



Evolutionary dynamics of FoxQ2 transcription factors across metazoans reveals three ancient paralogs



Giacomo Gattoni ^{1,5}, Che-Yi Lin², Joshua R. York³, Colin Shew ⁴, Daniel Keitley¹, Carole LaBonnie ³, Jr-Kai Yu ², J. Andrew Gillis^{1,6} & Elia Benito-Gutiérrez ^{1,4}

FoxQ2 is a highly conserved Forkhead-box transcription factor expressed anteriorly in cnidarians and bilaterians, yet its evolution is marked by rapid divergence and lineage-specific duplications or losses. Moreover, its presence and localization in vertebrate groups remains unclear. To reconcile these conflicting reports of conservation and divergence, we combine phylogenetic and synteny analyses of FoxQ2 sequences from 21 animal phyla. We uncover three ancient FoxQ2 paralogs in bilaterians—FoxQ2I, FoxQ2II, and FoxQ2III. All three were present in the chordate ancestor, and two are retained in vertebrates, indicating a richer FoxQ2 repertoire in vertebrates than previously recognized. To assess FoxQ2 expression, we analyzed mollusk, acoel, amphioxus, and zebrafish single-cell transcriptomic datasets, and conducted fluorescent *in situ* hybridization in amphioxus, lamprey, skate, zebrafish, and chicken. FoxQ2I and FoxQ2II show conserved anterior expression, while FoxQ2III is expressed in the gut endoderm in chordates, including amphioxus, lamprey, and skate. We also predict conserved transcription factor binding sites across amphioxus genera, revealing stage- and cell-type-specific regulatory interactions for FoxQ2I in deuterostomes. Overall, this work clarifies FoxQ2's evolutionary history, identifies the endodermally expressed paralog FoxQ2III, and proposes that early duplication of FoxQ2I/II enabled subfunctionalization, driving the fast evolutionary rate of FoxQ2 sequences observed in bilaterians.

FoxQ2 is a transcription factor that plays crucial roles in antero-posterior patterning and nervous system development across diverse animal groups, yet its evolutionary history and diversification remain to be fully understood. FoxQ2 is part of the ancient and highly conserved Forkhead-box (FOX) family of transcription factors, which likely originated in the common ancestor of Opisthokonta, as it has been identified in fungi, choanoflagellates and animals¹. Within the metazoan lineage, the repertoire of FOX proteins expanded dramatically to include ~26 classes, named with letters from A to S². These proteins are characterized by the presence of a conserved forkhead or winged-helix DNA-binding domain and are involved in virtually all developmental processes as well as in metabolism and in the regulation of cell cycle^{3,4}. The forkhead motif is conserved in metazoans, composed of three alpha helices and three beta-sheets, but each class can be distinguished by subtle differences within and outside the DNA-binding domain^{5–7}.

The FoxQ2 class was initially described in the cephalochordate amphioxus, when phylogenetic analysis of a newly discovered *FoxQ* sequence revealed it belonged to a distinct group from *FoxQ1* genes⁸. Since then, *FoxQ2* orthologs have been found in most animal phyla studied to date, suggesting they first originated during early metazoan evolution^{9–11} (Supplementary Fig. 1). Recent studies have also highlighted the complex evolutionary history of this Fox class, marked by numerous taxon-specific duplications and losses, as well as rapid sequence divergence. Together, these factors complicate the understanding of how many paralogs are ancestral to specific groups^{11–13}. In contrast to this highly dynamic sequence evolution, the analysis of *FoxQ2* expression in a variety of invertebrates has revealed a remarkable conservation in the localization of *FoxQ2* transcripts within the anterior portion of the body (aboral for cnidarians) during early development^{8,14–20}. This expression pattern appears to reflect FoxQ2's function in the specification of anterior ectodermal identities, which has

¹Department of Zoology, University of Cambridge, Cambridge, UK. ²Institute of Cellular and Organismic Biology, Academia Sinica, Taipei, Taiwan. ³Department of Molecular Biosciences, Northwestern University, Evanston, IL, USA. ⁴Neuroscience, Genentech, South San Francisco, CA, USA. ⁵Present address: Department of Biological Sciences, Columbia University, New York, NY, USA. ⁶Present address: Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biological Laboratory, Woods Hole, MA, USA. e-mail: gg2900@columbia.edu; eb647@cam.ac.uk

been investigated in detail only in a few deuterostomes (e.g., echinoderms) and protostomes (e.g., arthropods)^{21–31}. Notably, in many marine invertebrates *FoxQ2* has been shown to be part of anterior gene regulatory network (aGRN) involved in the specification of the anterior neuroectoderm^{16,28,32–37}.

