part B Mol. Dev. Evol.* 306, 460–469. doi: 10.1002/jez.b.21106
- Chiori, R., Jager, M., Denker, E., Wincker, P., Da Silva, C., Le Guyader, H., et al. (2009). Are Hox genes ancestrally involved in axial patterning? Evidence from the hydrozoan *Clytia hemisphaerica* (cnidaria). *PLoS ONE* 4:e4231. doi: 10.1371/journal.pone.0004231
- Chourrout, D., Delsuc, F., Chourrout, P., Edvardsen, R. B., Rentzsch, F., Renfer, E., et al. (2006). Minimal ProtoHox cluster inferred from bilaterian and cnidarian Hox complements. *Nature* 442, 684–687. doi: 10.1038/nature04863
- Clarke, M., Lohan, A. J., Liu, B., Lagkouvardos, I., Roy, S., Zafar, N., et al. (2013). Genome of *Acanthamoeba castellanii* highlights extensive lateral gene transfer and early evolution of tyrosine kinase signaling. *Genome Biol.* 14:R11. doi: 10.1186/gb-2013-14-2-r11
- Cole, A. G., Rizzo, F., Martinez, P., Fernandez-Serra, M., and Arnone, M. I. (2009). Two ParaHox genes, *SpLox* and *SpCdx*, interact to partition the posterior endoderm in the formation of a functional gut. *Development* 136, 541–549. doi: 10.1242/dev.029959
- Davidson, E. H., and Erwin, D. H. (2006). Gene regulatory networks and the evolution of animal body plans. *Science* 311, 796–800. doi: 10.1126/science.1113832
- de Jong, D. M., Hislop, N. R., Hayward, D. C., Reece-Hoyes, J. S., Pontynen, P. C., Ball, E. E., et al. (2006). Components of both major axial patterning systems of the Bilateria are differentially expressed along the primary axis of a “radiate” animal, the anthozoan cnidarian *Acropora millepora*. *Dev. Biol.* 298, 632–643. doi: 10.1016/j.ydbio.2006.07.034
- Duboule, D. (2007). The rise and fall of Hox gene clusters. *Development* 134, 2549–2560. doi: 10.1242/dev.001065
- Erwin, D. H. (2009). Early origin of the bilaterian developmental toolkit. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 364, 2253–2261. doi: 10.1098/rstb.2009.0038
- Finnerty, J. R., Pang, K., Burton, P., Paulson, D., and Martindale, M. Q. (2004). Origins of bilateral symmetry: Hox and dpp expression in a sea anemone. *Science* 304, 1335–1337. doi: 10.1126/science.1091946
- Finnerty, J. R., Paulson, D., Burton, P., Pang, K., and Martindale, M. Q. (2003). Early evolution of a homeobox gene: the parahox gene *Gsx* in the Cnidaria and the Bilateria. *Evol. Dev.* 5, 331–345. doi: 10.1046/j.1525-142X.2003.03041.x
- Foos, N., Maurel-Zaffran, C., Maté, M. J., Vincentelli, R., Hainaut, M., Berenger, H., et al. (2015). A flexible extension of the *Drosophila* ultrabithorax homeodomain defines a novel Hox/PBC interaction mode. *Structure* 23, 270–279. doi: 10.1016/j.str.2014.12.011
- Fortunato, S. A. V., Adamski, M., Ramos, O. M., Leininger, S., Liu, J., Ferrier, D. E. K., et al. (2014). Calcisponges have a ParaHox gene and dynamic expression of dispersed NK homeobox genes. *Nature* 514, 620–623. doi: 10.1038/nature13881
- Fujioka, M., Geblein, B., Cofer, Z. C., Mann, R. S., and Jaynes, J. B. (2012). Engrailed cooperates directly with Extrudenticle and Homothorax on a distinct class of homeodomain binding sites to repress sloppy paired. *Dev. Biol.* 366, 382–392. doi: 10.1016/j.ydbio.2012.04.004
- Galant, R., Walsh, C. M., and Carroll, S. B. (2002). Hox repression of a target gene: extrudenticle-independent, additive action through multiple monomer binding sites. *Development* 129, 3115–3126.
- Galliot, B., and de Vargas, C., Miller, D. (1999). Evolution of homeobox genes: *Q50* Paired-like genes founded the Paired class. *Dev. Genes Evol.* 209, 186–197. doi: 10.1007/s004270050243
- Garcia-Fernández, J. (2005). The genesis and evolution of homeobox gene clusters. *Nat. Rev. Genet.* 6, 881–892. doi: 10.1038/nrg1723
- Garstang, M., and Ferrier, D. E. K. (2013). Time is of the essence for ParaHox homeobox gene clustering. *BMC Biol.* 11:72. doi: 10.1186/1741-707-11-72
- Gauchat, D., Mazet, F., Berney, C., Schummer, M., Kreger, S., Pawłowski, J., et al. (2000). Evolution of Antp-class genes and differential expression of Hydra Hox/paraHox genes in anterior patterning. *Proc. Natl. Acad. Sci. U.S.A.* 97, 4493–4498. doi: 10.1073/pnas.97.9.4493
- Gould, S. J. (1992). Ontogeny and phylogeny—revisited and reunited. *Bioessays* 4, 275–279. doi: 10.1002/bies.950140413
- Hackbusch, J., Richter, K., Müller, J., Salamini, F., and Uhrig, J. F. (2005). A central role of *Arabidopsis thaliana* ovate family proteins in networking and subcellular localization of 3-aa loop extension homeodomain proteins. *Proc. Natl. Acad. Sci. U.S.A.* 102, 4908–4912. doi: 10.1073/pnas.0501181102
- Hay, A., and Tsiantis, M. (2010). KNOX genes: versatile regulators of plant development and diversity. *Development* 137, 3153–3165. doi: 10.1242/dev.030049
- Hayward, D. C., Catmull, J., Reece-Hoyes, J. S., Berghammer, H., Dodd, H., Hann, S. J., et al. (2001). Gene structure and larval expression of *cnox-2Am* from the coral *Acropora millepora*. *Dev. Genes Evol.* 211, 10–19. doi: 10.1007/s004270000112
- Heffer, A., Xiang, J., and Pick, L. (2013). Variation and constraint in *Hox* gene evolution. *Proc. Natl. Acad. Sci. U.S.A.* 110, 2211–2216. doi: 10.1073/pnas.1210847110

- Holland, P. W. H. (2013). Evolution of homeobox genes. *Wiley Interdiscip. Rev. Dev. Biol.* 2, 31–45. doi: 10.1002/wdev.78
- Hudry, B., Remacle, S., Delfini, M.-C., Rezsöhazy, R., Graba, Y., and Merabet, S. (2012). Hox proteins display a common and ancestral ability to diversify their interaction mode with the PBC class cofactors. *PLoS Biol.* 10:e1001351. doi: 10.1371/journal.pbio.1001351
- Hudry, B., Thomas-Chollier, M., Volovik, Y., Duffraisse, M., Dard, A., Frank, D., et al. (2014). Molecular insights into the origin of the Hox-TALE patterning system. *Elife* 3:e01939. doi: 10.7554/elife.01939
- Hudry, B., Viala, S., Graba, Y., and Merabet, S. (2011). Visualization of protein interactions in living *Drosophila* embryos by the bimolecular fluorescence complementation assay. *BMC Biol.* 9:5. doi: 10.1186/1741-7007-9-5
- Hueber, S. D., and Lohmann, I. (2008). Shaping segments: *Hox* gene function in the genomic age. *Bioessays* 30, 965–979. doi: 10.1002/bies.20823
- Ikuta, T., Chen, Y.-C., Annunziata, R., Ting, H.-C., Tung, C. H., Koyanagi, R., et al. (2013). Identification of an intact ParaHox cluster with temporal colinearity but altered spatial colinearity in the hemichordate *Ptychoderia flava*. *BMC Evol. Biol.* 13:129. doi: 10.1186/1471-2148-13-129
- In der Rieden, P. M., der, Mainguy, G., Woltering, J. M., and Durston, A. J. (2003). Homeodomain to hexapeptide or PBC-interactiondomain distance: size apparently matters. *Trends Genet.* 20, 76–79. doi: 10.1016/j.tig.2003.12.001
- Joshi, R., Passner, J. M., Rohs, R., Jain, R., Sosinsky, A., Crickmore, M., et al. (2007). Functional specificity of a Hox protein mediated by the recognition of minor groove structure. *Cell* 131, 530–543. doi: 10.1016/j.cell.2007.09.024
- Kamm, K., Schierwater, B., Jakob, W., Dellaporta, S. L., and Miller, D. J. (2006). Axial patterning and diversification in the cnidaria predate the Hox system. *Curr. Biol.* 16, 920–926. doi: 10.1016/j.cub.2006.03.036
- Kaneto, H., Miyatsuka, T., Kawamori, D., and Matsuoka, T. A. (2007). Pleiotropic roles of PDX-1 in the pancreas. *Rev. Diabet. Stud.* 4, 209–225. doi: 10.1900/RDS.2007.4.209
- Kanrar, S., Onguka, O., and Smith, H. M. S. (2006). Arabidopsis inflorescence architecture requires the activities of KNOX-BELL homeodomain heterodimers. *Planta* 224, 1163–1173. doi: 10.1007/s00425-006-0298-9
- Keleher, C. A., Passmore, S., and Johnson, A. D. (1989). Yeast repressor alpha 2 binds to its operator cooperatively with yeast protein Mcm1. *Mol. Cell. Biol.* 9, 5228–5230.
- King, N. (2004). The unicellular ancestry of animal development. *Dev. Cell* 7, 313–325. doi: 10.1016/j.devcel.2004.08.010
- King, N., Hittinger, C. T., and Carroll, S. B. (2003). Evolution of key cell signaling and adhesion protein families predates animal origins. *Science* 301, 361–363. doi: 10.1126/science.1083853
- King, N., Westbrook, M. J., Young, S. L., Kuo, A., Abedin, M., Chapman, J., et al. (2008). The genome of the choanoflagellate *Monosiga brevicollis* and the origin of metazoans. *Nature* 451, 783–788. doi: 10.1038/nature06617
- Kmita, M., and Duboule, D. (2003). Organizing axes in time and space; 25 years of colinear tinkering. *Science* 301, 331–333. doi: 10.1126/science.1085753
- Knoepfle, P. S., Bergstrom, D. A., Uetsuki, T., Dac-kortyko, I., Sun, Y. H., Wright, W. E., et al. (1999). A conserved motif N-terminal to the DNA-binding domains of myogenic bHLH transcription factors mediates cooperative DNA binding with Pbx – Meis1/Prep1. *27*, 3752–3761. doi: 10.1093/nar/27.18.3752
- LaRonde-LeBlanc, N. A., and Wolberger, C. (2003). Structure of HoxA9 and Pbx1 bound to DNA: Hox hexapeptide and DNA recognition anterior to posterior. *Genes Dev.* 17, 2060–2072. doi: 10.1101/gad.1103303
- Larroux, C., Fahey, B., Degnan, S. M., Adamski, M., Rokhsar, D. S., and Degnan, B. M. (2007). The NK Homeobox Gene Cluster Predates the Origin of Hox Genes. *Curr. Biol.* 17, 706–710. doi: 10.1016/j.cub.2007.03.008
- Larroux, C., Luke, G. N., Koopman, P., Rokhsar, D. S., Shimeld, S. M., and Degnan, B. M. (2008). Genesis and expansion of metazoan transcription factor gene classes. *Mol. Biol. Evol.* 25, 980–996. doi: 10.1093/molbev/msn047
- Lewis, E. B. (1978). A gene complex controlling segmentation in *Drosophila*. *Nature* 276, 565–570. doi: 10.1038/276565a0
- López-Juárez, A., Howard, J., Ullom, K., Howard, L., Grande, A., Pardo, A., et al. (2013). Gsx2 controls region-specific activation of neural stem cells and injury-induced neurogenesis in the adult subventricular zone. *Genes Dev.* 27, 1272–1287. doi: 10.1101/gad.217539.113
- Mann, R. S., and Affolter, M. (1998). Hox proteins meet more partners. *Curr. Opin. Genet. Dev.* doi: 10.1016/s0959-437x(98)80113-5
- Mann, R. S., and Carroll, S. B. (2002). Molecular mechanisms of selector gene function and evolution. *Curr. Opin. Genet. Dev.* 12, 592–600. doi: 10.1016/S0959-437X(02)00344-1
- Mann, R. S., Lelli, K. M., and Joshi, R. (2009). Chapter 3 Hox specificity. Unique roles for cofactors and collaborators. *Curr. Top. Dev. Biol.* 88, 63–101. doi: 10.1016/S0070-2153(09)88003-4
- Martindale, M. Q. (2005). The evolution of metazoan axial properties. *Nat. Rev. Genet.* 6, 917–927. doi: 10.1038/nrg1725
- Matus, D. Q., Pang, K., Marlow, H., Dunn, C. W., Thomsen, G. H., and Martindale, M. Q. (2006). Molecular evidence for deep evolutionary roots of bilaterality in animal development. *Proc. Natl. Acad. Sci. U.S.A.* 103, 11195–11200. doi: 10.1073/pnas.0601257103
- Maves, L., Tyler, A., Moens, C. B., and Tapscott, S. J. (2009). Pbx acts with Hand2 in early myocardial differentiation. *Dev. Biol.* 333, 409–418. doi: 10.1016/j.ydbio.2009.07.004
- Maves, L., Waskiewicz, A. J., Paul, B., Cao, Y., Tyler, A., Moens, C. B., et al. (2007). Pbx homeodomain proteins direct Myod activity to promote fast-muscle differentiation. *Development* 134, 3371–3382. doi: 10.1242/dev.003905
- McGinnis, W., Krumlauf, R. (1992). Homeobox genes and axial patterning. *Cell* 68, 283–302. doi: 10.1016/0092-8674(92)90471-N
- Mendivil Ramos, O., Barker, D., and Ferrier, D. E. K. (2012). Ghost loci imply Hox and parahox existence in the last common ancestor of animals. *Curr. Biol.* 22, 1951–1956. doi: 10.1016/j.cub.2012.08.023
- Merabet, S., and Dard, A. (2014). Tracking context-specific transcription factors regulating Hox activity. *Dev. Dyn.* 243, 16–23. doi: 10.1002/dvdy.24402
- Merabet, S., and Hudry, B. (2013). Hox transcriptional specificity despite a single class of cofactors: are flexible interaction modes the key? Plasticity in Hox/PBC interaction modes as a common molecular strategy for shaping Hox transcriptional activities. *Bioessays* 35, 88–92. doi: 10.1002/bies.201200146
- Merabet, S., Hudry, B., Saadaoui, M., and Graba, Y. (2009). Classification of sequence signatures: a guide to Hox protein function. *Bioessays* 31, 500–511. doi: 10.1002/bies.200800229
- Merabet, S., Kambris, Z., Capovilla, M., Bérenger, H., Pradel, J., and Graba, Y. (2003). The hexapeptide and linker regions of the AbdA Hox protein regulate its activating and repressive functions. *Dev. Cell* 4, 761–768. doi: 10.1016/S1534-5807(03)00126-6
- Merabet, S., Litim-Mecheri, I., Karlsson, D., Dixit, R., Saadaoui, M., Monier, B., et al. (2011). Insights into Hox protein function from a large scale combinatorial analysis of protein domains. *PLoS Genet.* 7:e1002302. doi: 10.1371/journal.pgen.1002302
- Merabet, S., Saadaoui, M., Sambrani, N., Hudry, B., Pradel, J., Affolter, M., et al. (2007). A unique Extradenticle recruitment mode in the *Drosophila* Hox protein Ultrabithorax. *Proc. Natl. Acad. Sci. U.S.A.* 104, 16946–16951. doi: 10.1073/pnas.0705832104
- Miljkovic-Licina, M., Chera, S., Ghila, L., and Galliot, B. (2007). Head regeneration in wild-type hydra requires de novo neurogenesis. *Development* 134, 1191–1201. doi: 10.1242/dev.02804
- Minguillón, C., and García-Fernández, J. (2003). Genesis and evolution of the *Evx* and *Mox* genes and the extended Hox and ParaHox gene clusters. *Genome Biol.* 4:R12. doi: 10.1186/gb-2003-4-2-r12
- Moens, C. B., and Selleri, L. (2006). Hox cofactors in vertebrate development. *Dev. Dyn.* 291, 193–206. doi: 10.1016/j.ydbio.2005.10.032
- Moreno, E., and Morata, G. (1999). *Caudal* is the *Hox* gene that specifies the most posterior *Drosophila* segment. *Nature* 400, 873–877. doi: 10.1038/23709
- Moroz, L. L., Kocot, K. M., Citarella, M. R., Dosung, S., Norekian, T. P., Povolotskaya, I. S., et al. (2014). The ctenophore genome and the evolutionary origins of neural systems. *Nature* 510, 109–114. doi: 10.1038/nature13400
- Mukherjee, K., and Bürglin, T. R. (2007). Comprehensive analysis of animal TALE homeobox genes: new conserved motifs and cases of accelerated evolution. *J. Mol. Evol.* 65, 137–153. doi: 10.1007/s00239-006-0023-0
- Ogishima, S., and Tanaka, H. (2007). Missing link in the evolution of *Hox* clusters. *Gene* 387, 21–30. doi: 10.1016/j.gene.2006.08.011
- Passner, J. M., Ryoo, H. D., Shen, L., Mann, R. S., and Aggarwal, A. K. (1999). Structure of a DNA-bound Ultrabithorax–Extradenticle homeodomain complex. *Nature* 397, 714–719.

- Pearson, J. C., Lemons, D., and McGinnis, W. (2005). Modulating Hox gene functions during animal body patterning. *Nat. Rev. Genet.* 6, 893–904. doi: 10.1038/nrg1726
- Piper, D. E., Batchelor, A. H., Chang, C. P., Cleary, M. L., and Wolberger, C. (1999). Structure of a HoxB1–Pbx1 heterodimer bound to DNA: role of the hexapeptide and a fourth homeodomain helix in complex formation. *Cell* 96, 587–97. doi: 10.1016/S0092-8674(00)80662-5
- Quiquand, M., Yanze, N., Schmich, J., Schmid, V., Galliot, B., and Piraino, S. (2009). More constraint on ParaHox than Hox gene families in early metazoan evolution. *Dev. Biol.* 328, 173–187. doi: 10.1016/j.ydbio.2009.01.022
- Rhee, J. W., Arata, A., Selleri, L., Jacobs, Y., Arata, S., Onimaru, H., et al. (2004). Pbx3 deficiency results in central hypoventilation. *Am. J. Pathol.* 165, 1343–1350. doi: 10.1016/S0002-9440(10)63392-5
- Ryan, J. F., and Baxevanis, A. D. (2007). Hox, Wnt, and the evolution of the primary body axis: insights from the early-divergent phyla. *Biol. Direct* 2, 37–48. doi: 10.1186/1745-6150-2-37
- Ryan, J. F., Mazza, M. E., Pang, K., Matus, D. Q., Baxevanis, A. D., Martindale, M. Q., et al. (2007). Pre-bilaterian origins of the Hox cluster and the Hox code: evidence from the sea anemone, *Nematostella vectensis*. *PLoS ONE* 2:e153. doi: 10.1371/journal.pone.0000153
- Ryan, J. F., Pang, K., Mullikin, J. C., Martindale, M. Q., and Baxevanis, A. D. (2010). The homeodomain complement of the ctenophore *Mnemiopsis leidyi* suggests that Ctenophora and Porifera diverged prior to the ParaHoxozoa. *Evodevo* 1:9. doi: 10.1186/2041-9139-1-9
- Saleh, M., Rambaldi, I., Yang, X. J., and Featherstone, M. S. (2000). Cell signaling switches HOX-PBX complexes from repressors to activators of transcription mediated by histone deacetylases and histone acetyltransferases. *Mol. Cell. Biol.* 20, 8623–33. doi: 10.1128/MCB.20.22.8623-8633.2000
- Schierwater, B., Kamm, K., Srivastava, M., Rokhsar, D., Rosengarten, R. D., and Dellaporta, S. L. (2008a). The early ANTP gene repertoire: insights from the placozoan genome. *PLoS ONE* 3:e2457. doi: 10.1371/journal.pone.0002457
- Schierwater, B., Kamm, K., Srivastava, M., Rokhsar, D., Rosengarten, R. D., and Dellaporta, S. L. (2008b). The early ANTP gene repertoire: insights from the placozoan genome. *PLoS ONE* 3:e2457. doi: 10.1371/journal.pone.0002457
- Schierwater, B., Kamm, K., Srivastava, M., Rokhsar, D., Rosengarten, R. D., and Dellaporta, S. L. (1998). Homology of Hox genes and the zootype concept in early metazoan evolution. *Mol. Phylogenet. Evol.* 9, 375–381. doi: 10.1006/mpev.1998.0489
- Schummer, M., Scheurlen, I., Schaller, C., and Galliot, B. (1992). HOM/HOX homeobox genes are present in hydra (*Chlorohydra viridissima*) and are differentially expressed during regeneration. *EMBO J.* 11, 1815–1823.
- Shen, W. F., Montgomery, J. C., Rozenfeld, S., Moskow, J. J., Lawrence, H. J., Buchberg, A. M., et al. (1997). AbdB-like Hox proteins stabilize DNA binding by the Meis1 homeodomain proteins. *Mol. Cell. Biol.* 17, 6448–58.
- Shinmyo, Y., Mito, T., Matsushita, T., Sarashina, I., Miyawaki, K., Ohuchi, H., et al. (2005). *Caudal* is required for gnathal and thoracic patterning and for posterior elongation in the intermediate-germband cricket *Gryllus bimaculatus*. *Mech. Dev.* 122, 231–239. doi: 10.1016/j.mod.2004.10.001
- Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcalá, P., Dror, I., et al. (2011). Cofactor binding evokes latent differences in DNA binding specificity between Hox proteins. *Cell* 147, 1270–1282. doi: 10.1016/j.cell.2011.10.053
- Srivastava, M., Simakov, O., Chapman, J., Fahey, B., Gauthier, M. E. A., Mitros, T., et al. (2010). The *Amphimedon queenslandica* genome and the evolution of animal complexity. *Nature* 466, 720–726. doi: 10.1038/nature09201
- Stark, M. R., and Johnson, A. D. (1994). Interaction between two homeodomain proteins is specified by a short C-terminal tail. *Nature* 371, 429–432. doi: 10.1038/371429a0
- Steele, R. E., David, C. N., and Technau, U. (2011). A genomic view of 500 million years of cnidarian evolution. *Trends Genet.* 27, 7–13. doi: 10.1016/j.tig.2010.10.002
- Stevens, K. E., and Mann, R. S. (2007). A balance between two nuclear localization sequences and a nuclear export sequence governs extradenticle subcellular localization. *Genetics* 175, 1625–1636. doi: 10.1534/genetics.106.066449
- Suga, H., Chen, Z., de Mendoza, A., Sebé-Pedrós, A., Brown, M. W., Kramer, E., et al. (2013). The *Capsaspora* genome reveals a complex unicellular prehistory of animals. *Nat. Commun.* 4, 2325. doi: 10.1038/ncomms3325
- Tompa, P., Davey, N. E., Gibson, T. J., and Babu, M. M. (2014). A million peptide motifs for the molecular biologist. *Mol. Cell* 55, 161–169. doi: 10.1016/j.molcel.2014.05.032
- True, J. R., and Carroll, S. B. (2002). Gene co-option in physiological and morphological evolution. *Annu. Rev. Cell Dev. Biol.* 18, 53–80. doi: 10.1146/annurev.cellbio.18.020402.140619
- Van den Akker, E., Forlani, S., Chawengsaksophak, K., de Graaff, W., Beck, F., Meyer, B. I., et al. (2002). *Cdx1* and *Cdx2* have overlapping functions in anteroposterior patterning and posterior axis elongation. *Development* 129, 2181–2193.
- Van Roey, K., Dinkel, H., Weatheritt, R. J., Gibson, T. J., and Davey, N. E. (2013). The switches.ELM resource: a compendium of conditional regulatory interaction interfaces. *Sci. Signal.* 6, rs7. doi: 10.1126/scisignal.2003345
- Waclaw, R. R., Wang, B., Pei, Z., Ehrman, L. A., and Campbell, K. (2009). Distinct temporal requirements for the homeobox gene *Gsx2* in specifying striatal and olfactory bulb neuronal fates. *Neuron* 63, 451–465. doi: 10.1016/j.neuron.2009.07.015
- Wagner, G. P. (2007). The developmental genetics of homology. *Nat. Rev. Genet.* 8, 473–479. doi: 10.1038/nrg2099
- Weiss, J. B., Von Ohlen, T., Mellerick, D. M., Dressler, G., Doe, C. Q., and Scott, M. P. (1998). Dorsoventral patterning in the *Drosophila* central nervous system: the intermediate neuroblasts defective homeobox gene specifies intermediate column identity. *Genes Dev.* 12, 3591–3602. doi: 10.1101/gad.12.22.3591
- Wenger, Y., and Galliot, B. (2013). Punctuated emergences of genetic and phenotypic innovations in eumetazoan, bilaterian, euteleostome, and hominidae ancestors. *Genome Biol. Evol.* 5, 1949–1968. doi: 10.1093/gbe/evt142
- Williams, N. A., and Holland, P. W. (2000). An amphioxus *Emx* homeobox gene reveals duplication during vertebrate evolution. *Mol. Biol. Evol.* 17, 1520–1528. doi: 10.1093/oxfordjournals.molbev.a026251
- Williams, T. M., Williams, M. E., and Innis, J. W. (2005). Range of HOX/TALE superklass associations and protein domain requirements for HOXA13:MEIS interaction. *Dev. Biol.* 277, 457–471. doi: 10.1016/j.ydbio.2004.10.004
- Winterbottom, E. F., Illes, J. C., Faas, L., and Isaacs, H., V (2010). Conserved and novel roles for the *Gsh2* transcription factor in primary neurogenesis. *Development* 137, 2623–2631. doi: 10.1242/dev.047159
- Yanze, N., Spring, J., Schmidli, C., and Schmid, V. (2001). Conservation of Hox/ParaHox-related genes in the early development of a cnidarian. *Dev. Biol.* 236, 89–98. doi: 10.1006/dbio.2001.0299
- Yao, Z., Farr, G. H., Tapscott, S. J., and Maves, L. (2013). Pbx and Prdm1a transcription factors differentially regulate subsets of the fast skeletal muscle program in zebrafish. *Biol. Open* 2, 546–555. doi: 10.1242/bio.20133921
- Young, T., Rowland, J. E., van de Ven, C., Bialecka, M., Novoa, A., Carapuco, M., et al. (2009). *Cdx* and *Hox* genes differentially regulate posterior axial growth in mammalian embryos. *Dev. Cell* 17, 516–526. doi: 10.1016/j.devcel.2009.08.010

Conflict of Interest Statement: 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 © 2015 Merabet and Galliot. 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) or licensor 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.



A stranger in a strange land: the utility and interpretation of heterologous expression

Elena M. Kramer*

Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA, USA

OPEN ACCESS

Edited by:

Sylvain Marcellini,
University of Concepcion, Chile

Reviewed by:

Verónica S. Di Stilio,
University of Washington, USA

Kacy Gordon,
Duke University, USA

***Correspondence:**

Elena M. Kramer,
Department of Organismic
and Evolutionary Biology,
Harvard University, 16 Divinity Avenue,
Cambridge, MA 02138, USA
ekramer@oeb.harvard.edu

Specialty section:

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

Received: 30 June 2015

Accepted: 29 August 2015

Published: 15 September 2015

Citation:

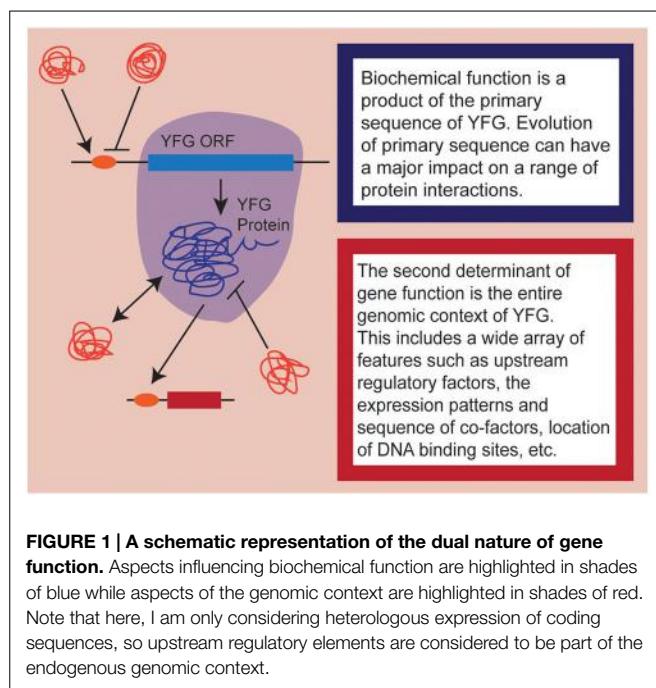
Kramer EM (2015) A stranger
in a strange land: the utility
and interpretation of heterologous
expression.
Front. Plant Sci. 6:734.
doi: 10.3389/fpls.2015.00734

One of the major goals of the modern study of evodevo is to understand the evolution of gene function across a range of contexts, including sub/neofunctionalization, co-option of genetic modules, and the evolution of morphological novelty. To these ends, comparative studies of gene expression can be useful for constructing hypotheses, but cannot provide direct evidence of functional evolution. Unfortunately, determining endogenous gene function in non-model species is often not an option. Faced with this dilemma, a common approach is to use heterologous expression (HE) in genetically tractable model species as a proxy for functional analyses. Such experiments have important limitations, however, and require caution in the interpretation of their results. How do we dissociate biochemical function from its original genomic context? In the end, what does HE actually tell us? Here, I argue that HE only sheds light on specific types of biochemical conservation, but can be useful when experiments are carefully interpreted.

Keywords: evo-devo, heterologous expression, functional evolution, biochemical evolution, developmental genetics

As developmental biologists, it is important to remember that when we speak of “gene function,” we are conflating, by necessity, a complex array of different factors. At a fundamental level, we can think of gene function as representing two complementary components: the first being biochemical function and the second being developmental role (**Figure 1**). The former is determined by the coding sequence of the gene itself and encompasses everything from secondary/tertiary protein structure, to enzymatic capacity, to co-factor and/or DNA binding site affinity. These aspects of gene function may change as the sequence of your favorite gene (YFG) itself evolves. As if this weren’t complicated enough, the actual developmental role played by YFG is the product of all of these primary components interacting with a wide array of *cis*- and *trans*-acting phenomena, including the expression patterns of YFG in relation to its co-factors, the epigenetic state of target loci, the position of binding sites within the genome, post-translation regulation of all interacting proteins, etc. Obviously, these secondary components evolve as well, to varying degrees in a coordinated fashion with YFG. So when we talk about the evolution of gene function, we are really considering the evolution of the whole genomic context of YFG—its protein sequence, *cis*- and *trans*-regulation, interacting partners, and target gene repertoire.

Heterologous expression (HE) takes the primary component of gene function—the sequence of the coding region itself—and plugs it into the second component—the genomic context—of a different species. We are essentially performing a site-directed mutagenesis experiment in which we ask whether the sequence differences between YFG and its endogenous homolog disrupt the functional roles normally played by the endogenous locus in its own genomic environment. Of course, HE can be conducted with varying degrees of rigor. The most rigorous approach is to drive expression with the endogenous promoter and ask whether YFG can rescue the phenotype of a null



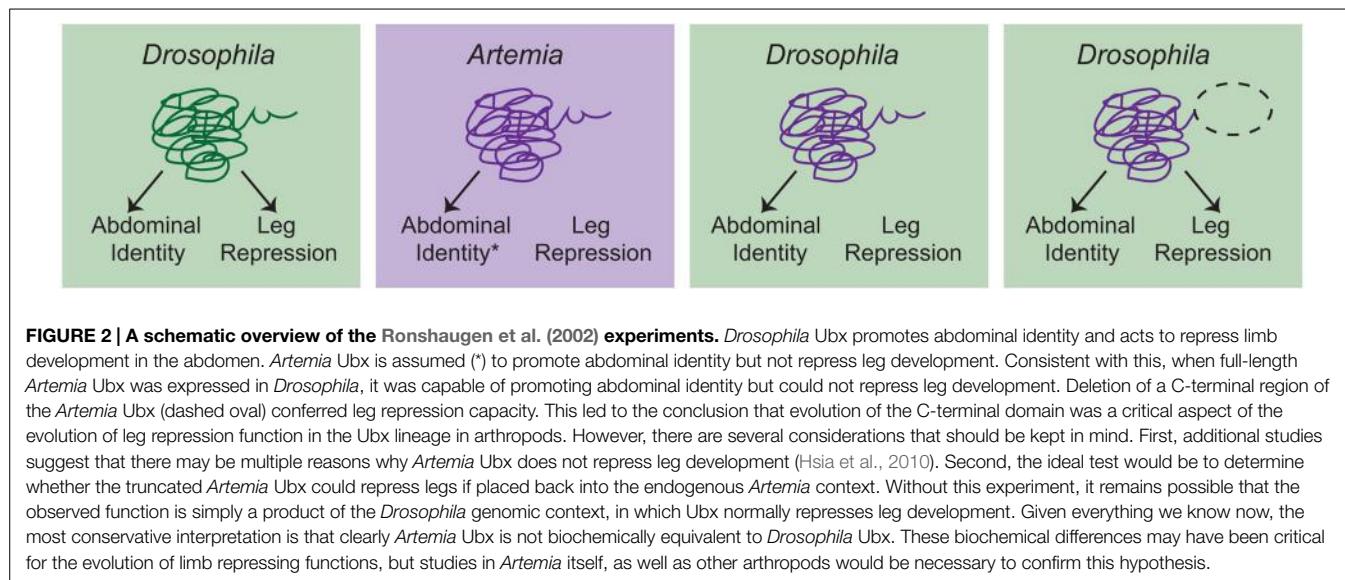
mutation in the endogenous locus. With surprising frequency, however, the heterologous locus is simply over-expressed in a wild type background (e.g., Lee et al., 2012; Perilleux et al., 2013; Lovisetto et al., 2015), such that the real question being asked is: Can this alien protein perturb development in the same manner as the endogenous protein when it is over-expressed? Such an approach creates new problems, including the nature of protein interactions, which are subject to reaction equilibria and therefore sensitive to the concentrations of the interacting factors.

Given this perspective, we should consider the variety of ways that HE is typically used in the evodevo field. These include to bolster evidence of genetic orthology (e.g., Serrano et al., 2009), to assess homology of a genetic module or an organ (e.g., Halder et al., 1995; Whipple et al., 2004), and to broadly assess conservation of gene “function” between taxa (e.g., Alvarez-Buylla et al., 2010; Kachroo et al., 2015). The first of these uses should be rejected since similarity of function is absolutely not a criterion for genetic homology in general or orthology in particular (Theissen, 2002; Gabaldon and Koonin, 2013). It is even true that positive HE results can be misleading when it comes to assessing orthology. Perhaps the best understood instance of this phenomenon is the *AGAMOUS* (*AG*) lineage of floral organ identity genes in flowering plants. The functions of *AG* homologs were first described in the core eudicot model systems *Arabidopsis* and *Antirrhinum* (snapdragon; Coen and Meyerowitz, 1991). In *Arabidopsis*, the *ag* mutant phenotype results in homeotic transformation of fertile organs into sterile organs and a loss of determinacy in the floral meristem. The *plena* (*ple*) mutant in *Antirrhinum* has the identical phenotype and *PLE* is clearly homologous to *AG*. However, *PLE* and *AG* are not orthologous but, rather, are derived from a whole genome duplication that occurred at the base of the core eudicots (Davies et al., 1999; Kramer et al., 2004). The orthologs of *PLE* in *Arabidopsis* are a pair of recent duplicates called *SHATTERPROOF1/2*, which participate

in fruit and ovule development (Liljegren et al., 2000), while the ortholog of *AG* in *Antirrhinum* is called *FARINELLI* (*FAR*), a gene that primarily contributes to stamen development (Davies et al., 1999; Causier et al., 2005). These distinct functions appear to be due to independent patterns of subfunctionalization that occurred along the lineages leading to the rosid *Arabidopsis* on the one hand and the asterid *Antirrhinum* on the other. Furthermore, while the paralogs *AG* and *PLE* are biochemically equivalent in *Arabidopsis*, the orthologs *AG* and *FAR* are not (Causier et al., 2005; Airolidi et al., 2010). This is most likely due to changes in selection as *FAR* became specialized to function in stamen identity. Therefore, while it may commonly be true that orthologs are more likely than not to have both functional similarity and biochemical conservation, we cannot take it for granted.

If you will permit me a digression, I would also like to strongly discourage the common use of the term “functional ortholog.” It is important to remember that function is generally considered not to be a criterion for homology, even among genes (Theissen, 2002). I actually agree with Mindell and Meyer (2001) on this point, that there should be some leeway for discussing the inheritance of genetic function, but we should recognize that it is widely held that functions of any kind cannot be homologous. What information are we trying to convey when we say “functional ortholog?” We want to say that we have a pair of genes that are genetic orthologs and also appear to play similar functional roles. This is an important piece of information; certainly, we often want to know if function is conserved among orthologs. However, this terminology seems to suggest that “functional” orthologs have an additional quality of greater orthology because they show conserved function. This is simply untrue. Orthology is a feature of genetic relationship, of inheritance and patterns of gene duplication. It does not increase or decrease based on functional similarity. It is much more informative to say that you have performed a rigorous phylogenetic and/or syntenic analysis and have determined that the genes in question are orthologs and, further, appear to share conserved functions. We must recognize that this statement can really only be made if you have conducted endogenous functional studies in the taxa being compared. If you have only performed HE, then the best you can say is that there is some degree of biochemical conservation.

The use of HE to assess homology of a genetic module or an organ is more complex and relates to the need to distinguish between process homology and morphological homology, which has been well-covered by many previous authors (Bolker and Raff, 1996; Abouheif, 1997; Abouheif et al., 1997). These authors recognized quite early during the molecular renaissance of our field that shared expression of genetic homologs, and even shared developmental control by homologous genetic modules, should not be used as the basis for assessment of morphological homology. Hodin (2000) succinctly addressed the issue while discussing the limited value of HE with *Pax6* homologs: “A positive result tells you only that the biochemical properties of the protein have been conserved, not necessarily that its function within a certain morphological structure has also been conserved. The commonplace use of the same gene within an organism performing distinct functions in a multitude of tissue



reveals why this experiment is generally uninformative with respect to evolutionary history (see also Abouheif et al., 1997).” Here, Hodin seeks to highlight the fact that conservation of biochemical interactions within a particular genetic module does not inform on the myriad of ways in which that module can be developmentally deployed. In this regard, I should note that HE can provide some relevant information if you are simply trying to assess homology of a genetic module, but I would argue that phylogeny-based homology assessment of the genes involved and tests of endogenous regulatory interactions are even more useful.

Process homology is especially relevant to cases of co-option of genetic modules to novel developmental functions. For instance, in butterflies *Distal-less* (*Dll*) orthologs have been recruited to promote the development of wing spots (Brunetti et al., 2001). The wing spot developmental program is very unlikely to be recapitulated by simply expressing the butterfly *Dll* in *Drosophila* because this developmental program is a product of what I defined as the second component of gene function, the endogenous genomic architecture of the butterfly. However, reciprocal HE of *Dll* orthologs between *Drosophila* and butterflies would be perfectly useful if your goal was to determine whether the evolution of the wing spot involved biochemical divergence in the butterfly *Dll* sequence. This type of co-option is just one extreme on a spectrum of evolutionary change that could also include morphological remodeling events such as the derivation of halteres from hindwings (Hersh et al., 2007), lodicules from petals (Whipple et al., 2007; Yoshida, 2012) or staminodia from stamens (Sharma and Kramer, 2013). Such evolutionary transitions may involve biochemical changes in upstream transcription factors but clearly also involve changes in target gene repertoires (e.g., Hersh et al., 2007). HE is much more likely to shed light on any biochemical changes rather than changes in target gene repertoires, which primarily depend on the positions of downstream binding sites dispersed throughout the genome.

The third common use of HE, to investigate conservation of “function,” is perfectly legitimate in many cases but less so in others. It is probably useful to start with a consideration of what

can go wrong with HE. For instance, a lack of rescue or the failure to produce a phenotype may simply be due to the divergence between your species of interest and the reference model system. Even proteins that are likely to serve conserved functions can experience the process of developmental system drift (True and Haag, 2001) at the level of primary sequence. In other words, this is a site directed mutagenesis experiment in which the altered protein cannot function in the model system’s genomic context but may be perfectly functional in its original environment. On occasion, HE results in novel or dominant negative phenotypes (e.g., Lee et al., 2012; Katahata et al., 2014; Sun et al., 2014). These may be due to the disruptive effects of an alien protein being introduced to a system for which it is not adapted. If the heterologous protein can interact with some co-factors but not others, it may act as a dominant negative allele, especially when over-expressed. Perhaps most surprisingly though, even positive results can be misleading. Zarrinpar et al. (2003) tested the ability of SH3-domain protein homologs to rescue the function of one specific family member in yeast. They found that while endogenous paralogs were highly functionally specific and could not rescue, diverse metazoan homologs showed higher frequency of rescue. These results reflect the fact that members of the same genome, especially when co-expressed, will tend to co-evolve for a high degree of functional specificity. Homologs from divergent genomic contexts that have not experienced the same patterns of co-evolution may actually be quite promiscuous in a heterologous genome. Thus, we see that a range of results from HE can be uninformative or misleading, especially when you do not have functional data from the original organism.

So am I suggesting that HE is never useful for examining the evolution of gene function? Certainly not. In cases where biochemical divergence is specifically being assessed, this approach can be the best experiment to use, albeit with some caveats. Let’s consider a classic HE experiment, Ronshaugen et al. (2002), in which they tested the ability of *Artemia* Ubx to suppress limb development in *Drosophila* (Figure 2). Interestingly, the authors found that while the full length *Artemia* Ubx had little

limb-suppressing capacity, a relatively minor C-terminal deletion allowed the *Artemia* protein to repress limbs in *Drosophila*. In light of this finding, the authors proposed a model in which the Ubx protein of a crustacean/insect ancestor experienced mutation in the C-terminus of the protein that uncovered a limb-repression function. This is certainly a plausible scenario that fits the presented data, but we should also recognize a weakness in that the experiment was only performed in the *Drosophila* genomic context where Ubx has a limb repressing function. If you could put the mutated *Artemia* Ubx back into *Artemia*, would it have the capacity to repress limbs or is that primarily a product of the *Drosophila* genome? As it turns out, further studies in *Artemia* have revealed a more complex situation that suggests that there may be multiple reasons why Ubx does not repress limbs in *Artemia* (Hsia et al., 2010). These findings underscore the fact that accurate interpretation of HE data really hinges on having as much information as possible in both taxa, including functional results whenever possible.

One especially elegant demonstration of how powerful HE can be when paired with functional studies in both the donor and recipient is work done on the control of flowering time in sugar beet (*Beta vulgaris* ssp. *vulgaris*, Pin et al., 2010). In flowering plants, homologs of the PEBP lineage defined by the *Arabidopsis* gene *FLOWERING LOCUS T* (*FT*) are broadly involved with promoting the transition from vegetative to reproductive development (reviewed Ballerini and Kramer, 2011). The *FT* protein has been identified as the classic Florigen factor that moves from leaves, where it is produced, to the apical meristem in order to change meristem identity. Consistent with this role, most *FT* homologs are only expressed at significant levels after the initiation of reproductive development. In cultivated sugar beet, however, a very recent gene duplication has given rise to two copies: *BvFT1*, which is primarily expressed during vegetative development, and *BvFT2*, which is expressed as expected during the reproductive stage (Pin et al., 2010). Using RNAi and overexpression in beet, Pin et al. (2010) clearly established that the *BvFT1* paralog had acquired a dominant negative effect that represses flowering until vernalization (cold treatment) represses *BvFT1* and allows expression of the floral promoting paralog *BvFT2*. This dramatic difference in function between the two paralogs can be recapitulated in *Arabidopsis*, where *BvFT2* activates flowering while *BvFT1* represses it. This demonstrates that there is a biochemical change in *BvFT1* relative to the otherwise highly conserved function of *FT* proteins. The use of chimeric proteins and site-directed mutagenesis in the more tractable *Arabidopsis* system allowed the authors to identify the specific amino acid changes that are responsible for the neofunctionalization, and further demonstrate that these changes are associated with *BvFT1* alleles that were selected during domestication. This kind of study relies heavily on HE but uses it in exactly the right way—by targeting an otherwise highly conserved genetic module, and in combination with detailed expression and functional studies in the original system, which allows the heterologous results to be accurately interpreted.

Another powerful application of HE is to use homologs from a series of diverging taxa to probe the conservation of specific

biochemical properties, such as recognition of DNA binding sites. This is essentially a matter of letting evolution do the site-directed mutagenesis for you: as you move out to more deeply diverging taxa, there are more non-synonymous mutations, allowing you to ask whether the endogenous biochemical function is still retained. The land plant-specific transcription factor *LEAFY* (*LFY*) is ideal for this type of study because unlike most plant gene lineages, it has very few retained paralogs. Maizel et al. (2005), tested the ability of *LFY* homologs from across the land plants to rescue the *lfy* mutation in *Arabidopsis*, and then further complemented the phenotypic analysis with microarray studies of gene expression. They found that there was a gradual decreasing degree of phenotypic rescue as they moved out to more distantly related taxa. When paired with tests of protein/DNA interaction, their results suggest that “the declining ability to replace *Arabidopsis LFY* . . . is caused by a progressive failure to interact with the canonical *LFY* binding sites,” which, of course, are defined based on work done in *Arabidopsis*. The microarray analysis of the various transgenic lines demonstrated that in the weakest cases of rescue, one of the last target interactions to be lost was with the floral meristem identity gene *APETALA1* (*AP1*). The authors quite correctly noted that this finding does not tell us anything about what the heterologous *LFY* homologs activate in their endogenous settings—*AP1* homologs are not even present outside angiosperms. Rather, this reflects the extraordinarily high affinity of the *LFY* binding site present in the *AP1* promoter, such that even deeply divergent homologs with many non-synonymous changes are still capable of recognizing it. This kind of study highlights evolutionary processes affecting both aspects of developmental gene function since it detects biochemical changes that have altered DNA affinity while also underscoring the fact that repertoires of target genes will simultaneously be evolving.

In summary, my argument is that HE can be very useful in specific cases where we want to investigate changes in the primary component of gene function, which is to say biochemical function. This includes enzymatic capacity as well as affinity for a range of interactions such as protein-DNA and protein–protein. It yields the best results when paired with functional studies in both the donor and recipient taxa so that potentially spurious phenotypes can be ruled out. I think it is also true that HE works best when you can target a genetic module that is otherwise very highly conserved, so that you can lessen the impact of drift and divergence in other components of the pathway (although this is hard to ever rule out completely!). HE does not inform upon homology in general or orthology in particular, nor does it give us much information on what developmental roles the gene may play in its original genomic context, so use it with care.

Acknowledgments

I would like to acknowledge several other scientists with whom I have discussed this question over the years, particularly Vivian Irish, who set me straight on the subject to begin with, as well as Amy Litt, and Sarah Mathews.

References

- Abouheif, E. (1997). Developmental genetics and homology: a hierarchical approach. *Tren. Ecol. Evol.* 12, 405–408. doi: 10.1016/S0169-5347(97)01125-7
- Abouheif, E., Akam, M., Dickinson, W. J., Holland, P. W. H., Meyer, A., Patel, N. H., et al. (1997). Homology and developmental genes. *Tren. Gen.* 13, 432–433. doi: 10.1016/S0168-9525(97)01271-7
- Airoldi, C. A., Bergonzi, S., and Davies, B. (2010). Single amino acid change alters the ability to specify male or female organ identity. *Proc. Natl. Acad. Sci. U.S.A.* 107, 18898–18902. doi: 10.1073/pnas.1009050107
- Alvarez-Buylla, E. R., Ambrose, B. A., Flores-Sandoval, E., Vergara-Silva, F., Englund, M., Garay-Arroyo, A., et al. (2010). B-function expression in the flower center underlies the homeotic phenotype of *Lacandonia schismatica* (Triuridaceae). *Plant Cell* 22, 3543–3559. doi: 10.1105/tpc.109.096153
- Ballerini, E. S., and Kramer, E. M. (2011). In the light of evolution: a reevaluation of conservation in the CO-FT regulon and its role in photoperiodic regulation of flowering time. *Front. Plant Sci.* 2:81. doi: 10.3389/fpls.2011.00081
- Bolker, J. A., and Raff, R. A. (1996). Developmental genetics and traditional homology. *Bioessays* 18, 489–494. doi: 10.1002/bies.950180611
- Brunetti, C. R., Selegue, J. E., Monteiro, A., French, V., Brakefield, P. M., and Carroll, S. B. (2001). The generation and diversification of butterfly eyespot color patterns. *Curr. Biol.* 11, 1578–1585. doi: 10.1016/S0960-9822(01)00502-4
- Causier, B., Castillo, R., Zhou, J. L., Ingram, R., Xue, Y. B., Schwarz-Sommer, Z., et al. (2005). Evolution in action: following function in duplicated floral homeotic genes. *Curr. Biol.* 15, 1508–1512. doi: 10.1016/j.cub.2005.07.063
- Coen, E. S., and Meyerowitz, E. M. (1991). The war of the whorls: genetic interactions controlling flower development. *Nature* 353, 31–37. doi: 10.1038/353031a0
- Davies, B., Motte, P., Keck, E., Saedler, H., Sommer, H., and Schwarz-Sommer, Z. (1999). PLENA and FARINELLI: redundancy and regulatory interactions between two *Antirrhinum* MADS-box factors controlling flower development. *EMBO J.* 18, 4023–4034. doi: 10.1093/emboj/18.14.4023
- Gabaldon, T., and Koonin, E. V. (2013). Functional and evolutionary implications of gene orthology. *Nat. Rev. Genet.* 14, 360–366. doi: 10.1038/nrg3456
- Halder, G., Callaerts, P., and Gehring, W. J. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in *Drosophila*. *Science* 267, 1788–1792.
- Hersh, B. M., Nelson, C. E., Stoll, S. J., Norton, J. E., Albert, T. J., and Carroll, S. B. (2007). The UBX-regulated network in the haltere imaginal disc of *D. melanogaster*. *Dev. Biol.* 302, 717–727. doi: 10.1016/j.ydbio.2006.11.011
- Hodin, J. (2000). Plasticity and constraints in development and evolution. *J. Exp. Zool. (Mol. Dev. Evol.)* 288, 1–20. doi: 10.1002/(SICI)1097-010X(20000415)288:1<1::AID-JEZ1>3.0.CO;2-7
- Hsia, C. C., Pare, A. C., Hannon, M., Ronshaugen, M., and McGinnis, W. (2010). Silencing of an abdominal Hox gene during early development is correlated with limb development in a crustacean trunk. *Evol. Dev.* 12, 131–143. doi: 10.1111/j.1525-142X.2010.00399.x
- Kachroo, A. H., Laurent, J. M., Yellman, C. M., Meyer, A. G., Wilke, C. O., and Marcotte, E. M. (2015). Systematic humanization of yeast genes reveals conserved functions and genetic modularity. *Science* 348, 921–925. doi: 10.1126/science.aaa0769
- Katahata, S. I., Futamura, N., Igasaki, T., and Shinohara, K. (2014). Functional analysis of SOC1-like and AGL6-like MADS-box genes of the gymnosperm *Cryptomeria japonica*. *Tree Genet. Genomes* 10, 317–327. doi: 10.1007/s11295-013-0686-9
- Kramer, E. M., Jaramillo, M. A., and Di Stilio, V. S. (2004). Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS-box genes in angiosperms. *Genetics* 166, 1011–1023. doi: 10.1534/genetics.166.2.1011
- Lee, J. H., Park, S. H., and Ahn, J. H. (2012). Functional conservation and diversification between rice OsMADS22/OsMADS55 and *Arabidopsis* SVP proteins. *Plant Sci.* 185, 97–104. doi: 10.1016/j.plantsci.2011.09.003
- Liljegren, S. J., Ditta, G. S., Eshed, Y., Savidge, B., Bowman, J. L., and Yanofsky, M. F. (2000). SHATTERPROOF MADS-box genes control seed dispersal in *Arabidopsis*. *Nature* 404, 766–770. doi: 10.1038/35008089
- Lovisotto, A., Baldan, B., Pavanello, A., and Casadoro, G. (2015). Characterization of an AGAMOUS gene expressed throughout development of the fleshy fruit-like structure produced by *Ginkgo biloba* around its seeds. *BMC Evol. Biol.* 15:139. doi: 10.1186/s12862-015-0418-x
- Maizel, A., Busch, M. A., Tanahashi, T., Perkovic, J., Kato, M., Hasebe, M., et al. (2005). The floral regulator LEAFY evolves by substitutions in the DNA binding domain. *Science* 308, 260–263. doi: 10.1126/science.1108229
- Mindell, D. P., and Meyer, A. (2001). Homology evolving. *Trends Ecol. Evol.* 16, 434–440. doi: 10.1016/S0169-5347(01)02206-6
- Perilleux, C., Pieltain, A., Jacquemin, G., Bouche, F., Detry, N., D'aloia, M., et al. (2013). A root chicory MADS box sequence and the *Arabidopsis* flowering repressor FLC share common features that suggest conserved function in vernalization and de-vernalization responses. *Plant J.* 75, 390–402. doi: 10.1111/tpj.12208
- Pin, P. A., Benlloch, R., Bonnet, D., Wremerth-Weich, E., Kraft, T., Gielen, J. J. L., et al. (2010). An antagonistic pair of FT homologs mediates the control of flowering time in sugar beet. *Science* 330, 1397–1400. doi: 10.1126/science.1197004
- Ronshaugen, M., McGinnis, N., and McGinnis, W. (2002). Hox protein mutation and macroevolution of the insect body plan. *Nature* 415, 914–917. doi: 10.1038/nature716
- Serrano, G., Herrera-Palau, R., Romero, J. M., Serrano, A., Coupland, G., and Valverde, F. (2009). *Chlamydomonas CONSTANS* and the evolution of plant photoperiodic signaling. *Curr. Biol.* 19, 359–368. doi: 10.1016/j.cub.2009.01.044
- Sharma, B., and Kramer, E. M. (2013). Sub- and neofunctionalization of APETALA3 paralogs have contributed to the evolution of novel floral organ identity in *Aquilegia* (columbine, Ranunculaceae). *New Phytol.* 197, 949–957. doi: 10.1111/nph.12078
- Sun, W., Huang, W. J., Li, Z. N., Song, C., Liu, D., Liu, Y. L., et al. (2014). Functional and evolutionary analysis of the API/SEP/AGL6 superclade of MADS-box genes in the basal eudicot *Epimedium sagittatum*. *Ann. Bot.* 113, 653–668. doi: 10.1093/aob/mct301
- Theissen, G. (2002). Secret life of genes. *Nature* 415, 741. doi: 10.1038/415741a
- True, J. R., and Haag, E. S. (2001). Developmental system drift and flexibility in evolutionary trajectories. *Evol. Dev.* 3, 109–119. doi: 10.1046/j.1525-142x.2001.003002109.x
- Whipple, C. J., Ciceri, P., Padilla, C. M., Ambrose, B. A., Bandong, S. L., and Schmidt, R. J. (2004). Conservation of B-class floral homeotic gene function between maize and *Arabidopsis*. *Development* 131, 6083–6091. doi: 10.1242/dev.01523
- Whipple, C. J., Zanis, M. J., Kellogg, E. A., and Schmidt, R. J. (2007). Conservation of B class gene expression in the second whorl of a basal grass and outgroups links the origin of lodicles and petals. *Proc. Natl. Acad. Sci. U.S.A.* 104, 1081–1086. doi: 10.1073/pnas.0606434104
- Yoshida, H. (2012). Is the lodicule a petal: molecular evidence? *Plant Sci.* 184, 121–128. doi: 10.1016/j.plantsci.2011.12.016
- Zarrinpar, A., Park, S. H., and Lim, W. A. (2003). Optimization of specificity in a cellular protein interaction network by negative selection. *Nature* 426, 676–680. doi: 10.1038/nature02178

Conflict of Interest Statement: The author declares 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 © 2015 Kramer. 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) or licensor 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.