With the expansion of developmental analyses across an increasing number of taxa, the interest in the expression and function of *FoxQ2* has grown in recent years, though information about this conserved class remains fragmented. Here we comprehensively examine the complex evolutionary history and expression pattern of *FoxQ2* genes across metazoans, with a specific focus on chordates - the phylum in which *FoxQ2* was initially discovered but remains scarcely explored. Our phylogenetic analysis, which includes sequences from most animal phyla and is coupled with synteny analysis, identified three *FoxQ2* paralogs shared by bilaterians and cnidarians with numerous subsequent duplications and losses. We further characterize the expression of *FoxQ2* paralogs across bilaterians, focusing on different chordate lineages, revealing the existence of two ancestral *FoxQ2* genes in vertebrates, one with a conserved anterior ectodermal expression and another, previously unidentified, expressed in the endoderm. Finally, the here augmented information on *FoxQ2* expression and function is re-examined in the context of our phylogenetic findings.

Results

High conservation but dynamic evolution of the *FoxQ2* class

To date, *FoxQ2* genes have been identified in species belonging to 14 phyla (Supplementary Fig. 1), including cnidarians^{14,15,38}, spirilians (mollusks, annelids, phoronids, brachiopods, nemerteans, orthonectida and platyhelminths)^{16,17,19,20,39–42}, ecdysozoans (arthropods, onychophorans and nematodes)^{18,27,29,43} and deuterostomes (echinoderms, hemichordates, chordates)^{8,13,24,32,44–47}. Their presence in poriferans and ctenophores has also been suggested⁴⁸.

To expand the repertoire of known *FoxQ2* genes, and to further test the conservation of this class across Metazoa, we searched for orthologs in published genomes of taxa where *FoxQ2* had not been previously described. Among invertebrates, we found *FoxQ2* orthologs in spirilians (gastropod mollusks, clitellate annelids, rotifers, bryozoans), ecdysozoans (tardigrada) and xenacoelomorphs. Additionally, we identified *FoxQ2* paralogs in most non-bilaterian metazoans, including poriferans, ctenophores and cnidarians (Supplementary Data 1). We then turned our attention to vertebrates, for which information on this gene was surprisingly scarce until recently, leading some researchers to speculate that this gene was lost in tetrapods or amniotes^{11,12}. Motivated by the recent discovery of *FoxQ2* orthologs in teleosts, coelacanths, sauropsids and monotreme mammals⁴⁹, we BLAST-searched the genomes of the lamprey *Petromyzon marinus*⁵⁰, the skate *Leucoraja erinacea*⁵¹, and two amphibians, *Xenopus laevis* and *Pleurodeles walti*⁵². We found a single *FoxQ2* copy in the genomes of the lamprey and skate, while no ortholog was found in amphibians. Contrary to the previously accepted scenario, our search indicates that most vertebrate groups possess *FoxQ2* genes, and that these were secondarily lost in amphibians and placental mammals independently.

In sum, combining our BLAST-based search with sequences identified in previous publications, we retrieved candidate *FoxQ2* orthologs from 70 species belonging to 21 phyla. All these contained forkhead domains clearly identified as belonging to the *FoxQ2* class (Supplementary Fig. 1A, Supplementary Data 1, see “Materials and Methods”). The number of *FoxQ2* paralogs in each species varies considerably both between and within phyla (Supplementary Fig. 1B). The genome of four sponge and two ctenophore species analyzed contained a single copy of *FoxQ2*, while most cnidarian species analyzed had more than one paralog. All xenacoelomorph species in our analysis had only a single copy. Within protostomes, all ecdysozoans analyzed only have a single *FoxQ2* copy, while in spirilians the number of paralogs is highly variable between and within each phylum, with the annelid *Paraescarpia echinospica* currently holding the record with 13 *FoxQ2* genes¹¹. Among deuterostomes, echinoderms have one or two copies and hemichordates vary from one to four, while most chordates have a single *FoxQ2* gene, with the exception of amphioxus that has three paralogs.