Evolution of epithelial morphogenesis: phenotypic integration across multiple levels of biological organization

Thorsten Horn, Maarten Hilbrant and Kristen A. Panfilio*

Institute for Developmental Biology, University of Cologne, Cologne, Germany

OPEN ACCESS

Edited by:

Sylvain Marcellini,
University of Concepcion, Chile

Reviewed by:

David Q. Matus,
Stony Brook University, USA
Stefanie D. Hueber,
University of Konstanz, Germany

*Correspondence:

Kristen A. Panfilio,
Institute for Developmental Biology,
University of Cologne, Zülpicher Str.
47b, 50674 Cologne, Germany
kristen.panfilio@alum.swarthmore.edu

Specialty section:

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

Received: 30 June 2015

Accepted: 11 September 2015

Published: 29 September 2015

Citation:

Horn T, Hilbrant M and Panfilio KA
(2015) Evolution of epithelial morphogenesis: phenotypic integration across multiple levels of biological organization.
Front. Genet. 6:303.
doi: 10.3389/fgene.2015.00303

Morphogenesis involves the dynamic reorganization of cell and tissue shapes to create the three-dimensional body. Intriguingly, different species have evolved different morphogenetic processes to achieve the same general outcomes during embryonic development. How are meaningful comparisons between species made, and where do the differences lie? In this Perspective, we argue that examining the evolution of embryonic morphogenesis requires the simultaneous consideration of different levels of biological organization: (1) genes, (2) cells, (3) tissues, and (4) the entire egg, or other gestational context. To illustrate the importance of integrating these levels, we use the extraembryonic epithelia of insects—a lineage-specific innovation and evolutionary hotspot—as an exemplary case study. We discuss how recent functional data, primarily from RNAi experiments targeting the *Hox3/Zen* and U-shaped group transcription factors, provide insights into developmental processes at all four levels. Comparisons of these data from several species both challenge and inform our understanding of homology, in assessing how the process of epithelial morphogenesis has itself evolved.

Keywords: epithelial morphogenesis, evolution of development, insects, extraembryonic tissues, *Hox3/zен*, *Tribolium castaneum*, *Megaselia abdita*, *Oncopeltus fasciatus*

Introduction

In the rapidly developing fruit fly *Drosophila melanogaster*, the predominant insect model for developmental genetics, embryonic morphogenesis occurs largely after cell fates are determined. Indeed, there is extensive literature on *Drosophila* early tissue patterning, including axis specification and segmentation, preceding morphogenesis. Perhaps as a result of our profound knowledge in *Drosophila*, many evolutionary developmental (evo-devo) studies in arthropods take a gene-centered approach and focus on early patterning, as early fate specification is often a powerful signal for comparisons of species that are separated by long periods of evolutionary time (e.g., Peel et al., 2005; Sachs et al., 2015).

In this Perspective article, however, we highlight the importance of studying the morphogenetic movements that occur during animal development and of integrating multiple levels of biological organization when making interspecific comparisons. For doing so, we distinguish between four increasingly inclusive levels of biological organization. (1) Genetic regulation of development comprises information about the specific genes and their protein products that are involved in transcriptional control, signaling cascades, and the molecular basis of cytoskeletal structure and remodeling. (2) Individual cells differentiate to acquire a particular identity, including

the transcriptional state as well as cell shape and structure. (3) More broadly, cells coordinate with their neighbors within tissues. In epithelial tissues for example, cells retain contact with their neighbors via adherens junctions, such that cell shape changes affect the entire tissue's geometry. (4) Finally, the egg is a global system, where tissue integrity and inter-tissue adhesion need to be precisely controlled during morphogenesis to achieve the final form.

Overall, integration of different biological levels is as much a conceptual framework for understanding the physical context of a given gene's role in a developmental process as for interpreting how morphogenesis has evolved. To illustrate this, here we discuss recent advances in the study of extraembryonic (EE) development in a range of insect model species. We show that the insect EE epithelia provide a case study with a particularly rich evolutionary history, making them well suited to assessing the evolution of morphogenesis.

Development and Evolution of the Insect Extraembryonic Membranes

Many arthropod eggs include an EE tissue component, but in the insects this feature has become a specific structural innovation (reviewed in Panfilio, 2008). At the base of the winged insect lineage, the EE epithelial membranes evolved to form discrete compartments within the egg. In most winged insects, the serosa lines the eggshell, providing the outermost cellular layer and enclosing all other contents, including the yolk. The amnion, analogous to its namesake in vertebrates, forms a fluid-filled cavity ventral to the embryo, retaining a connection to the embryo along the latter's dorsal margin (**Figure 1**: "most insects" schematic).

The ability of EE membranes to form these compartments early in development has allowed the insects to exploit diverse ecological niches, largely due to the manifold functions of the serosa as a protective outer layer that buffers the embryo against environmental fluctuations and assaults. Recent work has shown that serosal cuticle secretion correlates with the acquisition of desiccation resistance, and that the cuticle itself provides mechanical support to the egg (Rezende et al., 2008; Jacobs et al., 2013; Panfilio et al., 2013). At the same time, recent experimental evidence demonstrates the long hypothesized ability of the serosa to protect the embryo after wounding and pathogen infection via upregulation of the innate immune system (Chen et al., 2000; Jacobs et al., 2014). Furthermore, the serosa's ultrastructure is consistent with physiological roles in water and solute processing, and it has acquired additional mechanical and physiological functions during hatching and early larval life in species with oviposition sites within plant and animal tissues (citations in Panfilio, 2008).

The functional importance of the amnion remains far more enigmatic, despite early recognition of the potential value of an insect amniotic cavity (Zeh et al., 1989). Indeed, the amniotic cavity has been lost independently during the evolution of apocritan wasps and cyclorrhaphan flies (Fleig and Sander, 1988; Rafiqi et al., 2008), with the amniotic epithelium confined to a dorsal yolk cover (**Figure 1**: schematic for *Megaselia*). More extremely, in *Drosophila melanogaster* the serosa and amnion are

conflated into a single, dorsal amnioserosa, dispensing with EE compartments entirely (**Figure 1**: schematic for *Drosophila*), and some *Drosophila* species have decanализed development to the point where amnioserosal formation is variable, but still essential (Gavin-Smyth et al., 2013; Panfilio and Roth, 2013).

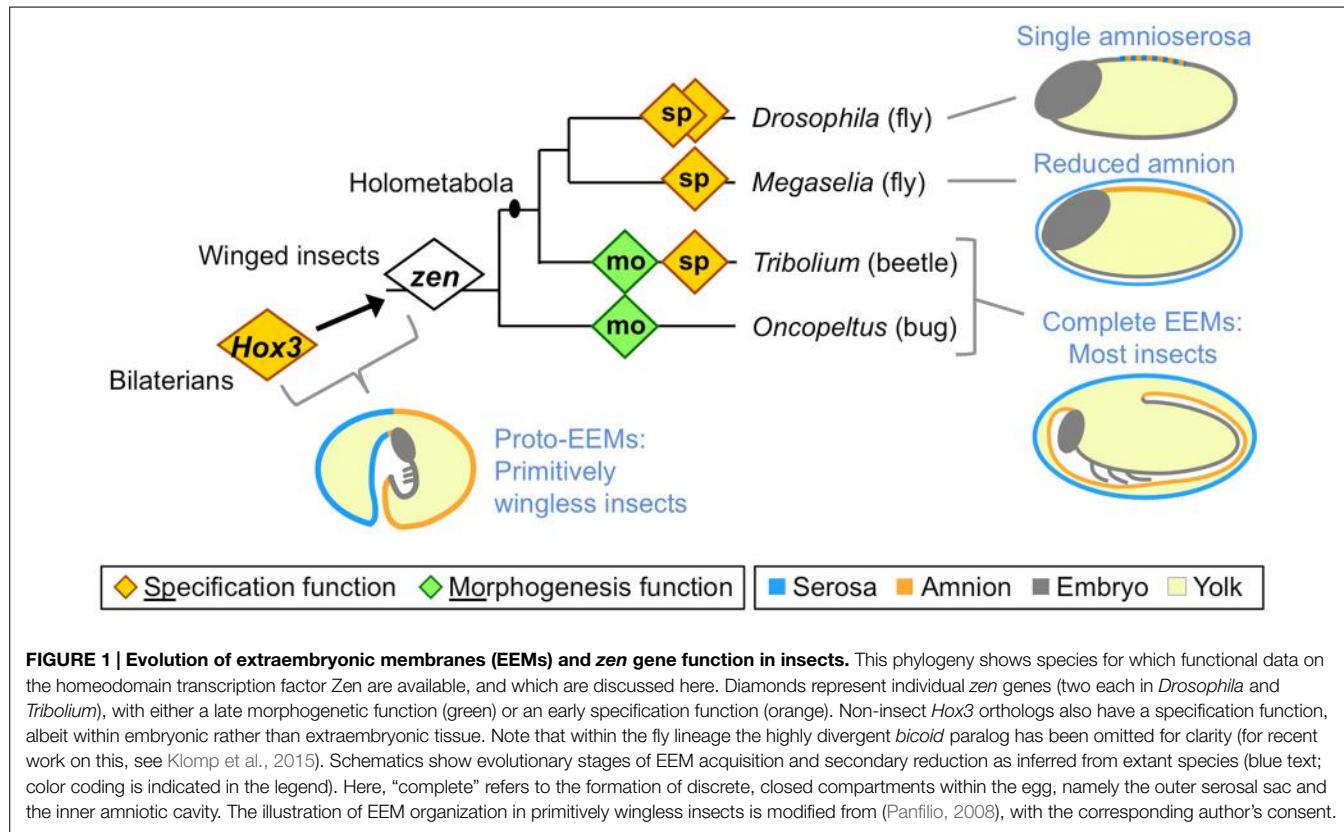
Using the evolution and development of insect EE membranes as a case study, in the next sections we discuss how the different levels of biological organization are interconnected. We show that, for example, changes at the gene level can induce dramatic changes in cell and tissue behavior that differ between species, even if the consequences at the whole egg level are similar. On the other hand, similar morphogenetic movements can be achieved by quite different mechanisms on the cellular and tissue organizational levels.

Linking zen and U-shaped Genes to Changes at the Egg Level

Extraembryonic tissue evolution is tightly linked to evolution of the Hox3/Zen transcription factor (**Figure 1**). The evolutionary origin of strictly EE expression of this gene coincides with the origin of complete EE compartments (Hughes et al., 2004; Panfilio et al., 2006). Hox genes are generally highly conserved in relative genomic position, protein sequence, copy number, and function in anterior-posterior patterning (Krumlauf, 1992; Cook et al., 2001). In contrast, arthropod *Hox3* orthologs are prone to duplication and marked sequence divergence, particularly the insect orthologs (known as "zen," after the original *Drosophila* mutants), with independent instances of duplication in beetles, flies, and lepidopterans (Pultz et al., 1988; Brown et al., 2002; Panfilio et al., 2006; Panfilio and Akam, 2007; Chai et al., 2008; Rafiqi, 2008; Ferguson et al., 2014). Interestingly, in the red flour beetle, *Tribolium castaneum*, the two *zen* paralogs have different functions (van der Zee et al., 2005), and these will be discussed in turn.

In all holometabolous insects studied so far *zen*, or *Tc-zen1* in *Tribolium*, has a conserved function in EE tissue specification (**Figure 1**: orange diamonds). However, while loss of function mutation in *Drosophila* is lethal (Wakimoto et al., 1984), the scuttle fly *Megaselia abdita* and *Tribolium* can survive after RNA interference (RNAi) knockdown (van der Zee et al., 2005; Rafiqi et al., 2008; Panfilio et al., 2013). Examining why the end-stage phenotypes differ after loss of a conserved gene's function provides a good example for the integration of the different levels of biological organization.

Key to understanding the different phenotypic outcomes of disrupting *zen/zen1* is the evolutionary change in EE membrane complement between these species. The single amnioserosa of *Drosophila* exhibits features of both early serosa and late amnion (reviewed in Schmidt-Ott et al., 2010), and specification of the entire EE domain is under the control of *Dm-zen* (the paralog *Dm-z2* is not essential during embryogenesis, Pultz et al., 1988). In the less derived situation in *Megaselia* and *Tribolium*, *Ma-zen/Tc-zen1* only specifies the serosa (van der Zee et al., 2005; Rafiqi et al., 2008). The phenotypic outcome after *zen* knockdown could then be explained by loss of all EE tissue identity in *Drosophila*, while *Megaselia* and *Tribolium* retain an amnion.



However, on closer inspection the similarity in gene function, residual EE tissue complement, and end-stage phenotypic outcome in *Megaselia* and *Tribolium* is rather surprising if we consider the difference in EE membrane configuration (**Figure 1**: schematics). In both species, we observe a respecification from serosal to amniotic fate and in both species it is important to have a tissue covering the yolk dorsally during the dorsal closure stage. However, the underlying wild type configurations are different. In *Megaselia*, the amnion provides a persistent dorsal yolk cover, and its overall shape, size, and dorsal position are not changed dramatically by the *Ma-zen*^{RNAi} fate shift (Rafiqi et al., 2008). In contrast, in *Tribolium* the dorsal side of the egg is first covered by the serosa and only later in wild type development is the amnion pulled dorsally when the serosa contracts (Panfilio et al., 2013). How can *Tribolium* then survive without a serosa? Here, *Tc-zen*^{RNAi} not only produces a persistently dorsal amniotic region due to respecification (van der Zee et al., 2005), but also reveals novel cellular and tissue properties of the entire amnion in late development as it takes over the role of the serosa in providing a dorsal cover (Panfilio et al., 2013). Hence, the survival of *Ma-zen*^{RNAi} embryos is rather due to the dispensability of the serosa for dorsal closure, while in *Tribolium* developmental regulation—that is, compensation via plasticity of the amnion—enables survival after *Tc-zen*^{RNAi}. Thus, conserved, early gene functions can feed into different developmental routes, depending on tissue configuration and morphogenetic properties.

At the same time, other genes with EE roles have undergone changes in their particular function and in their interaction

partners during insect evolution. One example is the T-box transcription factor Dorsocross (Doc), a member of the U-shaped gene family (Frank and Rushlow, 1996; Reim et al., 2003). In *Drosophila*, *Dm-Doc* is necessary for the maintenance of the amnioserosa toward the end of germband extension, when Zen protein disappears (Reim et al., 2003, and references therein). In contrast, *Tc-Doc* has multiple roles in *Tribolium* EE morphogenesis, but no role in maintaining either EE tissue (TH, KAP unpublished observation). There is some evidence that *Ma-Doc* has a maintenance function in the *Megaselia* serosa (Rafiqi et al., 2008), but the end stage RNAi phenotype would also be consistent with an early morphogenetic role, as in *Tribolium*.

Consistent with this difference in the EE role of Doc, the molecular context of its function also differs between species. *Drosophila Doc* expression requires simultaneous inputs from *Dm-Zen* and *Dm-Dpp* (Reim et al., 2003). In contrast, in *Tribolium* these inputs are temporally and spatially distinct, and subsequent Dpp signaling is itself locally dependent on *Tc-Doc* (TH, KAP unpublished observation), a feature not known from *Drosophila*. Another example is *Doc*’s relation to *hindSight* (*hnt*), another U-shaped gene. In *Drosophila*, *Dm-hnt* is downstream of *Dm-Doc* and therefore shows a similar knockdown phenotype. In *Tribolium*, both genes also show a similar knockdown phenotype to one another, but they seem not to influence each other’s expression (TH, KAP unpublished observation).

Finally, *Dm-Doc* performs multiple functions within the body proper (Hamaguchi et al., 2012; Sui et al., 2012), such as for heart development (Reim and Frasch, 2005), that are not observed in *Tribolium* (Nunes da Fonseca et al., 2010).

Interestingly, one of these functions, bending of the *Drosophila* wing imaginal disc, directly links the transcription factor Dm-Doc to cellular and epithelial rearrangements (Sui et al., 2012). Here, Dm-Doc promotes intracellular microtubule web redistribution and degradation of the extracellular matrix through Matrix metalloproteinase. It remains to be seen if similar mechanisms are also employed downstream of Doc in EE morphogenesis across species.

In summary, disruption of *zen*, a gene with a conserved function in specification of the serosa, leads to lethality, compensation by the amnion, or simply loss of the serosa with no severe consequences for development, depending on the tissue topography of the species under investigation. Moreover, there are large differences in gene knockdown phenotypes, overall gene functions, and specific interaction partners between orthologous genes in different species, and these differences can only be understood if all other biological levels, from cells to the egg system, are taken into account.

In the next section we shift the focus from the genes themselves to tissue organization and function, again highlighting differences between species at different levels of biological organization.

Linking Cellular, Tissue, and Egg System Levels

In late embryogenesis, it is essential that insect EE tissue actively withdraws in a precise way to mediate dorsal closure, whereby the embryonic epidermis seals at the dorsal midline and EE tissue degenerates within the yolk. Indeed, amnioserosa-epidermal tissue coordination during *Drosophila* dorsal closure has been extensively studied over the last 15 years (e.g., Jacinto et al., 2000; Kiehart et al., 2000; Solon et al., 2009; Lada et al., 2012; Wells et al., 2014). While differences in tissue organization are expected between dorsal closure involving an amnioserosa and dorsal closure involving a serosa and amnion, we also find differences between species with both EE membranes (Figure 1: “most insects” schematic). To illustrate this point, here we compare late EE morphogenesis between *Tribolium* and the hemimetabolous milkweed bug, *Oncopeltus fasciatus*, charting a sequence of similarities and differences as morphogenesis proceeds.

Firstly, rupture of the EE tissues over the embryo’s head produces an opening through which the embryo passively emerges. In *Oncopeltus*, preparation for EE rupture within this specialized region involves apoptosis of the amniotic cells subjacent to the serosa, thinning the region to a single EE epithelium, while at the border of this region the amnion adheres strongly to the serosa (Panfilio and Roth, 2010). As this epithelial remodeling occurs locally, the entire egg system is subtly reorganized to ensure that the specialized EE region is centered at the egg pole, which appears to mechanically facilitate rupture via global contractile force exerted by the serosa (Panfilio, 2009; Panfilio and Roth, 2010). In contrast, in *Tribolium* the opening for EE rupture is not centered at the egg pole but occurs anterior-ventrally (Panfilio et al., 2013). Here, precision in determining the site of EE opening involves morphological specialization in a cap of amniotic cells. Furthermore, preparation for rupture in *Tribolium* involves the formation of an amnion-serosa epithelial

bilayer over most of the amnion’s surface area (Koelzer et al., 2015), not just the narrow ring of amnion-serosa contact seen in *Oncopeltus*. These differences in local behavior of the amnion and in the amnion-serosa connection are all the more striking given that *Of-zen* and *Tc-zen2*, the second *Tribolium* paralog, both act extraembryonically to ensure that EE rupture occurs (Figure 1: green diamonds; van der Zee et al., 2005; Panfilio et al., 2006).

In subsequent stages the EE tissues withdraw dorsally, but with the serosa ending up in the tapered dorsal-anterior in *Oncopeltus* compared to the flat dorsal-medial region in *Tribolium* (Figures 2A,E). Nonetheless, in both cases the serosa transforms from a squamous to a columnar epithelium and forms a hollow disc known as the dorsal organ (Panfilio, 2009; Panfilio and Roth, 2010; Panfilio et al., 2013; Figures 2B,C,F,G). Thus, cell shape and intra-tissue organization are conserved despite the geometrical difference resulting from the tissues’ positions within an anisotropic egg system.

However, as a consequence of the manner in which the amnion-serosa connection was prepared for rupture, the inter-tissue organization remains fundamentally different at the dorsal organ stage. The *Oncopeltus* amnion is only connected to the serosa at its margin, and sits on top of the yolk (Figure 2D). While both this attachment point and substrate also apply to the *Tribolium* amnion, the bilayer organization means that additionally a portion of the amnion has the serosa as a substrate (Figure 2H). As the serosa degenerates, tissue continuity over the yolk surface is essential for successful dorsal closure (Panfilio et al., 2013). The planar (lateral-lateral) nature of amnion-serosa attachment in *Oncopeltus* allows the serosa to efficiently pinch off and draw the edges of the amnion together above it (Figure 2B). In *Tribolium*, inter-tissue shearing is required so that the portion of the amnion over the serosa (apical-basal connection) can detach, enabling final serosal internalization (Koelzer et al., 2015).

Altogether, *zen*-mediated rupture, EE contraction and withdrawal, and the cellular structure of the serosal dorsal organ are shared between *Oncopeltus* and *Tribolium* even though the manner of amniotic regionalization (selective apoptosis or morphological alteration) and therefore the nature of the amnion-serosa inter-tissue connection differ.

Conclusions

In this Perspective, we use morphogenesis of the insect EE epithelia to show how different levels of biological organization can provide apparently contradictory signals as to the degree of evolutionary conservation across species. At first glance, these levels are hierarchically ordered, with increasing complexity toward the whole egg system: genes specify cell types and shapes, cells of similar type form tissues, and different tissues shape the whole egg system morphology. However, any pattern of congruence across levels is possible. For example, *zen* orthologs are necessary to specify the serosa in holometabolous insects, but the loss of *zen* function is lethal in some species, while others survive—variously due to serosal dispensability or morphogenetic compensation by the amnion. These phenotypic

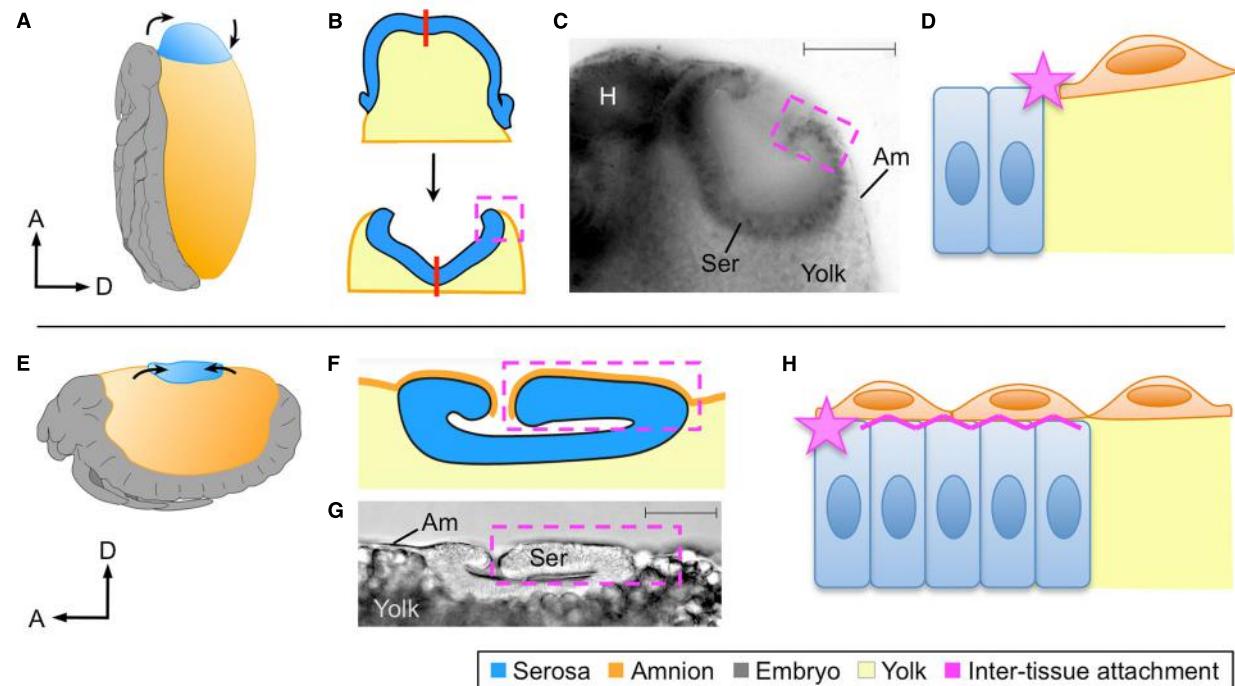


FIGURE 2 | Different tissue organizations achieve the same morphogenetic outcomes. At the dorsal organ stage the serosa compacts into a hollow disc that sinks into the yolk, shown here for *Oncopeltus fasciatus* (A–D) and *Tribolium castaneum* (E–H) as representative of hemi- and holometabolous insects, respectively. Although the site of dorsal organ formation differs in relative position and geometry within the global egg system (A,E), in both cases serosal cells become columnar as the tissue everts (B,C,F,G). The process of serosal eversion is shown schematically in B, where the red line indicates the center of the serosa during this process. In yet another difference, the nature of amnion-serosa attachment consists of a lateral junction within the plane of the extraembryonic epithelium in *Oncopeltus* (D: star), while the bilayered arrangement in *Tribolium* additionally involves basal-basal contact of the two tissues (H: zigzag line). All views are lateral except (B), which is dorsal and omits the embryonic tissue for simplification. Dashed boxes indicate the region of inter-tissue attachment, which is shown schematically at the cellular level in (D) and (H). Micrographs show fixed embryos with a nuclear stain (C) or DIC illumination (D), with scale bars of 100 µm and 50 µm, respectively. Abbreviations: A, anterior; Am, amnion; D, dorsal; H, head; Ser, serosa. Images A,B,C,E are reproduced with minor modification from (Panfilio, 2008, 2009; Panfilio and Roth, 2010; Panfilio et al., 2013), with the corresponding author's consent.

outcomes can be explained by a conserved gene function being embedded in the context of differences in EE tissue complement and topographical configuration across species. In the case of EE epithelial withdrawal in *Oncopeltus* and *Tribolium*, the nature of tissue regionalization and inter-tissue attachment differ dramatically even while gene function, intra-tissue structure, and gross morphogenesis are similar. Only the integration of all biological levels can provide the full picture and give insight into the evolution not just of epithelial morphogenesis but of embryogenesis in general, which ultimately depends on cell shape changes and coordinated tissue reorganization.

In the past, these levels have predominantly been studied separately or in limited combinations. In the watershed Heidelberg screen of *Drosophila* embryonic patterning mutants (Nüsslein-Volhard and Wieschaus, 1980), gene function was linked to final phenotype as determined from larval cuticle preparations, a method still widely employed, especially for large scale screening (Schmitt-Engel et al., 2015). However, even as the initial link between gene and egg system levels is being established, the aim is to refine this information to more precise phenotypic analysis. At this point, a misexpressed gene itself becomes a tool to further explore cell and tissue properties.

With new techniques available, we are increasingly able to investigate multiple levels of biological organization at the same time. For example, gene silencing via RNAi combined with live imaging of fluorescent constructs that afford cell and tissue resolution allows us to visualize the full developmental phenotype resulting from a given genetic manipulation, with *Tribolium* serving as a particularly amenable comparative model among the insects (Sarrazin et al., 2012; Benton et al., 2013; Panfilio et al., 2013; Koelzer et al., 2014). Also, as pioneered in *Drosophila*, mechanical manipulations provide a means of circumventing genetic manipulation when examining cell, tissue, and egg system levels (e.g., Ma et al., 2009; Monier et al., 2010; Wells et al., 2014), and clonal analysis approaches test cellular behaviors at tissue boundaries (Külschammer and Uhlírova, 2013). From all of these studies it becomes increasingly clear that the interplay between the levels is rather similar to a regulatory network (as known from gene interactions), including various interactions and feedback loops, than to a hierarchical structure based on increasing complexity.

Having understood the interplay of the developmental levels within a species, we can now start comparing different species and additional levels. For example, a key aspect of epithelial morphogenesis is the structure of boundaries between different

tissues, where mechanical forces are transmitted and inter-tissue attachments are made. To what extent are mechanical, geometric properties of tissues and the egg system a better predictor than phylogenetic relatedness of how similar two species' morphogenetic processes will be? Moving beyond the confines of the egg system, an even more integrated view of the phenotype can be extended to the influence of the external environment, as addressed in the growing field of eco-evo-devo (Gilbert and Epel, 2008; Abouheif et al., 2014). Ultimately, as the number of comparative animal models and accessibility of experimental tools increases, so too should the sophistication of our phenotypic understanding of how development has evolved.

References

- Abouheif, E., Favé, M.-J., Ibarrarán-Viniegra, A. S., Lesoway, M. P., Rafiqi, A. M., and Rajakumar, R. (2014). Eco-evo-devo: the time has come. *Adv. Exp. Med. Biol.* 781, 107–125. doi: 10.1007/978-94-007-7347-9_6
- Benton, M. A., Akam, M., and Pavlopoulos, A. (2013). Cell and tissue dynamics during *Tribolium castaneum* embryogenesis revealed by versatile fluorescence labeling approaches. *Development* 140, 3210–3220. doi: 10.1242/dev.096271
- Brown, S. J., Fellers, J. P., Shippy, T. D., Richardson, E. A., Maxwell, M., Stuart, J. J., et al. (2002). Sequence of the *Tribolium castaneum* homeotic complex: the region corresponding to the *Drosophila melanogaster* Antennapedia Complex. *Genetics* 160, 1067–1074.
- Chai, C.-L., Zhang, Z., Huang, F.-F., Wang, X.-Y., Yu, Q.-Y., Liu, B.-B., et al. (2008). A genomewide survey of homeobox genes and identification of novel structure of the Hox cluster in the silkworm, *Bombyx mori*. *Insect. Biochem. Mol. Biol.* 38, 1111–1120. doi: 10.1016/j.ibmb.2008.06.008
- Chen, G., Handel, K., and Roth, S. (2000). The maternal NF-kappaB/Dorsal gradient of *Tribolium castaneum*: dynamics of early dorsoventral patterning in a short-germ beetle. *Development* 127, 5145–5156.
- Cook, C. E., Smith, M. L., Telford, M. J., Bastianello, A., and Akam, M. (2001). Hox genes and the phylogeny of the arthropods. *Curr. Biol.* 11, 759–763. doi: 10.1016/S0960-9822(01)00222-6
- Ferguson, L., Marlétaz, F., Carter, J.-M., Taylor, W. R., Gibbs, M., Breuker, C. J., et al. (2014). Ancient expansion of the Hox cluster in Lepidoptera generated four homeobox genes implicated in extra-embryonic tissue formation. *PLoS Genet.* 10:e1004698. doi: 10.1371/journal.pgen.1004698
- Fleig, R., and Sander, K. (1988). Honeybee morphogenesis: embryonic cell movements that shape the larval body. *Development* 103, 525–534.
- Frank, L. H., and Rushlow, C. (1996). A group of genes required for maintenance of the amniocerosa tissue in *Drosophila*. *Development* 122, 1343–1352.
- Gavin-Smyth, J., Wang, Y.-C., Butler, I., and Ferguson, E. L. (2013). A genetic network conferring canalization to a bistable patterning system in *Drosophila*. *Curr. Biol.* 23, 2296–2302. doi: 10.1016/j.cub.2013.09.055
- Gilbert, S. F., and Epel, D. (2008). *Ecological Developmental Biology: Integrating Epigenetics, Medicine, and Evolution*. Sunderland, MA: Sinauer Associates, Inc.
- Hamaguchi, T., Takashima, S., Okamoto, A., Imaoka, M., Okumura, T., and Murakami, R. (2012). Dorsoventral patterning of the *Drosophila* hindgut is determined by interaction of genes under the control of two independent gene regulatory systems, the dorsal and terminal systems. *Mech. Dev.* 129, 236–243. doi: 10.1016/j.mod.2012.07.006
- Hughes, C. L., Liu, P. Z., and Kaufman, T. C. (2004). Expression patterns of the rogue Hox genes Hox3/zen and fushi tarazu in the apterygote insect *Thermobia domestica*. *Evol. Dev.* 6, 393–401. doi: 10.1111/j.1525-142X.2004.04048.x
- Jacinto, A., Wood, W., Balayo, T., Turmaine, M., Martinez-Arias, A., and Martin, P. (2000). Dynamic actin-based epithelial adhesion and cell matching during *Drosophila* dorsal closure. *Curr. Biol.* 10, 1420–1426. doi: 10.1016/S0960-9822(00)00796-X
- Jacobs, C. G. C., Rezende, G. L., Lamers, G. E. M., and van der Zee, M. (2013). The extraembryonic serosa protects the insect egg against desiccation. *Proc. R. Soc. B* 280, 20131082. doi: 10.1098/rspb.2013.1082
- Jacobs, C. G. C., Spaink, H. P., and Zee, M. V. D. (2014). The extraembryonic serosa is a frontier epithelium providing the insect egg with a full-range innate immune response. *eLife* 3, e04111. doi: 10.7554/eLife.04111
- Kiehart, D. P., Galbraith, C. G., Edwards, K. A., Rickolla, W. L., and Montague, R. A. (2000). Multiple forces contribute to cell sheet morphogenesis for dorsal closure in *Drosophila*. *J. Cell Biol.* 149, 471–490. doi: 10.1083/jcb.149.2.471
- Klomp, J., Athy, D., Kwan, C. W., Bloch, N. I., Sandmann, T., Lemke, S., et al. (2015). A cysteine-clamp gene drives embryo polarity in the midge *Chironomus*. *Science* 348, 1040–1042. doi: 10.1126/science.aaa7105
- Koelzer, S., Hilbrant, M., Horn, T., and Panfilio, K. A. (2015). The beetle amnion and serosa functionally interact as apposed epithelia. *bioRxiv* doi: 10.1101/025247
- Koelzer, S., Kölsch, Y., and Panfilio, K. A. (2014). Visualizing late insect embryogenesis: extraembryonic and mesodermal enhancer trap expression in the beetle *Tribolium castaneum*. *PLoS ONE* 9:e103967. doi: 10.1371/journal.pone.0103967
- Krumlauf, R. (1992). Evolution of the vertebrate Hox homeobox genes. *Bioessays* 14, 245–252. doi: 10.1002/bies.950140408
- Külshammer, E., and Uhlírova, M. (2013). The actin cross-linker Filamin/Cheerio mediates tumor malignancy downstream of JNK signaling. *J. Cell Sci.* 126, 927–938. doi: 10.1242/jcs.114462
- Lada, K., Gorfinkel, N., and Arias, A. M. (2012). Interactions between the amniocerosa and epidermis revealed by the function of the u-shaped gene. *Biol. Open* 1, 353–361. doi: 10.1242/bio.2012497
- Ma, X., Lynch, H. E., Scully, P. C., and Hutson, M. S. (2009). Probing embryonic tissue mechanics with laser hole drilling. *Phys. Biol.* 6:036004. doi: 10.1088/1478-3975/6/3/036004
- Monier, B., Pélissiére-Monier, A., Brand, A. H., and Sanson, B. (2010). An actomyosin-based barrier inhibits cell mixing at compartmental boundaries in *Drosophila* embryos. *Nat. Cell Biol.* 12, 60–65. doi: 10.1038/ncb2005
- Nunes da Fonseca, R., van der Zee, M., and Roth, S. (2010). Evolution of extracellular Dpp modulators in insects: the roles of tolloid and twisted-gastrulation in dorsoventral patterning of the *Tribolium* embryo. *Dev. Biol.* 345, 80–93. doi: 10.1016/j.ydbio.2010.05.019
- Nüsslein-Volhard, C., and Wieschaus, E. (1980). Mutations affecting segment number and polarity in *Drosophila*. *Nature* 287, 795–801. doi: 10.1038/287795a0
- Panfilio, K. A. (2008). Extraembryonic development in insects and the acrobatics of blastokinesis. *Dev. Biol.* 313, 471–491. doi: 10.1016/j.ydbio.2007.11.004
- Panfilio, K. A. (2009). Late extraembryonic development and its zen-RNAi-induced failure in the milkweed bug *Oncopeltus fasciatus*. *Dev. Biol.* 333, 297–311. doi: 10.1016/j.ydbio.2009.06.036
- Panfilio, K. A., and Akam, M. (2007). A comparison of Hox3 and Zen protein coding sequences in taxa that span the Hox3/zen divergence. *Dev. Genes Evol.* 217, 323–329. doi: 10.1007/s00427-007-0133-8
- Panfilio, K. A., Liu, P. Z., Akam, M., and Kaufman, T. C. (2006). *Oncopeltus fasciatus* zen is essential for serosal tissue function in katatrepsis. *Dev. Biol.* 292, 226–243. doi: 10.1016/j.ydbio.2005.12.028
- Panfilio, K. A., Oberhofer, G., and Roth, S. (2013). High plasticity in epithelial morphogenesis during insect dorsal closure. *Biol. Open* 2, 1108–1118. doi: 10.1242/bio.20136072
- Panfilio, K. A., and Roth, S. (2010). Epithelial reorganization events during late extraembryonic development in a hemimetabolous insect. *Dev. Biol.* 340, 100–115. doi: 10.1016/j.ydbio.2009.12.034
- Panfilio, K. A., and Roth, S. (2013). Development: getting into the groove, or evolving off the rails? *Curr. Biol.* 23, R1101–R1103. doi: 10.1016/j.cub.2013.10.073

Author Contributions

KAP conceived the idea for the manuscript. KAP, TH, and MH jointly wrote the manuscript.

Acknowledgments

We thank the German Research Foundation (Deutsche Forschungsgemeinschaft) for financial support for conducting some of the experimental investigations described here and during the preparation of this manuscript (grant PA 2044/1-1 to KAP).

- Peel, A. D., Chipman, A. D., and Akam, M. (2005). Arthropod segmentation: beyond the *Drosophila* paradigm. *Nat. Rev. Genet.* 6, 905–916. doi: 10.1038/nrg1724
- Pultz, M. A., Diederich, R. J., Cribbs, D. L., and Kaufman, T. C. (1988). The proboscipedia locus of the Antennapedia complex: a molecular genetic analysis. *Genes Dev.* 2, 901–920. doi: 10.1101/gad.2.7.901
- Rafiqi, A. M. (2008). *Morphological Transitions and the Genetic Basis of the Evolution of Extraembryonic Tissues in Flies* WU Plant Sciences, vol. Ph.D. thesis, Wageningen University, Wageningen.
- Rafiqi, A. M., Lemke, S., Ferguson, S., Stauber, M., and Schmidt-Ott, U. (2008). Evolutionary origin of the amnioserosa in cyclorrhaphan flies correlates with spatial and temporal expression changes of zen. *Proc. Natl Acad. Sci. U.S.A.* 105, 234–239. doi: 10.1073/pnas.0709145105
- Reim, I., and Frasch, M. (2005). The Dorsocross T-box genes are key components of the regulatory network controlling early cardiogenesis in *Drosophila*. *Development* 132, 4911–4925. doi: 10.1242/dev.02077
- Reim, I., Lee, H.-H., and Frasch, M. (2003). The T-box encoding Dorsocross genes function in amnioserosa development and the patterning of the dorsolateral germ band downstream of Dpp. *Development* 130, 3187–3204. doi: 10.1242/dev.00548
- Rezende, G. L., Martins, A. J., Gentile, C., Farnesi, L. C., Pelajo-Machado, M., Peixoto, A. A., et al. (2008). Embryonic desiccation resistance in *Aedes aegypti*: presumptive role of the chitinized serosal cuticle. *BMC Dev. Biol.* 8:82. doi: 10.1186/1471-213X-8-82
- Sachs, L., Chen, Y.-T., Drechsler, A., Lynch, J. A., Panfilio, K. A., Lässig, M., et al. (2015). Dynamic BMP signaling polarized by Toll patterns the dorsoventral axis in a hemimetabolous insect. *eLife* 4:e05502. doi: 10.7554/eLife.05502
- Sarrazin, A. F., Peel, A. D., and Averof, M. (2012). A segmentation clock with two-segment periodicity in insects. *Science* 336, 338–341. doi: 10.1126/science.1218256
- Schmidt-Ott, U., Rafiqi, A. M., and Lemke, S. (2010). “Hox3/zen and the evolution of extraembryonic epithelia in insects,” in *Hox Genes: Studies from the 20th to the 21st Century*, vol. 689, ed. J. S. Deutsch (Austin: Landes Bioscience), 133–144.
- Schmitt-Engel, C., Schultheis, D., Schwirz, J., Ströhlein, N., Troelenberg, N., Majumdar, U., et al. (2015). The iBeetle large scale RNAi screen reveals novel gene functions for insect development and physiology. *Nat. Commun.* 6, 7822. doi: 10.1038/ncomms8822
- Solon, J., Kaya-Çopur, A., Colombelli, J., and Brunner, D. (2009). Pulsed forces timed by a ratchet-like mechanism drive directed tissue movement during dorsal closure. *Cell* 137, 1331–1342. doi: 10.1016/j.cell.2009.03.050
- Sui, L., Pflugfelder, G. O., and Shen, J. (2012). The Dorsocross T-box transcription factors promote tissue morphogenesis in the *Drosophila* wing imaginal disc. *Development* 139, 2773–2782. doi: 10.1242/dev.079384
- van der Zee, M., Berns, N., and Roth, S. (2005). Distinct functions of the *Tribolium* zerknüllt genes in serosa specification and dorsal closure. *Curr. Biol.* 15, 624–636. doi: 10.1016/j.cub.2005.02.057
- Wakimoto, B. T., Turner, F. R., and Kaufman, T. C. (1984). Defects in embryogenesis in mutants associated with the Antennapedia complex of *Drosophila melanogaster*. *Dev. Biol.* 102, 147–172. doi: 10.1016/0012-1606(84)90182-9
- Wells, A. R., Zou, R. S., Tulu, U. S., Sokolow, A. C., Crawford, J. M., Edwards, G. S., et al. (2014). Complete canthi removal reveals that forces from the amnioserosa alone are sufficient to drive dorsal closure in *Drosophila*. *Mol. Cell. Biol.* 25, 3552–3568. doi: 10.1091/mcb.E14-07-1190
- Zeh, D. W., Zeh, J. A., and Smith, R. L. (1989). Ovipositors, amniota and eggshell architecture in the diversification of terrestrial arthropods. *Q. Rev. Biol.* 64, 147–168. doi: 10.1086/416238

Conflict of Interest Statement: 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 © 2015 Horn, Hilbrant and Panfilio. 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) or licensor 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.



From limbs to leaves: common themes in evolutionary diversification of organ form

Remco A. Mentink and Miltos Tsiantis*

Department of Comparative Development and Genetics, Max Planck Institute for Plant Breeding Research, Cologne, Germany

OPEN ACCESS

Edited by:

Sylvain Marcellini,
University of Concepcion, Chile

Reviewed by:

Barbara Ambrose,
The New York Botanical Garden, USA
Cédric Finet,
University of Cambridge, UK

***Correspondence:**

Miltos Tsiantis,
Department of Comparative
Development and Genetics, Max
Planck Institute for Plant Breeding
Research, Carl-von-Linné-Weg 10,
50829 Cologne, Germany
tsiantis@mpipz.mpg.de

Specialty section:

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

Received: 16 June 2015

Accepted: 27 August 2015

Published: 08 September 2015

Citation:

Mentink RA and Tsiantis M (2015)
From limbs to leaves: common
themes in evolutionary
diversification of organ form.
Front. Genet. 6:284.
doi: 10.3389/fgene.2015.00284

An open problem in biology is to derive general principles that capture how morphogenesis evolved to generate diverse forms in different organisms. Here we discuss recent work investigating the morphogenetic basis for digit loss in vertebrate limbs and variation in form of marginal outgrowths of angiosperm (flowering plant) leaves. Two pathways underlie digit loss in vertebrate limbs. First, alterations to digit patterning arise through modification of expression of the *Patched 1* receptor, which senses the *Sonic Hedgehog* morphogen and limits its mobility in the limb bud. Second, evolutionary changes to the degree of programmed cell death between digits influence their development after their initiation. Similarly, evolutionary modification of leaf margin outgrowths occurs via two broad pathways. First, species-specific transcription factor expression modulates outgrowth patterning dependent on regulated transport of the hormone auxin. Second, species-specific expression of the newly discovered REDUCED COMPLEXITY homeodomain transcription factor influences growth between individual outgrowths after their initiation. These findings demonstrate that in both plants and animals tinkering with either patterning or post-patterning processes can cause morphological change. They also highlight the considerable flexibility of morphological evolution and indicate that it may be possible to derive broad principles that capture how morphogenesis evolved across complex eukaryotes.

Keywords: evolution and development, leaflet formation, digit formation, patterning versus post-patterning, morphological diversity

A key question in biology is how morphological diversity is generated. Although plants and animals evolved multicellularity independently, within each kingdom conserved gene regulatory networks (hereafter termed networks) control the development of one or more body parts. In this context evolution operates as a “tinkerer,” being strongly influenced by the materials currently at hand as well as prior history (Jacob, 1977; Davidson and Erwin, 2006; Pajoro et al., 2014; Sorrells et al., 2015). Consequently, considerable constraints exist on the evolution of new traits (Pires-daSilva and Sommer, 2003; Davidson and Erwin, 2006; Carroll, 2008; Pires and Dolan, 2012) raising the question of how evolutionary changes to networks that control development may circumvent these constraints.

Both theoretical arguments and empirical evidence suggest that regulatory sequence variation has greater potential for the generation of morphological change than coding sequence variation (Stern, 2000; Carroll, 2008). This is likely because regulatory sequences tend to be organized in highly modular *cis* elements, leading to their mutation having a lower propensity to generate pleiotropic effects that would compromise development (Stern, 2000; Carroll, 2008; Rebeiz et al., 2015). However, to what degree this broad principle manifests itself in different

evolutionary lineages and how precisely the balance of conservation versus divergence of different networks creates morphological diversity remain open questions (Stern and Orgogozo, 2008, 2009; Rebeiz et al., 2015).

Other than knowing the types of genetic changes underlying the generation of morphological diversity, an understanding of evolution requires determining how, when and where those genetic changes influence morphogenesis. For example, it remains largely unclear if particular stages or aspects of development tend to be preferentially amenable to evolutionary tinkering. Does evolution primarily target developmental processes that are active during early stages of organ development; or are such early stages less favored by evolution, owing to the potential risk of causing pleiotropic effects that will influence later stages of development? Does diversity largely arise through tinkering with later acting developmental programs that fine-tune organ form after the more fundamental patterns have been laid down? Studies on the emergence of novel insect pigmentation patterns in closely related species suggest that later developmental stages (e.g., the insect pupal stage) might be more readily available for evolutionary tinkering (Wittkopp and Beldade, 2009). However, definitive answers to these questions are likely still to come and will depend on the particularities of the system that is under investigation, including its evolutionary history, its modularity, the type of trait being studied and its degree of integration with other traits. Nevertheless, one way to approach these problems in a unified fashion when comparing diverse organisms is to consider whether and how evolution influences patterning and post-patterning modes of development. Patterning processes act to impart positional information, for example through the use of morphogen concentration gradients, and facilitate correct distribution of cellular identities within tissues (Kondo and Miura, 2010; Rogers and Schier, 2011). Post-patterning processes, on the other hand, serve to sculpt emerging tissues and organs typically after their identity has previously been determined. For example, post-patterning processes may operate by removing superfluous cells through apoptosis or by adjusting the growth rates of specific populations of cells within the organ (Coen et al., 2004; Suzanne and Steller, 2013). Notably, this distinction between “post-patterning” and patterning does not exclude the possibility that patterning genes may have persistent effects in developmental time, including post-patterning stages (Salazar-Ciudad et al., 2003; McGregor et al., 2007; Werner et al., 2010).

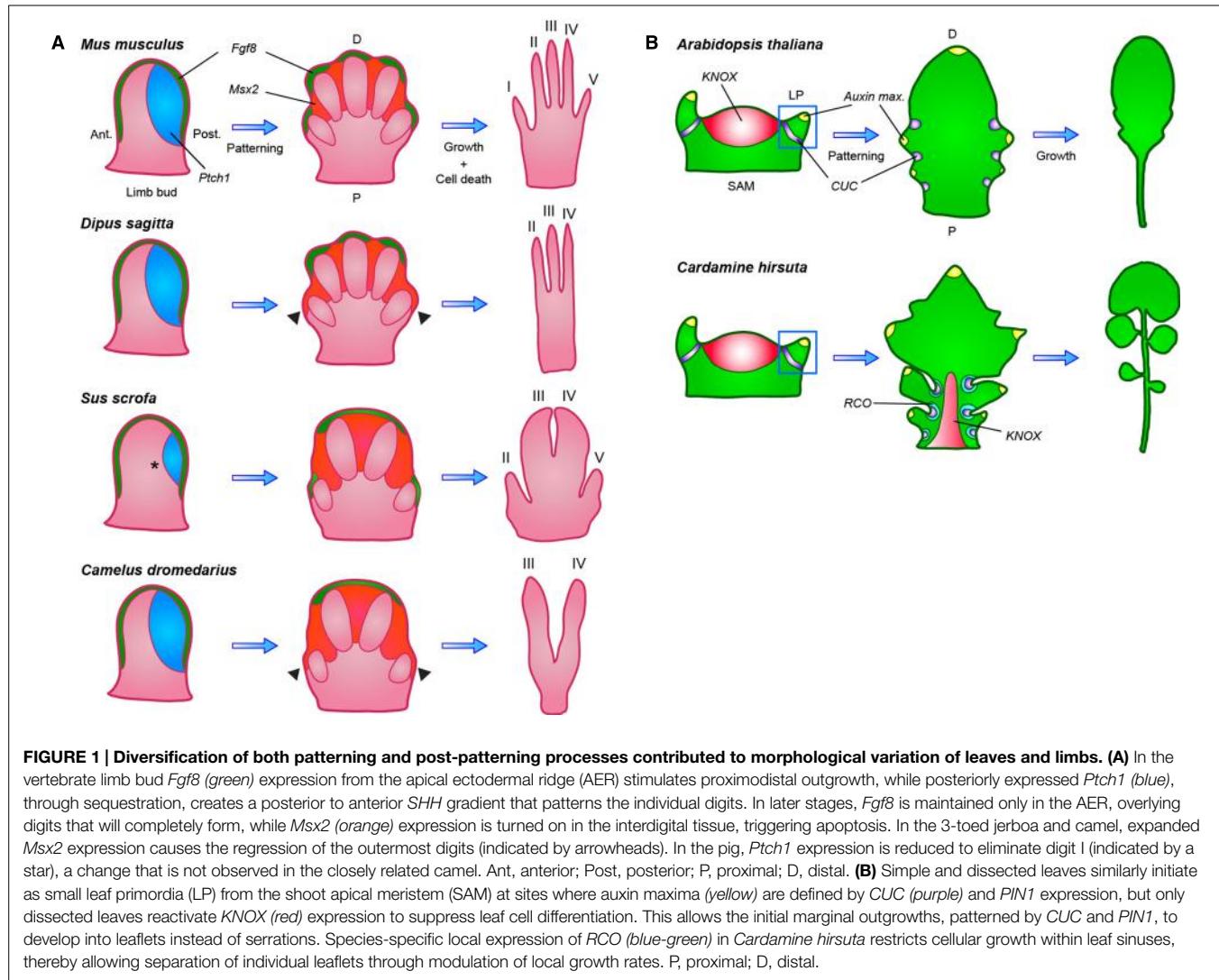
Two recent papers have explored the significance of patterning versus post-patterning events on development by studying digit loss in mammals and leaf shape formation in angiosperms and revealed a strong link between altered, species-specific gene expression domains and morphological variation. Both studies suggest considerable versatility in how evolutionary tinkering with developmental processes can ultimately arrive at similar phenotypes.

Cooper et al. (2014) studied the evolutionary changes that resulted in convergent digit loss in different mammalian species. A mammalian limb (such as a leg) is attached to the body at one (proximal) end and has 1 to 5 anteroposteriorly distinct digits (e.g., toes) at the other (distal) end. Limbs develop from the limb bud through the sequential action of several distinct signaling centers

(**Figure 1A**; Butterfield et al., 2010). Bone morphogenetic proteins (BMPs) specify the formation of the apical ectodermal ridge (AER) at the distal end of the limb bud, from which fibroblast growth factors (FGFs) are secreted to stimulate proximodistal outgrowth (Lewandoski et al., 2000; Pizette et al., 2001; Boulet et al., 2004). The morphogen Sonic hedgehog (SHH) is secreted from the posterior limb bud to direct both digit patterning and expansion of the hand- or footplate to accommodate all digits (Harfe et al., 2004; Towers et al., 2008). Subsequent digit elongation is controlled by FGFs secreted from the AER and in later stages BMPs sculpt the limb by inducing apoptotic cell death within interdigital tissue in concert with the transcription factor *Msx2* (Marazzi et al., 1997; Ferrari et al., 1998; Sanz-Ezquerro and Tickle, 2003).

Cooper et al. (2014) studied the possible relevance of these pathways to digit loss in 3- and 5-toed jerboas (small, desert-dwelling rodents that develop a varying number of digits on the hind limb between different species) and mice, as well as in ungulates (hoofed animals with 1 to 4 toes). In jerboas, they observed no differences in patterning gene expression but rather found expanded domains of apoptotic cells in 3-toed jerboa hind limbs, surrounding tissue otherwise destined to form digits I and V. Specifically, they found that expression of *Msx2* was expanded in the 3-toed jerboa hind limb, likely causing increased cell death (**Figure 1A**). They obtained comparable results with 1-toed horse embryos, where *Msx2* expansion correlated with removal of digits II and IV. This shows a convergent evolutionary event in which an apoptotic pathway normally used to remove interdigital tissue was co-opted by regulatory changes to act in truncating digit outgrowth.

By expanding their study to even-toed ungulate species, Cooper et al. (2014) found a striking flexibility in modes of digit loss. In pigs the expression of *Patched 1* (*Ptch1*), a *Shh* receptor, is reduced toward the posterior limb bud. *Ptch1* acts to restrict the spread of *Shh* by sequestration, thus reduction in *Ptch1* expression leads to an expanded region of *Shh* activity and more uniform expression of its target genes, presumably causing a shift in limb axis symmetry to the space between digits III and IV (**Figure 1A**; Chen and Struhl, 1996; Butterfield et al., 2009). These findings were corroborated by a second group that showed a similar reduction of *Ptch1* expression in cow limb buds (Lopez-Rios et al., 2014). These authors also demonstrated that *cis* regulatory divergence of *Ptch1* renders it unresponsive to *Shh* signaling in a negative feedback loop. Remarkably, when Cooper et al. (2014) examined embryos of camel, a third ungulate, they observed no modification of *Ptch1* expression, but instead an expansion of apoptosis and *Msx2* expression, resembling the case in 3-toed jerboas and horses (**Figure 1A**; Cooper et al., 2014). These results indicate that in species of the same taxonomic order, such as camels and pigs (both members of the Artiodactyla or even-toed ungulates), fundamentally different mechanisms can be modified to achieve similar organ modifications, revealing considerable flexibility in evolutionary pathways. A conclusion that is additionally in line with the fact that Cooper et al. (2014) did not recover any evidence for evolutionary tinkering with the *HoxD* regulatory landscape, which has previously been identified and hypothesized to be a good candidate for vertebrate digit



diversification, owing to its highly modular nature (Montavon et al., 2011). These findings raise the question of how broadly this flexibility in evolutionary tinkering with either growth or patterning occurred during evolution of complex eukaryotes.

A recent paper by Vlad et al. (2014) establishes that a comparable logic helps explain diversification of leaf shapes in plants of the Brassicaceae family. Brassicaceae, like other flowering plants, form either simple leaves, consisting of an entire blade with smooth, serrated or lobed margins, or dissected leaves, comprising individual leaflets. Both types of leaves develop from leaf primordia that initiate from the pluripotent shoot apical meristem (SAM). *KNOTTED-LIKE HOMEobox (KNOX)* transcription factors are expressed in the meristem to maintain its organ-generating potential (Hay and Tsiantis, 2006; Barkoulas et al., 2007). Transport of the plant hormone auxin through the *PIN-FORMED 1 (PIN1)* efflux transporter, coupled to a self-reinforcing feedback of auxin on PIN1 expression and polarization, likely creates sequential local auxin activity maxima at the flanks of the SAM. This process appears to be self-organizing and the resulting auxin maxima are required for sequential

primordium development (Reinhardt et al., 2003; Heisler et al., 2005; Jonsson et al., 2006; Smith et al., 2006). *CUP-SHAPED COTYLEDON (CUC)* genes mark the leaf primordium boundary and allow its separation by repressing growth at the flanks (Aida et al., 1997; Hibara et al., 2006). *CUCs* and *PIN1* also function together to pattern the leaf margin, as *CUCs* likely repress growth at the boundaries of serrations or leaflets, while *PIN1* generates auxin maxima at the sites of their outgrowth. Notably, in this context *CUCs* likely both repress growth at the flanks of marginal outgrowths and promote their outgrowth at least in part via promoting generation of an auxin maximum at their tip (Nikovics et al., 2006; Barkoulas et al., 2008; Blein et al., 2008; Koenig et al., 2009; Kawamura et al., 2010; Ben-Gera et al., 2012). In the *Arabidopsis thaliana* leaf margin, *CUC2* directs *PIN1* localization to form local auxin maxima while auxin feeds back to repress *CUC2*, creating the repeated pattern of leaf serrations along the leaf margin (Bilsborough et al., 2011). *KNOX* genes, then, are expressed in dissected leaves and differentiate these from simple leaves by retarding cellular differentiation, thus rendering the leaf competent to form leaflets in response to *PIN1*

dependent auxin maxima (Hay and Tsiantis, 2006; Barkoulas et al., 2008; Kimura et al., 2008; Bar and Ori, 2014). Similarly *CUC1*, a redundantly acting parologue of *CUC2*, is expressed in the dissected leaves of *Cardamine hirsuta* but is confined to the leaf meristem boundary in its simple-leaved relative *A. thaliana*. These observations indicate that evolutionary tinkering with auxin-based patterning mechanisms through alterations in expression of upstream transcription factors such as *KNOX* and *CUC* may be a major route for generating diversity in leaf shapes (**Figure 1B**; Barkoulas et al., 2008; Blein et al., 2008; Piazza et al., 2010; Hasson et al., 2011; Finet and Jaillais, 2012; Bar and Ori, 2014).

Until recently, no genes had been identified that specifically influence leaflet formation without also affecting meristem function or leaf initiation. Such findings suggested that leaflets form through the redeployment of processes that acted earlier in development during leaf initiation (Bar and Ori, 2014; Vlad et al., 2014). To identify novel regulators of leaf complexity, Vlad et al. (2014) conducted a forward genetic screen for genes required for leaflet formation in *C. hirsuta*. They identified the *REDUCED COMPLEXITY (RCO)* homeobox gene, of which a loss of function allele simplifies the leaf without causing pleiotropic phenotypes, suggesting a specific requirement for *RCO* in leaflet formation. *RCO* evolved in the Brassicaceae family from a gene duplication of *LATE MERISTEM IDENTITY 1 (LMI1)*; originally identified in *A. thaliana* as a floral regulator (Saddic et al., 2006). They found that *RCO* is specifically expressed at the base of leaflets (**Figure 1B**), while *LMI1* is expressed more distally, in a complementary pattern, along the leaf margins. *RCO* does not appear to influence PIN1-mediated auxin patterning, but instead functions by repressing cellular growth between individual leaflets in *C. hirsuta*, a post-patterning process that allows leaflet separation. *RCO* was lost in *A. thaliana* during evolution, contributing to its leaf simplification, but re-introducing *RCO* into *A. thaliana* drives expression in basal regions of the leaf and increases leaf complexity, partially reversing the consequences of evolution. These results, together with a follow-up study in the sister species *Capsella rubella* and *Capsella grandiflora* by Sicard et al. (2014), suggest that *RCO* is a key regulator of leaf shape and diversity in the Brassicaceae and provide a striking example of organ shape diversification by tinkering with local growth regulation at the flanks of a growing organ primordium (Vlad et al., 2014). Another notable aspect of the *RCO* study is that this gene was discovered through performing a forward genetics study in *C. hirsuta* and could not have been found in *A. thaliana*, where the gene has been lost, thus highlighting the importance of unbiased studies in diverse taxa for understanding the genetic basis for the evolution of form.

Taken together, these two studies illustrate how evolution can exploit both patterning and post-patterning processes to

create morphological diversity in both plants and animals. It will be interesting to explore whether bias might exist for variations of either kind or for particular developmental pathways across different kingdoms. For example, plants and animals have evolved distinct biophysical properties and morphogenetic strategies that pose different constraints for evolution. Whereas animal morphogenesis involves the use of large-scale apoptosis and cell migration, these mechanisms are used to more limited extent (Gunawardena, 2008; Fendrych et al., 2014) or not at all respectively in plants. This is because rigid cell walls in plants somewhat complicate the use of both these mechanisms during development: cell walls typically remain after apoptosis, thereby constraining developmental options, while they make cell migration impossible by preventing the sliding of cells alongside each other. These fundamental differences in the cellular underpinnings of development suggest that morphological diversity in plants mostly arises through tinkering with regional growth rates and growth directionality (Coen et al., 2004), consistent with the findings of Vlad et al. (2014). These particularities of plants, however, do not preclude that changes to such growth-related processes can also contribute to the evolution of animal form (Abzhanov et al., 2004; Wu et al., 2004). In any event, independent of the organism studied, morphology is determined by processes that take place at different levels of organization and yield the final form through complex feedback loops of genetic regulation, signaling and tissue growth (Salazar-Ciudad and Jernvall, 2010; Kennaway et al., 2011; Prusinkiewicz and Runions, 2012). Conceptualizing how activity of gene regulatory networks creates organ shape is consequently not solely intuitive. The computer science and developmental biology interface offers a promising path for resolving such problems in a predictive fashion (Lewis, 2008; Green et al., 2010; Prusinkiewicz and Runions, 2012; Sheth et al., 2012). Quantitative investigations of morphogenesis and the genetic basis of its variation in different organismal lineages will allow us to build a general picture of how organ diversity is generated and maintained. Such studies should also help us understand the basis for and limits of predictability of morphological evolution.

Acknowledgments

Work in the Tsiantis lab on diversification of leaf shape is supported by a Deutsche Forschungsgemeinschaft (DFG) “Adaptomics” grant TS 229/1-1, a DFG Collaborative Research Center SFB 680 grant on Evolutionary Innovations and a core grant from the Max Planck Society. MT also acknowledges support of the Cluster of Excellence on Plant Sciences. We thank Sheila McCormick for comments on the manuscript.

References

- Abzhanov, A., Protas, M., Grant, B. R., Grant, P. R., and Tabin, C. J. (2004). Bmp4 and morphological variation of beaks in Darwin's finches. *Science* 305, 1462–1465. doi: 10.1126/science.1098095
- Aida, M., Ishida, T., Fukaki, H., Fujisawa, H., and Tasaka, M. (1997). Genes involved in organ separation in *Arabidopsis*: an analysis of the cup-shaped cotyledon mutant. *Plant Cell* 9, 841–857. doi: 10.1105/tpc.9.6.841
- Barkoulas, M., Galinha, C., Grigg, S. P., and Tsiantis, M. (2007). From genes to shape: regulatory interactions in leaf development. *Curr. Opin. Plant Biol.* 10, 660–666. doi: 10.1016/j.pbi.2007.07.012
- Barkoulas, M., Hay, A., Kougoumoutzi, E., and Tsiantis, M. (2008). A developmental framework for dissected leaf formation in the *Arabidopsis* relative *Cardamine hirsuta*. *Nat. Genet.* 40, 1136–1141. doi: 10.1038/ng.189
- Bar, M., and Ori, N. (2014). Leaf development and morphogenesis. *Development* 141, 4219–4230. doi: 10.1242/dev.106195

- Ben-Gera, H., Shwartz, I., Shao, M. R., Shani, E., Estelle, M., and Ori, N. (2012). ENTIRE and GOBLET promote leaflet development in tomato by modulating auxin response. *Plant J.* 70, 903–915. doi: 10.1111/j.1365-313X.2012.04939.x
- Bilsborough, G. D., Runions, A., Barkoulas, M., Jenkins, H. W., Hasson, A., Galinha, C., et al. (2011). Model for the regulation of *Arabidopsis thaliana* leaf margin development. *Proc. Natl. Acad. Sci. U.S.A.* 108, 3424–3429. doi: 10.1073/pnas.1015162108
- Blein, T., Pulido, A., Viallette-Guiraud, A., Nikovics, K., Morin, H., Hay, A., et al. (2008). A conserved molecular framework for compound leaf development. *Science* 322, 1835–1839. doi: 10.1126/science.1166168
- Boulet, A. M., Moon, A. M., Arenkiel, B. R., and Capecci, M. R. (2004). The roles of Fgf4 and Fgf8 in limb bud initiation and outgrowth. *Dev. Biol.* 273, 361–372. doi: 10.1016/j.ydbio.2004.06.012
- Butterfield, N. C., Mcglinn, E., and Wicking, C. (2010). The molecular regulation of vertebrate limb patterning. *Curr. Top. Dev. Biol.* 90, 319–341. doi: 10.1016/S0070-2153(10)90009-4
- Butterfield, N. C., Metzis, V., Mcglinn, E., Bruce, S. J., Wainwright, B. J., and Wicking, C. (2009). Patched 1 is a crucial determinant of asymmetry and digit number in the vertebrate limb. *Development* 136, 3515–3524. doi: 10.1242/dev.037507
- Carroll, S. B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134, 25–36. doi: 10.1016/j.cell.2008.06.030
- Chen, Y., and Struhl, G. (1996). Dual roles for patched in sequestering and transducing Hedgehog. *Cell* 87, 553–563. doi: 10.1016/S0092-8674(00)81374-4
- Coen, E., Rolland-Lagan, A. G., Matthews, M., Bangham, J. A., and Prusinkiewicz, P. (2004). The genetics of geometry. *Proc. Natl. Acad. Sci. U.S.A.* 101, 4728–4735. doi: 10.1073/pnas.0306308101
- Cooper, K. L., Sears, K. E., Uygur, A., Maier, J., Baczkowski, K. S., Brosnahan, M., et al. (2014). Patterning and post-patterning modes of evolutionary digit loss in mammals. *Nature* 511, 41–45. doi: 10.1038/nature13496
- Davidson, E. H., and Erwin, D. H. (2006). Gene regulatory networks and the evolution of animal body plans. *Science* 311, 796–800. doi: 10.1126/science.1113832
- Fendrych, M., Van Hautegem, T., Van Durme, M., Olvera-Carrillo, Y., Huysmans, M., Karimi, M., et al. (2014). Programmed cell death controlled by ANAC033/SOMBRERO determines root cap organ size in *Arabidopsis*. *Curr. Biol.* 24, 931–940. doi: 10.1016/j.cub.2014.03.025
- Ferrari, D., Lichtler, A. C., Pan, Z. Z., Dealy, C. N., Upholt, W. B., and Kosher, R. A. (1998). Ectopic expression of Msx-2 in posterior limb bud mesoderm impairs limb morphogenesis while inducing BMP-4 expression, inhibiting cell proliferation, and promoting apoptosis. *Dev. Biol.* 197, 12–24. doi: 10.1006/dbio.1998.8880
- Finet, C., and Jaillais, Y. (2012). Auxology: when auxin meets plant evo-devo. *Dev. Biol.* 369, 19–31. doi: 10.1016/j.ydbio.2012.05.039
- Green, A. A., Kennaway, J. R., Hanna, A. I., Bangham, J. A., and Coen, E. (2010). Genetic control of organ shape and tissue polarity. *PLoS Biol.* 8:e1000537. doi: 10.1371/journal.pbio.1000537
- Gunawardena, A. H. (2008). Programmed cell death and tissue remodelling in plants. *J. Exp. Bot.* 59, 445–451. doi: 10.1093/jxb/erm189
- Harfe, B. D., Scherz, P. J., Nissim, S., Tian, H., McMahon, A. P., and Tabin, C. J. (2004). Evidence for an expansion-based temporal Shh gradient in specifying vertebrate digit identities. *Cell* 118, 517–528. doi: 10.1016/j.cell.2004.07.024
- Hasson, A., Plessis, A., Blein, T., Adroher, B., Grigg, S., Tsiantis, M., et al. (2011). Evolution and diverse roles of the CUP-SHAPED COTYLEDON genes in *Arabidopsis* leaf development. *Plant Cell* 23, 54–68. doi: 10.1105/tpc.110.081448
- Hay, A., and Tsiantis, M. (2006). The genetic basis for differences in leaf form between *Arabidopsis thaliana* and its wild relative *Cardamine hirsuta*. *Nat. Genet.* 38, 942–947. doi: 10.1038/ng1835
- Heisler, M. G., Ohno, C., Das, P., Sieber, P., Reddy, G. V., Long, J. A., et al. (2005). Patterns of auxin transport and gene expression during primordium development revealed by live imaging of the *Arabidopsis* inflorescence meristem. *Curr. Biol.* 15, 1899–1911. doi: 10.1016/j.cub.2005.09.052
- Hibara, K., Karim, M. R., Takada, S., Taoka, K., Furutani, M., Aida, M., et al. (2006). *Arabidopsis* CUP-SHAPED COTYLEDON3 regulates postembryonic shoot meristem and organ boundary formation. *Plant Cell* 18, 2946–2957. doi: 10.1105/tpc.106.045716
- Jacob, F. (1977). Evolution and tinkering. *Science* 196, 1161–1166. doi: 10.1126/science.860134
- Jonsson, H., Heisler, M. G., Shapiro, B. E., Meyerowitz, E. M., and Mjolsness, E. (2006). An auxin-driven polarized transport model for phyllotaxis. *Proc. Natl. Acad. Sci. U.S.A.* 103, 1633–1638. doi: 10.1073/pnas.0509839103
- Kawamura, E., Horiguchi, G., and Tsukaya, H. (2010). Mechanisms of leaf tooth formation in *Arabidopsis*. *Plant J.* 62, 429–441. doi: 10.1111/j.1365-313X.2010.04156.x
- Kennaway, R., Coen, E., Green, A., and Bangham, A. (2011). Generation of diverse biological forms through combinatorial interactions between tissue polarity and growth. *PLoS Comput. Biol.* 7:e1002071. doi: 10.1371/journal.pcbi.1002071
- Kimura, S., Koenig, D., Kang, J., Yoong, F. Y., and Sinha, N. (2008). Natural variation in leaf morphology results from mutation of a novel KNOX gene. *Curr. Biol.* 18, 672–677. doi: 10.1016/j.cub.2008.04.008
- Koenig, D., Bayer, E., Kang, J., Kuhlemeier, C., and Sinha, N. (2009). Auxin patterns *Solanum lycopersicum* leaf morphogenesis. *Development* 136, 2997–3006. doi: 10.1242/dev.033811
- Kondo, S., and Miura, T. (2010). Reaction-diffusion model as a framework for understanding biological pattern formation. *Science* 329, 1616–1620. doi: 10.1126/science.1179047
- Lewandoski, M., Sun, X., and Martin, G. R. (2000). Fgf8 signalling from the AER is essential for normal limb development. *Nat. Genet.* 26, 460–463. doi: 10.1038/82609
- Lewis, J. (2008). From signals to patterns: space, time, and mathematics in developmental biology. *Science* 322, 399–403. doi: 10.1126/science.1166154
- Lopez-Rios, J., Duchesne, A., Speziale, D., Andrey, G., Peterson, K. A., Germann, P., et al. (2014). Attenuated sensing of SHH by Ptch1 underlies evolution of bovine limbs. *Nature* 511, 46–51. doi: 10.1038/nature13289
- Marazzi, G., Wang, Y., and Sassoon, D. (1997). Msx2 is a transcriptional regulator in the BMP4-mediated programmed cell death pathway. *Dev. Biol.* 186, 127–138. doi: 10.1006/dbio.1997.8576
- McGregor, A. P., Orgogozo, V., Delon, I., Zanet, J., Srinivasan, D. G., Payre, F., et al. (2007). Morphological evolution through multiple cis-regulatory mutations at a single gene. *Nature* 448, 587–590. doi: 10.1038/nature05988
- Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., et al. (2011). A regulatory archipelago controls Hox genes transcription in digits. *Cell* 147, 1132–1145. doi: 10.1016/j.cell.2011.10.023
- Nikovics, K., Blein, T., Peaucelle, A., Ishida, T., Morin, H., Aida, M., et al. (2006). The balance between the MIR164A and CUC2 genes controls leaf margin serration in *Arabidopsis*. *Plant Cell* 18, 2929–2945. doi: 10.1105/tpc.106.045617
- Pajoro, A., Biewers, S., Dougal, E., Leal Valentim, F., Mendes, M. A., Porri, A., et al. (2014). The (r)evolution of gene regulatory networks controlling *Arabidopsis* plant reproduction: a two-decade history. *J. Exp. Bot.* 65, 4731–4745. doi: 10.1093/jxb/eru233
- Piazza, P., Bailey, C. D., Cartolano, M., Krieger, J., Cao, J., Ossowski, S., et al. (2010). *Arabidopsis thaliana* leaf form evolved via loss of KNOX expression in leaves in association with a selective sweep. *Curr. Biol.* 20, 2223–2228. doi: 10.1016/j.cub.2010.11.037
- Pires-daSilva, A., and Sommer, R. J. (2003). The evolution of signalling pathways in animal development. *Nat. Rev. Genet.* 4, 39–49. doi: 10.1038/nrg977
- Pires, N. D., and Dolan, L. (2012). Morphological evolution in land plants: new designs with old genes. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 367, 508–518. doi: 10.1098/rstb.2011.0252
- Pizette, S., Abate-Shen, C., and Niswander, L. (2001). BMP controls proximodistal outgrowth, via induction of the apical ectodermal ridge, and dorsoventral patterning in the vertebrate limb. *Development* 128, 4463–4474.
- Prusinkiewicz, P., and Runions, A. (2012). Computational models of plant development and form. *New Phytol.* 193, 549–569. doi: 10.1111/j.1469-8137.2011.04009.x
- Rebeiz, M., Patel, N. H., and Hinman, V. F. (2015). Unraveling the tangled skein: the evolution of transcriptional regulatory networks in development. *Annu. Rev. Genomics Hum. Genet.* doi: 10.1146/annurev-genom-091212-153423 [Epub ahead of print].
- Reinhardt, D., Pesce, E. R., Stieger, P., Mandel, T., Baltensperger, K., Bennett, M., et al. (2003). Regulation of phyllotaxis by polar auxin transport. *Nature* 426, 255–260. doi: 10.1038/nature02081
- Rogers, K. W., and Schier, A. F. (2011). Morphogen gradients: from generation to interpretation. *Annu. Rev. Cell Dev. Biol.* 27, 377–407. doi: 10.1146/annurev-cellbio-092910-154148

- Saddic, L. A., Huvermann, B., Bezhani, S., Su, Y., Winter, C. M., Kwon, C. S., et al. (2006). The LEAFY target LM11 is a meristem identity regulator and acts together with LEAFY to regulate expression of CAULIFLOWER. *Development* 133, 1673–1682. doi: 10.1242/dev.02331
- Salazar-Ciudad, I., and Jernvall, J. (2010). A computational model of teeth and the developmental origins of morphological variation. *Nature* 464, 583–586. doi: 10.1038/nature08838
- Salazar-Ciudad, I., Jernvall, J., and Newman, S. A. (2003). Mechanisms of pattern formation in development and evolution. *Development* 130, 2027–2037. doi: 10.1242/dev.00425
- Sanz-Ezquerro, J. J., and Tickle, C. (2003). Fgf signaling controls the number of phalanges and tip formation in developing digits. *Curr. Biol.* 13, 1830–1836. doi: 10.1016/j.cub.2003.09.040
- Sheth, R., Marcon, L., Bastida, M. F., Junco, M., Quintana, L., Dahn, R., et al. (2012). Hox genes regulate digit patterning by controlling the wavelength of a Turing-type mechanism. *Science* 338, 1476–1480. doi: 10.1126/science.1226804
- Sicard, A., Thamm, A., Marona, C., Lee, Y. W., Wahl, V., Stinchcombe, J. R., et al. (2014). Repeated evolutionary changes of leaf morphology caused by mutations to a homeobox gene. *Curr. Biol.* 24, 1880–1886. doi: 10.1016/j.cub.2014.06.061
- Smith, R. S., Guyomarc'h, S., Mandel, T., Reinhardt, D., Kuhlemeier, C., and Prusinkiewicz, P. (2006). A plausible model of phyllotaxis. *Proc. Natl. Acad. Sci. U.S.A.* 103, 1301–1306. doi: 10.1073/pnas.0510457103
- Sorrells, T. R., Booth, L. N., Tuch, B. B., and Johnson, A. D. (2015). Intersecting transcription networks constrain gene regulatory evolution. *Nature* 523, 361–365. doi: 10.1038/nature14613
- Stern, D. L. (2000). Evolutionary developmental biology and the problem of variation. *Evolution* 54, 1079–1091. doi: 10.1111/j.0014-3820.2000.tb00544.x
- Stern, D. L., and Orgogozo, V. (2008). The loci of evolution: how predictable is genetic evolution? *Evolution* 62, 2155–2177. doi: 10.1111/j.1558-5646.2008.00450.x
- Stern, D. L., and Orgogozo, V. (2009). Is genetic evolution predictable? *Science* 323, 746–751. doi: 10.1126/science.1158997
- Suzanne, M., and Steller, H. (2013). Shaping organisms with apoptosis. *Cell Death Differ.* 20, 669–675. doi: 10.1038/cdd.2013.11
- Towers, M., Mahood, R., Yin, Y., and Tickle, C. (2008). Integration of growth and specification in chick wing digit-patterning. *Nature* 452, 882–886. doi: 10.1038/nature06718
- Vlad, D., Kierzkowski, D., Rast, M. I., Vuolo, F., Dello Ioio, R., Galinha, C., et al. (2014). Leaf shape evolution through duplication, regulatory diversification, and loss of a homeobox gene. *Science* 343, 780–783. doi: 10.1126/science.1248384
- Werner, T., Koshikawa, S., Williams, T. M., and Carroll, S. B. (2010). Generation of a novel wing colour pattern by the Wingless morphogen. *Nature* 464, 1143–1148. doi: 10.1038/nature08896
- Wittkopp, P. J., and Beldade, P. (2009). Development and evolution of insect pigmentation: genetic mechanisms and the potential consequences of pleiotropy. *Semin. Cell Dev. Biol.* 20, 65–71. doi: 10.1016/j.semcd.2008.10.002
- Wu, P., Jiang, T. X., Suksaweang, S., Widelitz, R. B., and Chuong, C. M. (2004). Molecular shaping of the beak. *Science* 305, 1465–1466. doi: 10.1126/science.1098109

Conflict of Interest Statement: 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 © 2015 Mentink and Tsiantis. 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) or licensor 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.

The differential view of genotype–phenotype relationships

Virginie Orgogozo^{1*}, Baptiste Morizot² and Arnaud Martin³

¹ CNRS, UMR 7592, Institut Jacques Monod, Université Paris Diderot, Paris, France, ² Aix Marseille Université, CNRS, CEPERC UMR 7304, Aix en Provence, France, ³ Department of Molecular Cell Biology, University of California, Berkeley, CA, USA

An integrative view of diversity and singularity in the living world requires a better understanding of the intricate link between genotypes and phenotypes. Here we re-emphasize the old standpoint that the genotype–phenotype (GP) relationship is best viewed as a connection between two differences, one at the genetic level and one at the phenotypic level. As of today, predominant thinking in biology research is that multiple genes interact with multiple environmental variables (such as abiotic factors, culture, or symbionts) to produce the phenotype. Often, the problem of linking genotypes and phenotypes is framed in terms of genotype and phenotype maps, and such graphical representations implicitly bring us away from the differential view of GP relationships. Here we show that the differential view of GP relationships is a useful explanatory framework in the context of pervasive pleiotropy, epistasis, and environmental effects. In such cases, it is relevant to view GP relationships as differences embedded into differences. Thinking in terms of differences clarifies the comparison between environmental and genetic effects on phenotypes and helps to further understand the connection between genotypes and phenotypes.

OPEN ACCESS

Edited by:

Sylvain Marcellini,
 University of Concepcion, Chile

Reviewed by:

Jacob A. Tennesen,
 Oregon State University, USA
 Margarida Matos,
 University of Lisbon, Portugal

*Correspondence:

Virginie Orgogozo,
 CNRS, UMR 7592, Institut Jacques
 Monod, Université Paris Diderot,
 15 Rue Hélène Brion, 75013 Paris,
 France
 virginie.orgogozo@normalesup.org

Specialty section:

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

Received: 24 March 2015

Accepted: 28 April 2015

Published: 19 May 2015

Citation:

Orgogozo V, Morizot B and Martin A
 (2015) The differential view
 of genotype–phenotype relationships.
Front. Genet. 6:179.

doi: 10.3389/fgene.2015.00179

Introduction

We sometimes seem to have forgotten that the original question in genetics was not what makes a protein but rather ‘what makes a dog a dog, a man a man.’

(Noble, 2006)

One fundamental question in biology is to understand what makes individuals, populations, and species different from each other. The concept of *phenotype*, which corresponds to the observable attributes of an individual, was coined in opposition to the *genotype*, the inherited material transmitted by gametes. Since the early proposal that genotypes and phenotypes form two fundamentally different levels of biological abstraction (Johannsen, 1911), the challenge has been to understand how they articulate with each other, how genotypes map onto phenotypes. In the last 15 years, more than 1,000 examples of DNA sequence changes have been linked to naturally occurring non-deleterious phenotypic differences between individuals or species in Eukaryotes (Martin and Orgogozo, 2013b). In human, the OMIM® catalog (Online Mendelian Inheritance in Man, <http://omim.org/>) compiling the genetic determinants of disease-related phenotypes totals more than 4,300 entries and a total of 2,493 published Genome-Wide Association Studies (GWAS) have been uncovering a wealth of sites in the genome that are statistically associated to complex traits

Abbreviations: GP, genotype–phenotype.

(Welter et al., 2014). As the detection of causal links between genetic and phenotypic variation is accelerating, a reexamination of our conceptual tools may help us in finding unifying principles within the swarm of data. Here we reflect on the relationship between genotypes and phenotypes and we address this essay to biologists who are willing to try to challenge their current understanding of phenotypes. We single out one useful point of view, the differential view. We then show that this simple framework remains insightful in the context of pervasive pleiotropy, epistasis, and environmental effects.

Genes as Difference Makers

Mutations isolated from laboratory strains have been instrumental to the understanding of the GP map. Under the classical scheme, a mutation is compared to a wild-type reference, and its phenotypic effects are used to infer gene function. This framework often leads to a semantic shortcut: from a genetic change causing a *variation* in phenotype, it is often convenient to assimilate the corresponding gene as a causal determinant of a trait (Keller, 2010; **Figure 1A**). It is common to find headlines expressing these simplifications, trumpeting to wide audiences the discovery of the “longevity” or “well-being” gene, that sacrifice scientific accuracy to psychological impact. Along these lines, should a gene whose mutation is lethal be called a “life gene”? What these over-simplified formulations truly mean is that *variation* at a given gene causes *variation* in a given phenotype (Dawkins, 1982; Schwartz, 2000; Waters, 2007). In fact, a gene alone can neither cause an observable phenotypic trait, nor can it be necessary and sufficient to the emergence of observable characteristics. Genes need a cellular environment, the combined action of multiple other genes, as well as certain physico-chemical conditions to have an observable effect on organisms (**Figure 1B**). For example, brown hair pigmentation in one human being is not just a product of the genes coding for pigment synthesizing enzymes but also of the presence of cells producing pigments of relevant substrate molecules (such as tyrosine for melanin), and of the amount of received sun light (Liu et al., 2013). Thus, the genetic reductionist approach, which only explores a few genetic parameters among the variety of causal factors, is vain to fully address the broad question of what makes hair brown, of what brings forth a particular biological structure, or process in its entirety. Nevertheless, genetic reductionism can be perfectly appropriate for identifying genetic loci where a change causes a phenotypic difference (**Figure 1C**). A *difference* in hair color between two individuals could be due in some cases to their genetic difference. We note, however, that not all phenotypic changes can be attributed to genetic changes. A difference in hair color could also be caused by non-genetic factors such as age, intensity of solar radiation or hair dyeing, or by a combination of both genetic and non-genetic differences.

While modern genetics was in its infancy, Alfred Sturtevant formulated the question of the GP map in simple terms: “one of the central problems of biology is that of differentiation – how does an egg develop into a complex many-celled organism?

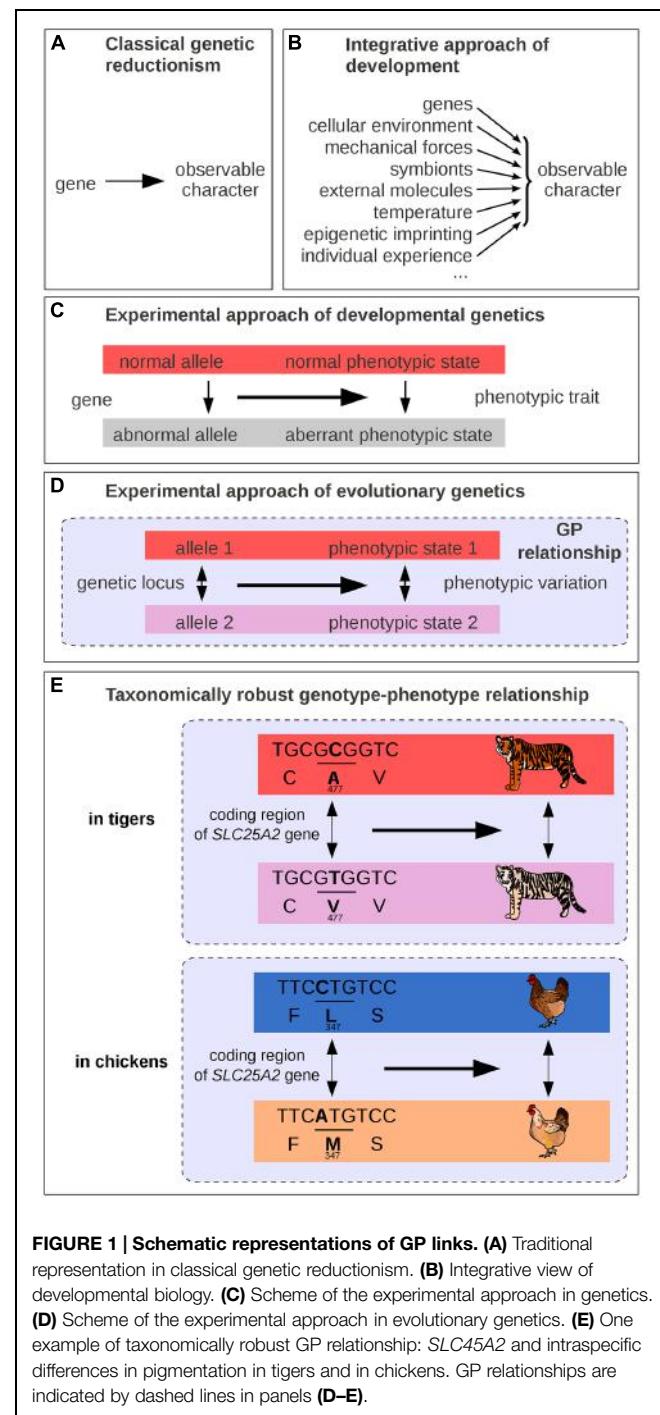


FIGURE 1 | Schematic representations of GP links. **(A)** Traditional representation in classical genetic reductionism. **(B)** Integrative view of developmental biology. **(C)** Scheme of the experimental approach in genetics. **(D)** Scheme of the experimental approach in evolutionary genetics. **(E)** One example of taxonomically robust GP relationship: *SLC45A2* and intraspecific differences in pigmentation in tigers and in chickens. GP relationships are indicated by dashed lines in panels **(D–E)**.

That is, of course, the traditional major problem of embryology; but it also appears in genetics in the form of the question, How do genes produce their effects?” (Sturtevant, 1932). For long some geneticists may have thought that they were dissecting the morphogenetic mechanisms underlying the *formation* of phenotypic traits, while their experimental approach were in fact uncovering genes whose absence or alteration (mutations, deletions, duplications, rearrangements, etc.) leads to phenotypic differences (compare **Figure 1A** with **Figure 1C**). In fact, the

sentence “your hair is brown” can be interpreted either as an absolute observation (a description of a particular assemblage of molecules containing defined levels of the dark pigment eumelanin and of the pale pigment pheomelanin) or with implicit reference to other possibilities (it is brown and not of another color). Misconceptions arise because phenotypes are usually defined relative to possibilities that are not formulated explicitly. Our minds and our language often tend to confuse the objects whose variation is under consideration with the variation itself (Keller, 2010), and it is essential to remind that, in genetics, the objects of interest (e.g., a given genotype, an allele or a phenotype) deserve to be defined *relatively* to another reference state.

In summary, the classical genetic reductionist approach is inherently unable to elucidate all the factors responsible for observable characteristics in the living world (Stotz, 2012) but is a powerful and relevant method for dissecting the genetic levers of heritable phenotypic variation. Focusing on phenotypic variation between individuals rather than on absolute characters present in single organisms is key to better comprehend the genetic causes of phenotypic diversity.

The GP Relationship is between Two Levels of Variation

Thinking in terms of differences makes apparent an abstract entity that encapsulates both genetic and phenotypic levels. This entity is composed of a variation at a genetic *locus* (two alleles), its associated phenotypic change (two distinct phenotypic states), and their relationships (**Figure 1D**). The three of us name the assemblage of these elements a “gephe,” but here we simply call it a “genotype–phenotype relationship” (GP relationship). We will show that the GP relationship is much more than a simple and loosely defined interaction between two levels of organization: it is a cause-and-effect connection that facilitates our understanding of phenotypic diversity.

The Genetic Part of a GP Relationship

In current genome annotation databases, a *gene* is usually defined as a stretch of nucleic acids that is transcribed and codes for an RNA or a polypeptide with a known or presumed function (Gerstein et al., 2007). The genetic locus underlying a phenotypic difference is not necessarily a gene in the strict sense; it can span a particular base-pair, a coding region, a *cis*-regulatory region, or extend to an entire gene with its *cis*-regulatory regions, or even to a gene cluster (**Table 1**). As previously noted by others (Falk, 1984; Gilbert, 2000; Stern, 2000; Moss, 2003; Griffiths and Stotz, 2013), the concept of gene in developmental biology and in current genome annotation databases is distinct from the concept of gene in evolutionary biology. Here the emphasis is not on the gene itself as defined in genome databases, but rather on a case-by-case functional partitioning of the genome into difference-making loci. The genotypic part of a GP relationship can take the form of various alleles: distinct codons coding for different amino acids, insertions/deletions within a protein coding sequence, diverging versions of a particular *cis*-regulatory element, presence/absence of transposon insertions,

number of gene copies within a gene cluster prone to structural variation, etc. Within a genome not all nucleotide sites are associated with phenotypic variation. For instance, there are probably fragments of nucleotide sequences, including the so-called junk DNA (Graur et al., 2015), whose presence does not have any consequence on observable characteristics of the organism, besides being replicated, and possibly transcribed. There are also genetic loci that may have been associated with phenotypic variation in the past and that are no longer associated with phenotypic variation. For example, genetic variation in histone DNA binding coding regions may have been important during the early evolution of eukaryotic cells, but these genetic loci no longer harbor phenotypically relevant variation besides lethal mutations. Within a genome, there are thus nucleotide sites that are absolutely required for life, but that do not harbor viable phenotypically relevant variation themselves.

The Phenotypic Part of a GP Relationship

The phenotypic counterpart of the GP relationship refers to a kind of variation (hair color, level of toxin resistance, etc.) rather than to a state (blond hair, taster of phenylthiocarbamide, etc.; **Table 1**).

The phenotype associated with a genetic change is not necessarily confined to the organism that harbors the genetic mutation. For example, the difference between left- and right-coiled shells in the snail *Lymnaea peregra* is determined by a single genetic locus with maternal effect: the genotype of the mother, but not of the individual itself, is responsible for the direction of shell coiling (Boycott et al., 1931). In other cases, the causal genetic change lies within symbiont bacteria: aphid thermal tolerance can vary between individuals due to a point mutation in their bacterial symbiont (Dunbar et al., 2007). Certain phenotypic effects can also come up at a level higher than the organism harboring the genetic change (Dawkins, 1982), one exemplary case being the social organization of an ant colony (Wang et al., 2013).

The Differential Part of a GP Relationship

As defined above, the GP relationship encompasses a genetic difference and a phenotypic difference. The relationship of *difference* at both the genetic and the phenotypic level is quite abstract, and it can correspond to three distinct differences within the living world: (#1) a difference between two reproductively isolated taxa (living or extinct), (#2) a difference segregating within a population, and (#3) the difference that first appeared during evolution, between an organism harboring the ancestral allele/trait and its direct descendant which evolved the new allele/trait. Of note, the variation in phenotype does not always immediately follow the emergence of the new causing mutation, but can appear later from the singular assortment of alleles that are segregating in the population. For example, a new phenotype of reduced armor plates appeared in a freshwater stickleback population when a recessive *EDA* allele already present at cryptic levels ended up in a homozygous state in one individual (Colosimo et al., 2005; Jones et al., 2012). A major conceptual advance made by Charles Darwin was to

TABLE 1 | A few examples of GP relationships.

Genetic locus	Phenotypic trait	Organisms
Two coding sites in the <i>ABO HBGG</i> gene	A/B blood group	Human, chimpanzee, gibbon
<i>Cis</i> -regulatory element in the <i>lactase</i> gene	Ability to digest milk	Various human populations
Number of duplications of the amylase genes <i>AMY1</i> or <i>AMY2B</i>	Ability to digest starch	Human, dog
Presence/absence of a complex of adjacent genes coding for carotenoid desaturases and carotenoid cyclase/synthases	Ability to produce carotenoids	Pea aphid, spider mite, gall midge fly, nematode
Coding sites in <i>opsin</i> genes	Color vision	Human, cetaceans, fishes, butterflies
Coding region of <i>F3'5'H</i>	Flower pigmentation	Soybean, pea, annual phlox, potato, <i>lochroma</i>
Coding region of <i>FRIGIDA</i>	Flowering time	Thale cress, oilseed rape
Coding region of <i>BADH2</i>	Fragrance	Soybean, rice
Coding region of <i>HMA3</i>	Heavy metal tolerance	Thale cress, rice
Coding and <i>cis</i> -regulatory regions of <i>myostatin</i>	Muscle size	Cattle, sheep, dog, pig, horse, human
<i>Cis</i> -regulatory regions in the <i>achaete-scute</i> complex	Number and position of sensory bristles	Fruitflies
<i>Cis</i> -regulatory element in the <i>pitx1</i> gene	Pelvis morphology	Stickleback fish
Coding region of <i>SLC45A2</i>	Pigmentation of eye, hair, and skin	Human, tiger, chicken
Coding region of <i>Mc1R</i>	Pigmentation of hair and skin, but not eye	Human, mouse, cattle, chicken, guinea pig, horse, fox, pig, sheep, dog, rabbit, bear, jaguar, jaguarundi, squirrel, birds, sand lizard
<i>Cis</i> -regulatory regions of <i>tan</i> and <i>ebony</i>	Pigmentation pattern	Fruitflies
<i>Cis</i> -regulatory elements in the <i>shavenbaby</i> gene	Position and number of trichomes	Fruitflies
<i>Cis</i> -regulatory elements in the <i>optix</i> gene	Red color pattern on butterfly wings	Longwing butterflies
Coding region of <i>hemoglobin alpha</i> and <i>beta</i> chain genes	Resistance to hypoxia	Human, llama, crocodile, deer mouse, waterfowl
Coding region of <i>Ace</i>	Resistance to organophosphate insecticides	Potato beetle, aphids, mosquitoes, house fly, fruit flies, oriental fruitfly
Presence/absence of a <i>CypA</i> insertion within the <i>TRIM5a</i> gene	Resistance to retrovirus	Owl monkey, macaque, Old World monkeys
Coding sites in the <i>Nav1.4</i> gene	Resistance to tetrodotoxin or saxotoxin	Snakes, pufferfish, clam
Coding region of <i>TAS2R38</i>	Sensitivity to bitterness	Human, chimpanzee
<i>Cis</i> -regulatory and coding regions in <i>Agouti</i>	Skin and coat pigmentation	Human, deer mouse, cattle, pig, dog, cat, horse, fox, domesticated fox, quail, sheep
Number of duplications of the glucose transporter gene <i>HXT6</i>	Survival in low-glucose environment	Yeast
Number of duplications of the <i>CCL3L1</i> gene	Susceptibility to HIV infection and progression rate of AIDS after infection	Human
<i>Cis</i> -regulatory element in the <i>WntA</i> gene	Wing pigmentation pattern	Butterflies

See (Martin and Orgogozo, 2013a) for references and for additional cases.

relate variation among individuals within an interbreeding group (difference #2) with variation between taxonomic groups in space and time (difference #1; Lewontin, 1974a).

Note also that certain phenotypic changes may appear at the level of the entire organism when the “causative” mutation is accompanied by additional somatic mutations that are highly likely. For example, in women carrying a wild-type allele and a mutant allele of *BRCA1*, cells can produce wild-type *BRCA1* proteins since they carry one copy of the wild-type *BRCA1* allele. Nevertheless, these women have up to an 80% risk of developing breast or ovarian cancer by age 70 compared to women carrying two wild-type copies of *BRCA1*, due to the appearance of additional deleterious mutations within the wild-type *BRCA1* allele in their somatic breast cells (Narod and Foulkes, 2004).

Importantly, the GP difference is always defined relative to a population, or taxon, of interest (Sober, 1988). In less medically developed countries, humans carrying two defective

copies of the phenylalanine hydroxylase gene have serious medical problems including seizure and intellectual disabilities. In contrast, in most developed countries, such humans are diagnosed at birth and have a normal life span with normal mental development thanks to a phenylalanine-restricted diet (Armstrong and Tyler, 1955). Therefore the GP relationship involving the phenylalanine hydroxylase defective mutation is context-dependent: the mutation is associated with health problems in less medically developed countries but not in other countries. This example shows that the causal relationship between a genetic change and its associated phenotypic change can hide multiple embedded parameters (such as medical practices for the phenylalanine hydroxylase case) within the *ceteris paribus* assumption of “all other things being equal.”

In summary, the GP relationship is best viewed as a relationship between two variations, one at the genotypic level, and one at the phenotypic level. The human mind can elaborate concepts of increasing abstraction: concepts of things (e.g., a cell),

concepts of change (e.g., evolution), and concepts of relations (e.g., homology; Cassirer, 1910; Simondon, 1968). Here the concept of GP relationship establishes a relation between two *changes* (genetic and phenotypic). In the next paragraphs we will show that, compared to the usage of intuitive concepts of things, this detour through increased abstraction may prove more efficient to better understand phenotypic diversity.

Several Current Representations of the Connection between Genotype and Phenotype Implicitly Dismiss the Differential View

We argued above that the differential view should always be kept in mind when thinking about the connection between genotypes and phenotypes. GWAS, which represent the most popular method to detect genomic loci that are associated with complex traits in populations, are based on the analysis of differences (Visscher et al., 2012). Nevertheless, in current research the differential view is sometimes implicitly dismissed. When multiple factors are observed to influence phenotypic traits (Figure 1B), the differential view is considered as too simplistic and researchers often prefer to focus back on phenotypes of single individuals, without explicitly relating them to a phenotypic reference.

In most current articles, the problem of connecting the genotype to the phenotype is framed in terms of genotype and phenotype maps. The first GP map was introduced by Richard Lewontin in his book “The genetic basis of evolutionary change” (Lewontin, 1974a; Figure 2A). He indicated the average genotype of a population as a point in the space of all possible genotypes (G space) and the average phenotype of the same population as a corresponding point in the space of all possible phenotypes (P space). The evolutionary process was thus decomposed into four steps: (1) the average phenotype is derived from the development of the distinct genotypes in various environments; (2) migration, mating, and natural selection acts in P space to change the average phenotype of the initial population into the average phenotype of the individuals

which will have progeny; (3) the identity of successful parents determines which genotypes are preserved; and (4) genetic processes such as mutation and recombination modify position in G space.

In another common graphical representation (Figure 2B), a point in the G space and its corresponding point in the P space correspond to the genotype and the phenotype of a single individual (Fontana, 2002; Landry and Rifkin, 2012). Under such a representation, the abstract object that we defined above as the GP relationship would correspond to a “move” in genotype space associated with a “move” in phenotype space (or, better, a sum of several “moves” in genotype, and phenotype spaces because several distinct genomes can carry the two alternative alleles of a given GP relationship). In a third representation put forward by Wagner (1996; Figure 2C), individual genes are connected to individual traits.

Although these three graphical representations of GP maps may facilitate our understanding of certain aspects of biology, in all of them the GP relationship and the differential view are not easy to grasp. It is quite perplexing that the first person to draw such a GP map was Richard Lewontin, an eloquent advocate of the differential view (see for example his preface to Oyama, 2000, a masterpiece of persuasion). Because these graphics focus on individual rather than differential objects, we believe that these three representations implicitly incite us to go back to the more intuitive idea of one genotype associated with one phenotype. Losing sight of the differential view might also come from the molecular biology perspective, where proteins are viewed as having causal effects on their own, such as phosphorylation of a substrate or binding to a DNA sequence. Because of the two entangled definitions of the gene, either as encoding a protein, or as causing a phenotypic change (Griffiths and Stotz, 2013), it is easy to move from a differential view to a non-differential view of the GP relationship.

In summary, many current mental representations of the connection between genotype and phenotype implicitly dismiss the differential view. We will now show that the differential view is compatible with the fact that phenotypic traits are influenced by a complex combination of multiple factors and that we can find a relevant schematic representation of GP relationships.

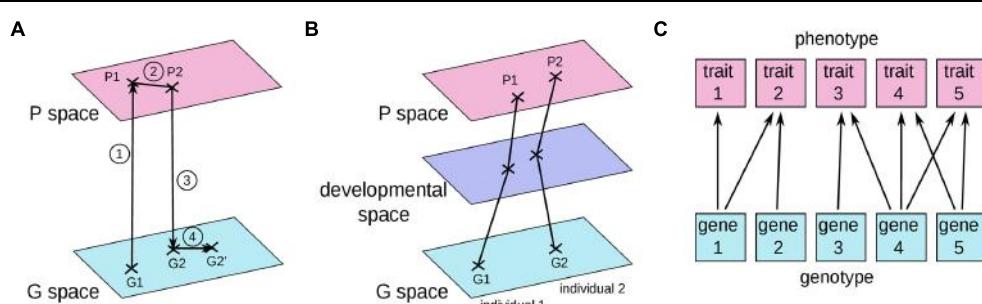


FIGURE 2 | Three current graphical representations of GP maps. **(A)** The early version of the GP map proposed by Lewontin (1974a). **(B)** A GP map where each point represents a single individual (Houle et al., 2010; Gjuvsland et al., 2013; Salazar-Ciudad and Marín-Riera, 2013). **(C)** The relationships between traits and genes, as depicted by Wagner (1996). See text for details.

The Problem of Pleiotropy

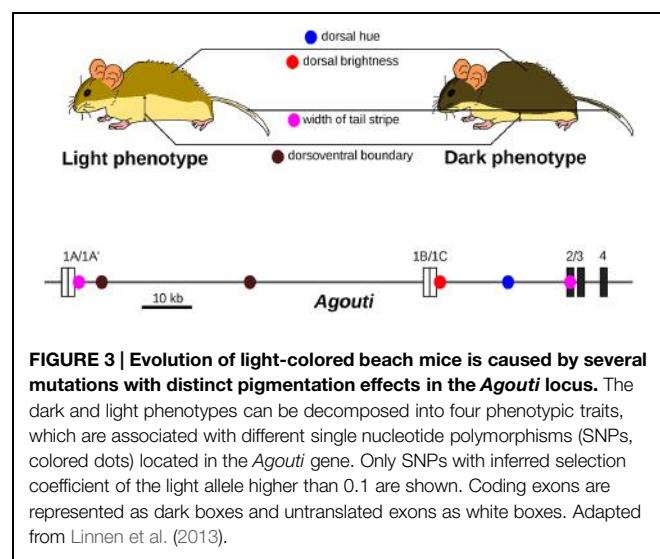
Decomposing an organism into elementary units such as anatomical structures has been instrumental in many biology disciplines such as physiology, paleontology and evolution. However, the issue is to identify the decomposition into characters that is most adequate for the question of interest. For questions related to relationships between organs of various individuals or species (such as homology), it might be appropriate to keep the traditional decomposition into anatomical structures (Wagner, 2014). Richard Lewontin and Günter Wagner defined characters as elements within an organism that answer to adaptive challenges and that represent quasi-independent units of evolutionary change (Lewontin, 1978; Wagner, 2000). Their definition deals with absolute traits observed in single organisms (for example the shape of a wing, or the number of digits in an individual) and is thus far from the differential view. Here, to better apprehend evolution and phenotypic diversity of the living world, we propose to decompose the observable attributes of an organism into multiple elementary GP variations that have accumulated through multiple generations, starting from an initial state. We insist that under this perspective, characters are not concrete objects (such as skin) but abstract entities defined by the existence of *differences* between two possible observable states (for example skin color). As an analogy, one can imagine two ways to produce a well-worn leather shoe of a particular shape. One can either assemble the different atoms into the same organization, or one can buy a shoe in a store and then subject it to a series of mechanical forces. We are naturally inclined to compare organisms to machines, and to think in terms of pieces that must be assembled to make a functional whole. However, the rampant metaphor of the designer or maker is inadequate for understanding the origin of present-day organisms (Coen, 2012). To understand the phenotypic features of a given organism it is more efficient to decompose it into abstract changes that occurred successively across evolutionary time, and not across developmental time. The initial state is a hypothetical ancestor of the organism under study.

Certain mutations (qualified as pleiotropic) are observed to affect several organs at once while others alter only one at a time (Paaby and Rockman, 2013; Zhang and Wagner, 2013). For pleiotropic mutations, we consider that the GP relationship should include all the phenotypic changes (in diverse organs, at various stages, etc.) associated with the genetic difference. For instance, the V370A mutation of the EDAR receptor is associated not only to hair thickness but also to changes in sweat gland and mammary gland density in Asian populations (Kamberov et al., 2013). The GP relationship is, in such cases, one-to-multi. Considering skin and eye as independent anatomical modules of the human body might seem appropriate for many evolutionary changes, but it is somewhat inadequate in cases where these two organs evolved a new pigmentation trait at once through a single mutation in the *SLC45A2* gene (Liu et al., 2013). Reasoning in terms of GP relationships strikes off the problem of finding a relevant decomposition into elementary anatomical structures. The elementary GP relationships themselves appear as adequate semi-independent

modules, whose combination can account for the observable characteristics of an organism.

The Problem of Continuous Complex Traits

Under the differential concept of GP relationships, one crucial point is to decompose observable traits into a series of semi-independent phenotypic variations, that is to identify the elementary changes that have occurred during evolution. Experimental approaches are available to decompose a given phenotypic difference into appropriate finer sub-variations. For example, crossing plants with different leaf shapes yields a progeny that exhibits a composite range of intermediate leaf shapes. Principal component analysis uncovered elementary leaf shape changes that can together account for the difference in shape between parental lines and that appear to be caused by distinct genomic regions (Langlade et al., 2005). This suggests to some extent that “the sum obscures the parts.” What we traditionally consider as complex traits can be made of simpler traits, more amenable to genetic analysis. Another illuminating example is the abdominal pigmentation in the *Drosophila dunni* group. Taken as a single variable, the levels of pigmentation show a complex genetic architecture, but decomposing adult patterns into anatomical sub-units unravels discrete genetic control for each sub-trait (Hollocher et al., 2000). A better known case is the evolution of body color in beach mice. The difference in color between light-colored beach mice and dark mice can be decomposed into distinct phenotypes (dorsal hue, dorsal brightness, width of tail stripe, and dorsoventral boundary), which are all associated with distinct mutations in the *Agouti* gene (Linnen et al., 2013; **Figure 3**). Each *Agouti* genetic locus appears to be dedicated to the specification of pigmentation in a given body part. Together, they form a group of tightly linked loci that are associated with changes in coat pigmentation.



While complex traits may not always be reducible to a suite of simple GP relationships, it is possible that traits such as adult human height, the most emblematic quantitative trait predicted to consist of many genetic effects of small size (Fisher, 1930), might also be decomposed into elementary variations, each explaining more discrete sub-trait. While some determinants of human height such as *LIN28B* have been associated to adult height at different ages, other genes have only reached statistical significance in stage-specific studies focusing on fetal growth and height velocity at puberty (Lettre, 2011). In other words, these data suggest that human height may be a composite trait that is modulated by several GP relationships, each acting at different phases of developmental growth.

The Problem of Epistasis and GxE

Gene-by-Environment (GxE) interaction occurs when the phenotypic effect of a given genetic change depends on environmental parameters. Similarly, epistasis, or GxG interaction, occurs when the phenotypic effect of a given genetic change depends on the allelic state of at least one other locus (Phillips, 2008; Hansen, 2013). There is increasing evidence that GxG and GxE interactions are of fundamental importance to understand evolution and inheritance of complex traits (Gilbert and Epel, 2009; Hansen, 2013). We propose that both phenomena can be integrated into the basic GP differential framework, where both GxG and GxE interactions inject a layer of context-dependence, and result in differences embedded within differences.

The difference in color pigmentation between dark and light-colored beach mice mentioned previously (Figure 3) is not only due to mutations in *Agouti* but also to a coding mutation in the *MC1R* gene that decreases pigmentation (Steiner et al., 2007; Figure 4B). The effect of the *MC1R* mutation is visible only in presence of the light-colored-associated derived *Agouti* haplotype. Here the *Mc1R* locus is considered to interact epistatically with the *Agouti* locus. In this case, we propose that the GP relationship does not comprise a single phenotypic difference but two possible phenotypic differences (a change in coat pigmentation or no change at all). The choice between these two phenotypic differences is determined by the genetic background (here at the *Agouti* locus). The differential view thus remains relatively straightforward for two-loci interactions: the context-dependence of the phenotype is translated into a choice between two possible phenotypic differences. We propose that a GP relationship involving a mutation subjected to multiple epistatic interactions should comprise all possible phenotypic differences that can result from the mutation in all genetic backgrounds. Among all possible phenotypic variations, the phenotypic difference that will be observed is determined by other genetic loci. In general, GxG interactions involve multiple sites that are dispersed across the genome (Bloom et al., 2013).

An example of GxE interaction (see also Figure 4A) is the naturally occurring loss-of-function allele of *brx* in *Arabidopsis*

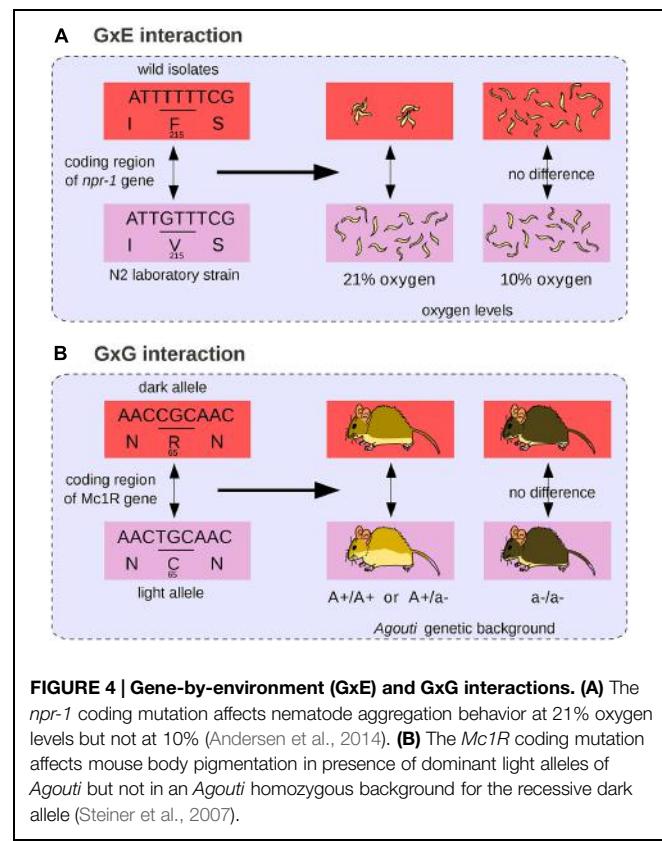


FIGURE 4 | Gene-by-environment (GxE) and GxG interactions. (A) The *npr-1* coding mutation affects nematode aggregation behavior at 21% oxygen levels but not at 10% (Andersen et al., 2014). **(B)** The *Mc1R* coding mutation affects mouse body pigmentation in presence of dominant light alleles of *Agouti* but not in an *Agouti* homozygous background for the recessive dark allele (Steiner et al., 2007).

plants, which is associated with accelerated growth and increased fitness in acidic soils, and with severely reduced root growth compared to wild-type in normal soils (Gujas et al., 2012). GxE interactions are usually analyzed in the form of a *norm of reaction*, which represents all the observable traits of a *single genotype* across a range of environments (Johannsen, 1911; Sarkar, 1999). In the case of GxE interactions, we propose that the GP relationship should comprise all the possible phenotypic changes that can be caused by the associated genetic change across various experimental conditions. The associated phenotypic change is thus a difference between two norms of reaction. A textbook example is the variation in temperature-size rule in *C. elegans*. Like most other animals, *C. elegans* nematodes grow larger at low temperature, but a wild-type laboratory strain of *C. elegans* originating from Hawaii shows no variation in body size across various temperatures. An amino acid change in a calcium-binding protein is responsible for the decreased ability of the Hawaiian strain to grow larger at low temperature (Kammenga et al., 2007). Here the norm of reaction (representing nematode body size across a range of temperatures) differs between nematodes and the associated GP relationship encompasses the difference between these two slopes.

The range of phenotypic variations embodied within GP relationships subjected to GxG and GxE interactions can be quite overwhelming, especially in cases when several tissues are affected by the same mutation, and when the phenotypic variation of each tissue is influenced by other genomic loci and by environmental conditions. In fact, the phenotypic effects of a mutation always

rely on other pieces of DNA from the same genome, so that any GP relationship can be considered to experience epistasis. In other words, a genetic locus affecting a phenotype never acts independently of other DNA sequences. For instance, a given opsin allele will only lead to particular color vision properties if an eye is formed and if this eye receives light during its development, allowing effective vision neural circuits to form. For the differential view to be tractable, we advise not to consider all possible genetic backgrounds and environmental conditions, but to restrict possibilities to potential environments, and segregating alleles that are relevant to the population of interest (Sober, 1988).

In summary, in presence of epistasis or GxE interactions, a genetic change is not associated with a single phenotypic difference but with multiple possible phenotypic differences, among which one will be achieved, depending on the environment and the genetic background. The context-dependence can be represented schematically as GP differences embedded into other genotype and environment differences.

The Differential View of Genetic and Environmental Effects on Phenotypes

As underlined by multiple authors (most notably Waddington, 1957; Oyama, 2000; Keller, 2010), genes and environment act jointly on the phenotype, and in most cases it is impossible to disentangle the effect of one from the other. Here we show that reasoning in terms of differences helps to clarify the comparison between genetic and environmental effects on phenotypes. However, we identify certain cases where the comparison remains difficult.

By analogy with the GP relationship, we can define the environment-phenotype relationship as an environmental variation (two environments), its associated phenotypic change (distinct phenotypic states), and their relationships. For example, in many turtle species, a change in temperature during egg development is associated with the male/female sex difference (**Figure 5A**) and at least six transitions from environmental to genetic sex determination (**Figure 5B**) occurred across the turtle phylogeny (Pokorná and Kratochvíl, 2009). In this case, environmental and genetic effects can be compared: sex chromosomes and temperature have the same phenotypic effect on turtles. Such observations led West-Eberhard (2003, 2005) to propose the “genes as followers” hypothesis, which suggests that novel phenotypic states are more likely to arise first from a change in the environment than from a genetic mutation, and that mutations occur only later, in modifying the threshold for expression of the novel trait. West-Eberhard (2003, 2005) extrapolated from differences segregating within populations (difference #2) to differences that arose temporally during the evolution of a population (difference #1).

The independent evolution of directional left-right asymmetry from symmetrical ancestors in multiple lineages has provided a major argument supporting the “gene as follower” hypothesis (Palmer, 2004). Under this framework, it is stipulated that directional asymmetry, where all individuals are same-sided, has often evolved from a “random asymmetry” state, where the directionality will depend on environmental factors and thus vary between genetically identical individuals. For instance,

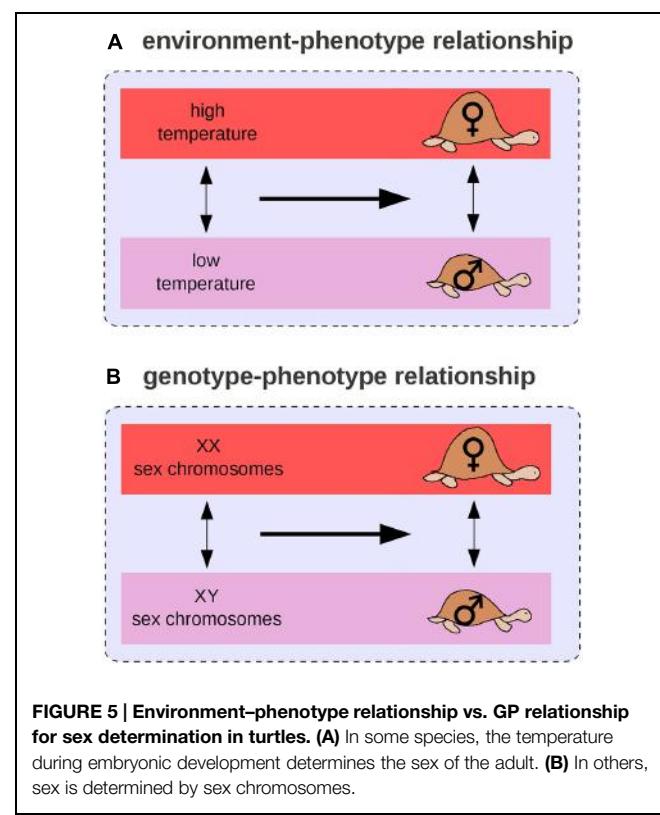


FIGURE 5 | Environment–phenotype relationship vs. GP relationship for sex determination in turtles. **(A)** In some species, the temperature during embryonic development determines the sex of the adult. **(B)** In others, sex is determined by sex chromosomes.

the strongest claw of a lobster will develop based on usage and has a priori equal probabilities to develop on the left or on the right side. We can see how the “genes as followers” formula applies here: the environment triggers an asymmetry, and later in evolution some genetic effects can bias its directionality on one side or the other. But while asymmetry “occurs before genetic variation exists to control it,” the differential view makes it clear that the genetic effect on directionality is not comparable to the environmental effect that triggers the asymmetry. The genetic change makes a switch between the final 100% same-sided condition and an initial condition where 50% of the cases are dextral and 50% sinistral. In contrast, the two alternative phenotypic states resulting from variation in the environment are considered to be 100% dextral and 100% sinistral. This example shows that for the sake of accuracy it is important to explicitly state the differences that are being considered within a GP relationship.

The differential view provides a theoretical framework that can help in designing experiments to investigate the proper variables: one can compare different genotypes in a fixed environment (classic GP relationship), compare the response of a fixed genotype to two different environments (phenotypic plasticity), or compare the sensitivity of two different genotypes to two different environments (wherein the phenotypic variation becomes a *difference in a difference*; see for example Engelmann et al., 2009; Thomas, 2010).

Various quantitative methods have been developed to disentangle genetic from environmental effects and to quantify GxE interactions (Lynch and Walsh, 1998). Yet in certain

situations it can be impossible to separate genetic from environmental effects in a biologically meaningful way, even when reasoning in terms of differences (Lewontin, 1974b). Populations of the beetle *Calathus melanocephalus* comprise two morphs, long-winged, and short-winged (Schwander and Leimar, 2011). The long-winged morph only develops from homozygous individuals for a recessive allele segregating in the population, and only when food conditions are good. In this case, the genetic and environmental effects are intermingled (Figures 6A,B). In the theoretical case of a population comprising only short-winged heterozygous animals that have been raised in starving conditions and long-winged ones, both genes and environment are responsible for the wing difference between individuals and it is impossible to estimate the proportion of environmental and genetic effects because genes and environment act on distinct levels along the complex causal link between genotypes and phenotypes.

Another case that questions the classical environment/genetic distinction is when the addition of certain symbiotic bacteria modifies the host phenotype. Mice fed with a *Lactobacillus* strain of bacteria show reduced anxiety-related behaviors compared to control mice fed with broth without bacteria (Bravo et al., 2011). Here the behavioral difference is caused by a switch between presence or absence of a particular gut symbiont. The cause of the phenotypic difference is not a simple change in a DNA sequence,

nor a simple environmental change disconnected from genetic changes, but a switch between presence and absence of a factor that can be considered as an environmental factor – the bacteria – which contains DNA whose mutations may also change the host phenotype.

In conclusion, reasoning in terms of differences can help to clarify the comparison between genetic and environmental effects on phenotypes. However, the issues are nothing but simple. Since genes and environment act on distinct levels along the complex causal link between genotypes and phenotypes, in certain cases it is impossible to disentangle both causes.

A Clarification on the Terminology Gain/Loss and Permissive/Instructive

Phenotypic differences appear to fall under two major categories, either the presence/absence of something (for example body hair or the ability to digest milk), or the shift between two alternatives that are both present (for example two hair colors). Similarly, on the genotype side, a mutation can correspond to the presence/absence of a relevant DNA sequence, or to a nucleotide polymorphism. The differential perspective makes it evident that a loss of phenotype is not necessarily associated with a loss of genetic material, and vice versa. For example, the evolutionary gain of dark pigments covering the entire coat of animals has often been associated with a loss of the *MclR* gene (Gompel and Prud'homme, 2009). Furthermore, as one of us noted previously (appendix of Stern and Orgogozo, 2008), gain or loss for a phenotype is subjective. For example, loss of hair might also be considered as gain of naked epidermis. Most insect epidermal cells differentiate into one of these two alternative states and both states involve large gene regulatory networks. It is not clear which phenotypic state represents a gain or loss relative to the other. Even on the genotypic side, defining losses and gains can be difficult. The insertion of a transposable element can knock down a gene, whereas a deletion can sometimes creates a new binding site for an activator of transcription. As a matter of fact, the evolutionary gain of *desatF* expression in *D. melanogaster* occurred through a series of three deletions, each creating an hexamer motif that is required for *desatF* expression (Shirangi et al., 2009).

Similarly, the differential perspective on environmental effects highlights the fallacy of the distinction between permissive and instructive signals. A permissive signal is associated with the presence/absence of a phenotype and an instructive signal with the shift between two alternatives that are both present. As argued above, these distinctions at the phenotypic level are not clear-cut.

In conclusion, we suggest that the gain/loss and instructive/permissive terminology should be used with caution.

Taxonomically Robust GP Relationships

A mutation is expected to produce a somewhat reproducible phenotypic variation within a population. Such reproducibility

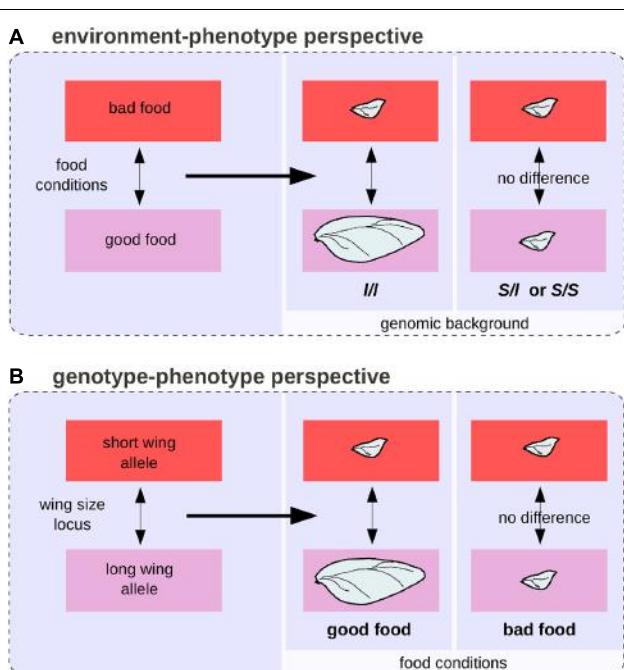


FIGURE 6 | The environment-phenotype relationship and GP relationship perspectives for wing length polymorphism in the beetle *Calathus melanocephalus*. **(A)** Under the environment-phenotype relationship perspective, a change in food conditions is associated with a change in wing size, but only in a homozygous background for the recessive allele (*I*) of the wing size locus. **(B)** Under the GP relationship perspective, a genetic change at the wing size locus is associated with a change in wing size, but only in good food conditions.

in phenotypic outcome is required to allow genetic evolution and adaptation by natural selection (Lewontin, 1974a; Kirschner and Gerhart, 1998). Indeed, a newly formed allele that would generate yet another phenotype each time it ends up in a different organism would not be subjected to natural selection. Reasoning in terms of variation, rather than considering alleles as isolated entities, makes it clear that competition occurs between alleles that span the same genetic locus. Natural selection acts directly on the allelic variation that is consistently associated with a given phenotypic variation, which is the GP relationship itself. The GP relationship is thus a basic unit of evolutionary change, on which natural selection acts.

A major discovery of the past 20 years is that variation at certain genetic loci produce comparable phenotypic variation not only in various individuals of one population, but also in extremely diverse taxa (Martin and Orgogozo, 2013b). In other words, certain GP relationships are taxonomically robust and present across a large range of species. This implies that the genetic and environmental backgrounds have remained relatively constant or have appeared repeatedly throughout evolution to allow for genetic loci to generate similar phenotypic changes in various taxonomic groups. This important finding was quite unsuspected some 50 years ago. For a long time the singularity observed in the living world was expected to reflect a comparable singularity at the genetic level, implicating disparate and non-conserved genes, specific to each lineage (Mayr, 1963). As Mayr (1963) once proposed in 1963, “Much that has been learned about gene physiology makes it evident that the search for homologous genes is quite futile except in very close relatives [...]. The saying “Many roads lead to Rome” is as true in evolution as in daily affairs” (Mayr, 1963). In other words, the genetic loci that make a man a man were expected to be different from the ones that make a dog, or a fish. Later, in the 80–90s, a few researchers suggested quite the contrary that evolution proceeds through mutations in conserved protein-coding genes (Romero-Herrera et al., 1978; Perutz, 1983; Stewart et al., 1987; Carroll et al., 2005) – but they had little experimental data at hand to support their view (Tautz and Schmid, 1998). As of today, the accumulating data on the mutations responsible for natural variation make it clear that the diversity in living organisms share a common genetic basis on at least three points. First, comparative developmental biology revealed that animals share common sets of key regulatory genes with conserved functions (Wilkins, 2002, 2014; Carroll et al., 2005). Second, most interspecific differences in animals and plants for which the underlying genetic basis has been at least partly identified (154 cases out of 160) are due to mutations at homologous genes, and very few (6/160) are due to new genes, which nevertheless represent duplicates of existing genes (Martin and Orgogozo, 2013b). Third, multiple cases of similar phenotypic changes have been shown to involve mutations of the same homologous genes in independent lineages (**Table 1**), sometimes across large phylogenetic distances. For instance, the difference in pigmentation between white and orange Bengal tigers has recently been mapped to a single mutation in the transporter protein gene *SLC45A2* (Xu et al., 2013), and this gene has also been associated with hypopigmented eyes, skin, hair, and feathers in humans and chickens (Xu et al., 2013;

Figure 1E). A more dramatic example is the recent evolution of a toxin resistance in three species that diverged more than 500 million years ago – a clam, a snake and a pufferfish – via the same amino acid substitution in a conserved gene (Bricelj et al., 2005; Geffeney et al., 2005; Venkatesh et al., 2005; Feldman et al., 2012). Such striking patterns of genetic repetition have now been found for more than 100 genes in animals and plants (Martin and Orgogozo, 2013b). Despite existing methodological biases favoring conserved genes in the search for quantitative trait loci (Rockman, 2012; Martin and Orgogozo, 2013b), the level of genetic repetition remains astounding and suggests that for the evolution of at least certain phenotypic differences, relatively few genetic roads lead to Rome (Stern, 2013). Nowadays, one should not be surprised that a piece of DNA associated with a complex wing color pattern in one *Heliconius* butterfly species provides similar wings and collective protection from the same predators when introduced into the genome of other butterflies (Supple et al., 2014). What makes a dog a dog or a man a man is now partly explained by singular assortments of taxonomically robust GP relationships, which are found in multiple lineage branches.

Certain environment–phenotype relationships are also taxonomically robust. For example, across most taxa body size is affected by nutrition; iron deficiency can cause anemia and certain toxic compounds can be lethal. In ectotherms the temperature of the organism depends on the environmental temperature. Given the daunting number of environmental conditions that can be conceived, it is probably impossible to determine whether taxonomically robust GP relationships or taxonomically robust environment–phenotype relationships are more prevalent. Furthermore, whether taxonomically robust GP relationships represent an exceptional and small fraction, or a significant proportion, of all GP relationships is a matter of debate. In any case, the existence of taxonomically robust GP relationships is now clear and should be broadly accepted by the research community.

Some of the most striking teachings of modern biology include the discovery that living beings share the same genetic material (DNA or RNA), the same genetic code (with few exceptions), and the same basic cellular machinery. It is thus far from paradoxical that individual differences are built upon similarities, and the finding that certain GP relationships persists over long evolutionary times completes the picture.

The precise predictive power resulting from the existence of taxonomically robust GP relationships is rare in biology, and is only starting to be exploited at its full potential. Long-range conservations of GP links now fully justify the use of comparative genetics approaches to tackle pragmatic problems. For instance, crop domestication took the form of similar selective pressures in many species, and we now have experimental evidence that this process has repeatedly involved mutations in the same set of conserved genes (Paterson et al., 1995; Martin and Orgogozo, 2013b). This observation opens up interesting applications, as we can use this emerging body of genetic expertise to assist the domestication of future crops, or to use marker-assisted strategies to produce and

maintain crop biodiversity (Lenser and Theissen, 2013). GP predictability is already used in the identification of strains that evolved resistance to different pest control strategies, with extreme cases targeting anti-malarial drugs tolerance in *Plasmodium* parasites (Manske et al., 2012), antibiotic resistance in bacteria and yeasts (Fischbach, 2009; MacCallum et al., 2010), or even more dramatically, the anthropogenic evolution of insecticide-resistance in diverse cohorts of insects, regardless of their pest status (Ffrench-Constant et al., 2004; Martin and Orgogozo, 2013b).

Furthermore, repeatability in the genetic basis of phenotypic variation suggests that clinical research is also likely to benefit from genetic studies of a large range of model species (Robinson and Webber, 2014). For instance, natural variation in the tolerance to methotrexate, a chemotherapy drug, was mapped in *Drosophila* fruitflies to genes whose human orthologs are also associated with the response of patients to this drug (Kislukhin et al., 2013), thus extending the use of model organisms as disease models.

Toward a Gene-Based Classification of Phenotypes

One original aspect of framing the GP connection in terms of individual GP relationships is that it allows to classify phenotypes according to their underlying genetic basis. On a first level, GP relationships implicating different regions within the same gene and producing comparable phenotypic outcomes can be grouped together. Simple cases of GxG interactions have been found between tightly linked mutations, generally within a coding sequence or within a *cis*-regulatory element, when they generate a non-additive effect on the phenotype. For example, a particular mutation in an enhancer was observed to produce different shifts in expression pattern of the downstream coding gene, depending on neighboring DNA sequence (Frankel et al., 2011; Rogers et al., 2013). Similarly, amino acid mutations in a hemoglobin gene was found to increase or decrease affinity to oxygen, depending on the allelic state of other sites (Natarajan et al., 2013). In such cases, it is intuitive to group such genetically linked sites together as they all affect the same kind of phenotypic trait.

Absence of melanin pigments in animals has been associated with mutations in several genes, including *OCA2*, *kit ligand* or *Mc1R* (reviewed in Gompel and Prud'homme, 2009; Liu et al., 2013). Whereas the absence of melanin is traditionally considered as one character state, albinism, irrespective of the underlying genetic basis, we propose here to distinguish *OCA2*-associated albinism from *Mc1R*-associated albinism, or from albinism associated with any other gene. One interest of decomposing the variation within the living world into these multiple elementary GP relationships is that these elements can then be grouped together into successively larger groups. Elementary phenotypic changes involving different genes that are part of the same genetic pathway could also be grouped together as concomitant components of the same phenotype-modulating mechanism. This is clearly the case for the TGF- β signaling molecules BMP15, GDF9, and the TGF- β receptor BMPR1B,

that have all been repeatedly associated to variations in ovarian function in humans and in domestic sheep breeds (reviewed in Luong et al., 2011).

Another important consequence of the GP relationship perspective is that apparently distinct phenotypic changes caused by similar genetic loci in various organisms can be examined further to uncover what might be a common basic phenotypic change (Deans et al., 2015). For example, fly larvae and nematode worms have distinctive food search behaviors but mutations in the same orthologous gene (*for/egl-4*) have been shown to alter the intensity of food search behaviors in both organisms (Osborne et al., 1997; Mery et al., 2007; Hong et al., 2008). It is thus plausible that a basic behavioral change, which underlies seemingly distinctive fly and nematode food search changes, represents a conserved GP relationship across nematodes and flies. This somewhat borderline example illustrates the challenge to incorporate widespread comparative thinking into our global understanding of biology. Is a mutation in a mouse model relevant to human disease? Can we consider that a mouse phenotype is similar to a human condition if its genetic basis is different? We and others predict that the search for orthologous phenotypes, or “phenologs” (McGary et al., 2010), will represent a major task for modern genetics and will require a fruitful alliance between applied and evolutionary biology.

Conclusion

In this paper, we bring back the differential concept of gene (Schwartz, 2000) into our framework for understanding the GP map. The differential view of the GP relationship helps to clarify the genetic and environmental effects on phenotypes and their connection. It also opens up new avenues of thinking, in particular regarding the decomposition of observable features within an organism and the representation of GP maps. Furthermore, the existence of taxonomically robust GP relationships encourages an unabashed use of comparative genetics to predict the genetic basis of phenotypic variation in diverse groups of organisms, and this predictive power has an important potential for translational research in agronomy and clinical research.

Acknowledgments

We deeply acknowledge Marie-Anne Félix for enlightening discussions and Thomas Pradeu for bringing several relevant papers to our attention. We also thank Giuseppe Baldacci, Marie-Anne Félix, Pierre-Henri Gouyon, Alexandre Peluffo, Mark Siegal, David Stern, and the reviewers for their insightful comments on the manuscript. The research leading to this paper has received funding from the European Research Council under the European Community's Seventh Framework Program (FP7/2007-2013 Grant Agreement no. 337579) and from the John Templeton Foundation (Grant no. 43903).

References

- Andersen, E. C., Bloom, J. S., Gerke, J. P., and Kruglyak, L. (2014). A variant in the neuropeptide receptor npr-1 is a major determinant of *Caenorhabditis elegans* growth and physiology. *PLoS Genet.* 10:e1004156. doi: 10.1371/journal.pgen.1004156
- Armstrong, M. D., and Tyler, F. H. (1955). Studies on phenylketonuria. I. Restricted phenylalanine intake in phenylketonuria. *J. Clin. Invest.* 34, 565–580. doi: 10.1172/JCI103105
- Bloom, J. S., Ehrenreich, I. M., Loo, W. T., Lite, T.-L. V., and Kruglyak, L. (2013). Finding the sources of missing heritability in a yeast cross. *Nature* 494, 234–237. doi: 10.1038/nature11867
- Boycott, A. E., Diver, C., Garstang, S. L., and Turner, F. M. (1931). The inheritance of sinistrality in limnaea peregra (Mollusca, Pulmonata). *Philos. Trans. R. Soc. Lond. Ser. B* 219, 51–131. doi: 10.1098/rstb.1931.0002
- Bravo, J. A., Forsythe, P., Chew, M. V., Escaravage, E., Savignac, H. M., Dinan, T. G., et al. (2011). Ingestion of *Lactobacillus* strain regulates emotional behavior and central GABA receptor expression in a mouse via the vagus nerve. *Proc. Natl. Acad. Sci. U.S.A.* 108, 16050–16055. doi: 10.1073/pnas.1102999108
- Bricelj, V. M., Connell, L., Konoki, K., Macquarrie, S. P., Scheuer, T., Catterall, W. A., et al. (2005). Sodium channel mutation leading to saxitoxin resistance in clams increases risk of PSP. *Nature* 434, 763–767. doi: 10.1038/nature03415
- Carroll, S. B., Grenier, J., and Weatherbee, S. (2005). *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design*. Malden, MA: John Wiley & Sons.
- Cassirer, E. (1910). *Substance et Fonction*, 2010 Edn. Paris: Nabu Press.
- Coen, E. (2012). *Cells to Civilizations: The Principles of Change That Shape Life*. Princeton, NJ: Princeton University Press. doi: 10.1515/9781400841653
- Colosimo, P. F., Hosemann, K. E., Balabhadra, S., Villarreal, G. Jr., Dickson, M., Grimwood, J., et al. (2005). Widespread parallel evolution in sticklebacks by repeated fixation of Ectodysplasin alleles. *Science* 307, 1928–1933. doi: 10.1126/science.1107239
- Dawkins, R. (1982). *The Extended Phenotype: The Long Reach of the Gene*. Oxford: Oxford University Press.
- Deans, A. R., Lewis, S. E., Huala, E., Anzaldo, S. S., Ashburner, M., Balhoff, J. P., et al. (2015). Finding our way through phenotypes. *PLoS Biol.* 13:e1002033. doi: 10.1371/journal.pbio.1002033
- Dunbar, H. E., Wilson, A. C. C., Ferguson, N. R., and Moran, N. A. (2007). Aphid thermal tolerance is governed by a point mutation in bacterial symbionts. *PLoS Biol.* 5:e96. doi: 10.1371/journal.pbio.0050096
- Engelman, C. D., Baurley, J. W., Chiu, Y.-F., Joubert, B. R., Lewinger, J. P., Maenner, M. J., et al. (2009). Detecting gene-environment interactions in genome-wide association data. *Genet. Epidemiol.* 33(Suppl. 1), S68–S73. doi: 10.1002/gepi.20475
- Falk, R. (1984). The gene in search of an identity. *Hum. Genet.* 68, 195–204. doi: 10.1007/BF00418388
- Feldman, C. R., Brodie, E. D. Jr., Brodie, E. D. III, and Pfrender, M. E. (2012). Constraint shapes convergence in tetrodotoxin-resistant sodium channels of snakes. *Proc. Natl. Acad. Sci. U.S.A.* 109, 4556–4561. doi: 10.1073/pnas.1113468109
- Ffrench-Constant, R. H., Daborn, P. J., and Le Goff, G. (2004). The genetics and genomics of insecticide resistance. *Trends Genet.* 20, 163–170. doi: 10.1016/j.tig.2004.01.003
- Fischbach, M. A. (2009). Antibiotics from microbes: converging to kill. *Curr. Opin. Microbiol.* 12, 520–527. doi: 10.1016/j.mib.2009.07.002
- Fisher, R. A. (1930). *The Genetical Theory of Natural Selection*. Oxford: Clarendon. doi: 10.5962/bhl.title.27468
- Fontana, W. (2002). Modelling “evo-devo” with RNA. *BioEssays News Rev. Mol. Cell. Dev. Biol.* 24, 1164–1177. doi: 10.1002/bies.10190
- Frankel, N., Erezymilaz, D. F., McGregor, A. P., Wang, S., Payre, F., and Stern, D. L. (2011). Morphological evolution caused by many subtle-effect substitutions in regulatory DNA. *Nature* 474, 598–603. doi: 10.1038/nature10200
- Geffeney, S. L., Fujimoto, E., Brodie, E. D. III, Brodie, E. D. Jr., and Ruben, P. C. (2005). Evolutionary diversification of TTX-resistant sodium channels in a predator-prey interaction. *Nature* 434, 759–763. doi: 10.1038/nature03444
- Gerstein, M. B., Bruce, C., Rozowsky, J. S., Zheng, D., Du, J., Korbel, J. O., et al. (2007). What is a gene, post-ENCODE? History and updated definition. *Genome Res.* 17, 669–681. doi: 10.1101/gr.6339607
- Gilbert, S. F. (2000). “Genes classical and genes developmental: the different uses of genes in evolutionary syntheses,” in *The Concept of the Gene in Development and Evolution*, eds P. Beurton, R. Falk, and H.-J. Rheinberger (Cambridge, NY: Cambridge University Press).
- Gilbert, S. F., and Epel, D. (2009). *Ecological Developmental Biology: Integrating Epigenetics, Medicine, and Evolution*. Sunderland, MA: Sinauer Associates.
- Gjuvsland, A. B., Vik, J. O., Beard, D. A., Hunter, P. J., and Omholt, S. W. (2013). Bridging the genotype-phenotype gap: what does it take? *J. Physiol.* 591, 2055–2066. doi: 10.1113/jphysiol.2012.248864
- Gompel, N., and Prud’homme, B. (2009). The causes of repeated genetic evolution. *Dev. Biol.* 332, 36–47. doi: 10.1016/j.ydbio.2009.04.040
- Graur, D., Zheng, Y., and Azevedo, R. B. R. (2015). An Evolutionary Classification of Genomic Function. *Genome Biol. Evol.* 7, 642–645. doi: 10.1093/gbe/evv021
- Griffiths, P., and Stotz, K. (2013). *Genetics and Philosophy: An Introduction*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511744082
- Gujas, B., Alonso-Blanco, C., and Hardtke, C. S. (2012). Natural *Arabidopsis* brx loss-of-function alleles confer root adaptation to acidic soil. *Curr. Biol.* 22, 1962–1968. doi: 10.1016/j.cub.2012.08.026
- Hansen, T. F. (2013). Why epistasis is important for selection and adaptation. *Evolution* 67, 3501–3511. doi: 10.1111/evo.12214
- Hollocher, H., Hatcher, J. L., and Dyreson, E. G. (2000). Genetic and developmental analysis of abdominal pigmentation differences across species in the *Drosophila dunni* subgroup. *Evol. Int. J. Org. Evol.* 54, 2057–2071. doi: 10.1111/j.0014-3820.2000.tb01249.x
- Hong, R. L., Witte, H., and Sommer, R. J. (2008). Natural variation in *Pristionchus pacificus* insect pheromone attraction involves the protein kinase EGL-4. *Proc. Natl. Acad. Sci. U.S.A.* 105, 7779–7784. doi: 10.1073/pnas.0708406105
- Houle, D., Govindaraju, D. R., and Omholt, S. (2010). Phenomics: the next challenge. *Nat. Rev. Genet.* 11, 855–866. doi: 10.1038/nrg2897
- Johannsen, W. (1911). The genotype conception of heredity. *Am. Nat.* 45, 129–159. doi: 10.1086/279202
- Jones, F. C., Grabherr, M. G., Chan, Y. F., Russell, P., Mauceli, E., Johnson, J., et al. (2012). The genomic basis of adaptive evolution in threespine sticklebacks. *Nature* 484, 55–61. doi: 10.1038/nature10944
- Kamberov, Y. G., Wang, S., Tan, J., Gerbault, P., Wark, A., Tan, L., et al. (2013). Modeling recent human evolution in mice by expression of a selected EDAR variant. *Cell* 152, 691–702. doi: 10.1016/j.cell.2013.01.016
- Kammenga, J. E., Doroszuk, A., Riksen, J. A. G., Hazendonk, E., Spiridon, L., Petrescu, A.-J., et al. (2007). A *Caenorhabditis elegans* wild type defies the temperature-size rule owing to a single nucleotide polymorphism in tra-3. *PLoS Genet* 3:e34. doi: 10.1371/journal.pgen.0030034
- Keller, E. F. (2010). *The Mirage of a Space between Nature and Nurture*. Durham, NC: Duke University Press. doi: 10.1215/9780822392811
- Kirschner, M., and Gerhart, J. (1998). Evolvability. *Proc. Natl. Acad. Sci. U.S.A.* 95, 8420–8427. doi: 10.1073/pnas.95.15.8420
- Kislukhin, G., King, E. G., Walters, K. N., Macdonald, S. J., and Long, A. D. (2013). The genetic architecture of methotrexate toxicity is similar in *Drosophila melanogaster* and humans. 3, 1301–1310. doi: 10.1534/g3.113.006619
- Landry, C. R., and Rifkin, S. A. (2012). The genotype-phenotype maps of systems biology and quantitative genetics: distinct and complementary. *Adv. Exp. Med. Biol.* 751, 371–398. doi: 10.1007/978-1-4614-3567-9_17
- Langlade, N. B., Feng, X., Dransfield, T., Copsey, L., Hanna, A. I., Thébaud, C., et al. (2005). Evolution through genetically controlled allometry space. *Proc. Natl. Acad. Sci. U.S.A.* 102, 10221–10226. doi: 10.1073/pnas.0504210102
- Lenser, T., and Theissen, G. (2013). Molecular mechanisms involved in convergent crop domestication. *Trends Plant Sci.* 18, 704–714. doi: 10.1016/j.tplants.2013.08.007
- Lettre, G. (2011). Recent progress in the study of the genetics of height. *Hum. Genet.* 129, 465–472. doi: 10.1007/s00439-011-0969-x
- Lewontin, R. (1974a). *The Genetic Basis of Evolutionary Change (Columbia Biological Series)*. Columbia: Columbia University Press.
- Lewontin, R. C. (1974b). The analysis of variance and the analysis of causes. *Am. J. Hum. Genet.* 26, 400–411.

- Lewontin, R. (1978). Adaptation. *Sci. Am.* 293, 212–228. doi: 10.1038/scientificamerican0978-212
- Linnen, C. R., Poh, Y.-P., Peterson, B. K., Barrett, R. D. H., Larson, J. G., Jensen, J. D., et al. (2013). Adaptive evolution of multiple traits through multiple mutations at a single gene. *Science* 339, 1312–1316. doi: 10.1126/science.1233213
- Liu, F., Wen, B., and Kayser, M. (2013). Colorful DNA polymorphisms in humans. *Semin. Cell Dev. Biol.* 24, 562–575. doi: 10.1016/j.semcd.2013.03.013
- Luong, H. T. T., Chaplin, J., McRae, A. F., Medland, S. E., Willemse, G., Nyholt, D. R., et al. (2011). Variation in BMPR1B, TGFRB1 and BMPR2 and control of dizygotic twinning. *Twin Res. Hum. Genet. Off. J. Int. Soc. Twin Stud.* 14, 408–416. doi: 10.1375/twin.14.5.408
- Lynch, M., and Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sunderland, MA: Sinauer Associates.
- MacCallum, D. M., Coste, A., Ischer, F., Jacobsen, M. D., Odds, F. C., and Sanglard, D. (2010). Genetic dissection of azole resistance mechanisms in *Candida albicans* and their validation in a mouse model of disseminated infection. *Antimicrob. Agents Chemother.* 54, 1476–1483. doi: 10.1128/AAC.01645-1649
- Manske, M., Miotti, O., Campino, S., Auburn, S., Almagro-Garcia, J., Maslen, G., et al. (2012). Analysis of *Plasmodium falciparum* diversity in natural infections by deep sequencing. *Nature* 487, 375–379. doi: 10.1038/nature11174
- Martin, A., and Orgogozo, V. (2013a). *Data from: The Loci of Repeated Evolution: A Catalogue of Genetic Hotspots of Phenotypic Variation*. Available at: <http://datadryad.org/resource/> doi: 10.5061/dryad.v66p0
- Martin, A., and Orgogozo, V. (2013b). The Loci of repeated evolution: a catalog of genetic hotspots of phenotypic variation. *Evol. Int. J. Orgn. Evol.* 67, 1235–1250. doi: 10.1111/evo.12081
- Mayr, E. (1963). *Animal Species and Evolution*. Harvard: Harvard University Press. doi: 10.4159/harvard.9780674865327
- McGary, K. L., Park, T. J., Woods, J. O., Cha, H. J., Wallingford, J. B., and Marcotte, E. M. (2010). Systematic discovery of nonobvious human disease models through orthologous phenotypes. *Proc. Natl. Acad. Sci. U.S.A.* 107, 6544–6549. doi: 10.1073/pnas.0910200107
- Mery, F., Belay, A. T., So, A. K.-C., Sokolowski, M. B., and Kawecki, T. J. (2007). Natural polymorphism affecting learning and memory in *Drosophila*. *Proc. Natl. Acad. Sci. U.S.A.* 104, 13051–13055. doi: 10.1073/pnas.0702923104
- Moss, L. (2003). One, Two (Too?), Many Genes? *Q. Rev. Biol.* 78, 57–67. doi: 10.1086/367581
- Narod, S. A., and Foulkes, W. D. (2004). BRCA1 and BRCA2: 1994 and beyond. *Nat. Rev. Cancer* 4, 665–676. doi: 10.1038/nrc1431
- Natarajan, C., Inoguchi, N., Weber, R. E., Fago, A., Moriyama, H., and Storz, J. F. (2013). Epistasis among adaptive mutations in deer mouse hemoglobin. *Science* 340, 1324–1327. doi: 10.1126/science.1236862
- Noble, D. (2006). *The Music of Life: Biology beyond the Genome*. Oxford: Oxford University Press.
- Osborne, K. A., Robichon, A., Burgess, E., Butland, S., Shaw, R. A., Coulthard, A., et al. (1997). Natural behavior polymorphism due to a cGMP-dependent protein kinase of *Drosophila*. *Science* 277, 834–836. doi: 10.1126/science.277.5327.834
- Oyama, S. (2000). *The Ontogeny of Information: Developmental Systems and Evolution*, 2nd Edn. Cambridge, NY: Duke University Press. doi: 10.1215/9780822380665
- Paaby, A. B., and Rockman, M. V. (2013). The many faces of pleiotropy. *Trends Genet.* 29, 66–73. doi: 10.1016/j.tig.2012.10.010
- Palmer, A. R. (2004). Symmetry breaking and the evolution of development. *Science* 306, 828–833. doi: 10.1126/science.1103707
- Paterson, A. H., Lin, Y. R., Li, Z., Schertz, K. F., Doebley, J. F., Pinson, S. R., et al. (1995). Convergent domestication of cereal crops by independent mutations at corresponding genetic Loci. *Science* 269, 1714–1718. doi: 10.1126/science.269.5231.1714
- Perutz, M. F. (1983). Species adaptation in a protein molecule. *Mol. Biol. Evol.* 1, 1–28.
- Phillips, P. C. (2008). Epistasis—the essential role of gene interactions in the structure and evolution of genetic systems. *Nat. Rev. Genet.* 9, 855–867. doi: 10.1038/nrg2452
- Pokorná, M., and Kratochvíl, L. (2009). Phylogeny of sex-determining mechanisms in squamate reptiles: are sex chromosomes an evolutionary trap? *Zool. J. Linn. Soc.* 156, 168–183. doi: 10.1111/j.1096-3642.2008.00481.x
- Robinson, P. N., and Webber, C. (2014). Phenotype ontologies and cross-species analysis for translational research. *PLoS Genet* 10:e1004268. doi: 10.1371/journal.pgen.1004268
- Rockman, M. V. (2012). The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evol. Int. J. Orgn. Evol.* 66, 1–17. doi: 10.1111/j.1558-5646.2011.01486.x
- Rogers, W. A., Salomone, J. R., Tacy, D. J., Camino, E. M., Davis, K. A., Rebeiz, M., et al. (2013). Recurrent modification of a conserved cis-regulatory element underlies fruit fly pigmentation diversity. *PLoS Genet* 9:e1003740. doi: 10.1371/journal.pgen.1003740
- Romero-Herrera, A. E., Lehmann, H., Joysey, K. A., and Friday, A. E. (1978). On the evolution of myoglobin. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 283, 61–163. doi: 10.1098/rstb.1978.0018
- Salazar-Ciudad, I., and Marín-Riera, M. (2013). Adaptive dynamics under development-based genotype-phenotype maps. *Nature* 497, 361–364. doi: 10.1038/nature12142
- Sarkar, S. (1999). From the reaktionsnorm to the adaptive norm: the norm of reaction, 1909–1960. *Biol. Philos.* 14, 235–252. doi: 10.1023/A:1006690502648
- Schwander, T., and Leimar, O. (2011). Genes as leaders and followers in evolution. *Trends Ecol. Evol.* 26, 143–151. doi: 10.1016/j.tree.2010.12.010
- Schwartz, S. (2000). “The differential concept of the gene: past and present,” in *The Concept of the Gene in Development and Evolution*, eds P. J. Beurton, R. Falk, and H. J. Rheinberger (Cambridge, NY: Cambridge University Press), 26–39. doi: 10.1017/CBO9780511527296.004
- Shirangi, T. R., Dufour, H. D., Williams, T. M., and Carroll, S. B. (2009). Rapid evolution of sex pheromone-producing enzyme expression in *Drosophila*. *PLoS Biol.* 7:e1000168. doi: 10.1371/journal.pbio.1000168
- Simondon, G. (1968). *L'individuation à la Lumière des Notions de Forme et d'Information*, 2005th Edn. Grenoble: Millon.
- Sober, E. (1988). Apportioning causal responsibility. *J. Philos.* 85, 303–318. doi: 10.2307/2026721
- Steiner, C. C., Weber, J. N., and Hoekstra, H. E. (2007). Adaptive variation in beach mice produced by two interacting pigmentation genes. *PLoS Biol.* 5:e219. doi: 10.1371/journal.pbio.0050219
- Stern, D. L. (2000). Evolutionary developmental biology and the problem of variation. *Evol. Int. J. Orgn. Evol.* 54, 1079–1091. doi: 10.1111/j.0014-3820.2000.tb00544.x
- Stern, D. L. (2013). The genetic causes of convergent evolution. *Nat. Rev. Genet.* 14, 751–764. doi: 10.1038/nrg3483
- Stern, D., and Orgogozo, V. (2008). The loci of evolution: How predictable is genetic evolution? *Evol. Int. J. Orgn. Evol.* 62, 2155–2177. doi: 10.1111/j.1558-5646.2008.00450.x
- Stewart, C. B., Schilling, J. W., and Wilson, A. C. (1987). Adaptive evolution in the stomach lysozymes of foregut fermenters. *Nature* 330, 401–404. doi: 10.1038/330401a0
- Stotz, K. (2012). Murder on the development express: who killed nature/nurture? *Biol. Philos.* 27, 919–929. doi: 10.1007/s10539-012-9343-1
- Sturtevant, A. H. (1932). The use of mosaics in the study of the developmental effects of genes. *Proc. Sixth Int. Congr. Genet. Ithaca N. Y.* 1, 304–307.
- Supple, M., Papa, R., Counterman, B., and McMillan, W. O. (2014). The genomics of an adaptive radiation: insights across the heliconius speciation continuum. *Adv. Exp. Med. Biol.* 781, 249–271. doi: 10.1007/978-94-007-7347-9_13
- Tautz, D., and Schmid, K. J. (1998). From genes to individuals: developmental genes and the generation of the phenotype. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* 353, 231–240. doi: 10.1098/rstb.1998.0205
- Thomas, D. (2010). Gene-environment-wide association studies: emerging approaches. *Nat. Rev. Genet.* 11, 259–272. doi: 10.1038/nrg2764
- Venkatesh, B., Lu, S. Q., Dandona, N., See, S. L., Brenner, S., and Soong, T. W. (2005). Genetic basis of tetrodotoxin resistance in pufferfishes. *Curr. Biol.* 15, 2069–2072. doi: 10.1016/j.cub.2005.10.068
- Visscher, P. M., Brown, M. A., McCarthy, M. I., and Yang, J. (2012). Five years of GWAS discovery. *Am. J. Hum. Genet.* 90, 7–24. doi: 10.1016/j.ajhg.2011.11.029

- Waddington, C. H. (1957). *The Strategy of the Genes*, 2014 Edn Reprint. London: Routledge.
- Wagner, G. P. (1996). Homologues, natural kinds and the evolution of modularity. *Am. Zool.* 36, 36–43. doi: 10.1093/icb/36.1.36
- Wagner, G. P. (2000). *The Character Concept in Evolutionary Biology*. San Diego, CA: Academic Press.
- Wagner, G. P. (2014). *Homology, Genes, and Evolutionary Innovation*. Princeton, NJ: Princeton University Press. doi: 10.1515/9781400851461
- Wang, J., Wurm, Y., Nipitwattanaphon, M., Riba-Grognuz, O., Huang, Y.-C., Shoemaker, D., et al. (2013). A Y-like social chromosome causes alternative colony organization in fire ants. *Nature* 493, 664–668. doi: 10.1038/nature11832
- Waters, C. K. (2007). Causes that make a difference. *J. Philos.* 104, 551–579.
- Welter, D., MacArthur, J., Morales, J., Burdett, T., Hall, P., Junkins, H., et al. (2014). The NHGRI GWAS catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 42, D1001–D1006. doi: 10.1093/nar/gkt1229
- West-Eberhard, M. J. (2003). *Developmental Plasticity and Evolution*. Oxford: Oxford University Press.
- West-Eberhard, M. J. (2005). Developmental plasticity and the origin of species differences. *Proc. Natl. Acad. Sci. U.S.A.* 102(Suppl. 1), 6543–6549. doi: 10.1073/pnas.0501844102
- Wilkins, A. (2014). “The genetic tool-kit: the life-history of an important metaphor,” in *Advances in Evolutionary Developmental Biology*, ed. J. Todd Streelman (Hoboken, NJ: John Wiley & Sons), 1–14.
- Wilkins, A. S. (2002). *The Evolution of Developmental Pathways*. Sunderland, MA: Sinauer Associates.
- Xu, X., Dong, G.-X., Hu, X.-S., Miao, L., Zhang, X.-L., Zhang, D.-L., et al. (2013). The genetic basis of white tigers. *Curr. Biol.* 23, 1031–1035. doi: 10.1016/j.cub.2013.04.054
- Zhang, J., and Wagner, G. P. (2013). On the definition and measurement of pleiotropy. *Trends Genet.* 29, 383–384. doi: 10.1016/j.tig.2013.05.002

Conflict of Interest Statement: 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 © 2015 Orgogozo, Morizot and Martin. 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) or licensor 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.

Advantages of publishing in Frontiers



OPEN ACCESS

Articles are free to read,
for greatest visibility



COLLABORATIVE PEER-REVIEW

Designed to be rigorous
– yet also collaborative,
fair and constructive



FAST PUBLICATION

Average 85 days from
submission to publication
(across all journals)



COPYRIGHT TO AUTHORS

No limit to article
distribution and re-use



TRANSPARENT

Editors and reviewers
acknowledged by name
on published articles



SUPPORT

By our Swiss-based
editorial team



IMPACT METRICS

Advanced metrics
track your article's impact



GLOBAL SPREAD

5'100'000+ monthly
article views
and downloads



LOOP RESEARCH NETWORK

Our network
increases readership
for your article

Frontiers

EPFL Innovation Park, Building I • 1015 Lausanne • Switzerland
Tel +41 21 510 17 00 • Fax +41 21 510 17 01 • info@frontiersin.org
www.frontiersin.org

Find us on