Despite the broad phylogenetic distances and the varying number of paralogs, we found the predicted secondary structures of *FoxQ2* forkhead domains are highly conserved in deuterostomes, protostomes and xenacoelomorphs (Supplementary Fig. 2). In each species, the domain features a sequence of (fold – sheet – fold – fold – sheet – sheet – fold) with only minor variations in length and amino acid composition. This suggests that differences in the expression pattern and function across species and paralogs might be related to sequence changes outside the forkhead domain or to the regulation of *FoxQ2* gene expression.

Phylogenetic and microsynteny analyses recover three *FoxQ2* paralogs in metazoans

The highly variable number of *FoxQ2* paralogs in several phyla underscores the complex evolutionary history of the *FoxQ2* class. This raises the question of how many paralogs can be traced back to the common ancestors of metazoans, bilaterians, protostomes and deuterostomes respectively, and which paralogs instead evolved independently in specific lineages. To address these questions, previous studies have examined the phylogenetic relationships of *Fox* genes in cnidarians and bilaterians using an increasingly high number of species^{11–13}. Initially, the *FoxQ2* class was hypothesized to consist of two ancestral *FoxQ2* groups, distinguished by the presence and position of an Engrailed Homology 1 (EH)-i-like motif at the N-terminal or C-terminal of the protein^{12,13}. Accordingly, these groups were named *FoxQ2-N* and *FoxQ2-C*, or *FoxQ2* and *FoxQD*, respectively. However, more recent analyses revealed that while *FoxQ2-C/FoxQD* proteins, with a C-terminal EH-i-like motif, appear to form a single monophyletic group, many other *FoxQ2* sequences are highly divergent, lacking an EH-i-like motif and distributing on multiple branches of the phylogenetic tree¹¹. This suggests a fast evolutionary rate, further evidenced by the presence of lineage-specific expansions that complicate the reconstruction of the evolutionary history of this class.

To reconcile these contrasting results, we aimed to reconstruct the evolutionary relationship of *FoxQ2* genes with a more comprehensive taxonomic sampling. We thus performed two parallel phylogenetic analyses incorporating the additional sequences we identified from bilaterians, cnidarians, ctenophores and poriferans. First, from our *FoxQ2* transcript database of 70 species, we selected those for which complete *FoxQ2* sequences were available. This resulted in the analysis of sequences from 33 species belonging to 17 different phyla: 7 Spiralia, 3 Ecdysozoa, 3 Deuterostomia, Xenacoelomorpha, Cnidaria, Ctenophora and Porifera (Fig. 1). To further expand the taxonomic sampling and include taxa for which only partial gene sequences were available, we also isolated and aligned the *FoxQ2* forkhead domain of 47 species from 21 animal phyla—9 Spiralia, 5 Ecdysozoa, 3 Deuterostomia, Xenacoelomorpha, Cnidaria, Ctenophora and Porifera (Supplementary Fig. 3). Phylogenetic trees were constructed using Maximum Likelihood (ML) (Fig. 1, Supplementary Fig. 3A) and Neighbor Joining (NJ) (Supplementary Fig. 3B, C) methods, using *FoxQ1*, *FoxL1* and *FoxP* as outgroups. All the sequences retrieved in this study branched within the *FoxQ2* clade, and the broad structure of the phylogenetic trees was highly robust across groups of sequences and phylogenetic methods, supporting their identification as *FoxQ2* orthologs (Fig. 1, Supplementary Fig. 3). Moreover, the tree structure remained consistent when considering only deuterostomes or only protostomes (Supplementary Fig. 4A, B) and aligned with the unrooted tree and results from dimensionality reduction approaches (Supplementary Fig. 4C, D). Our expanded analysis revealed the existence of three ancient groups of *FoxQ2* genes, all present in bilaterians and cnidarians, that likely originated near the root of the metazoan tree. These three branches were recovered in all analyses, although their relative position on the tree varied depending on whether full sequences or forkhead domains were considered (Supplementary Fig. 3). We define these as three distinct *FoxQ2* paralogs named *FoxQ2I*, *FoxQ2II*, and *FoxQ2III* (Fig. 1).

The *FoxQ2I* family includes many fast-evolving genes previously identified as *FoxQ2-N*¹¹. Both ML and NJ trees retrieved *FoxQ2I* as a monophyletic clade, supporting previous analyses by Pascual-Herrera et al.¹² (Fig. 1, Supplementary Fig. 3). However, the main node had lower

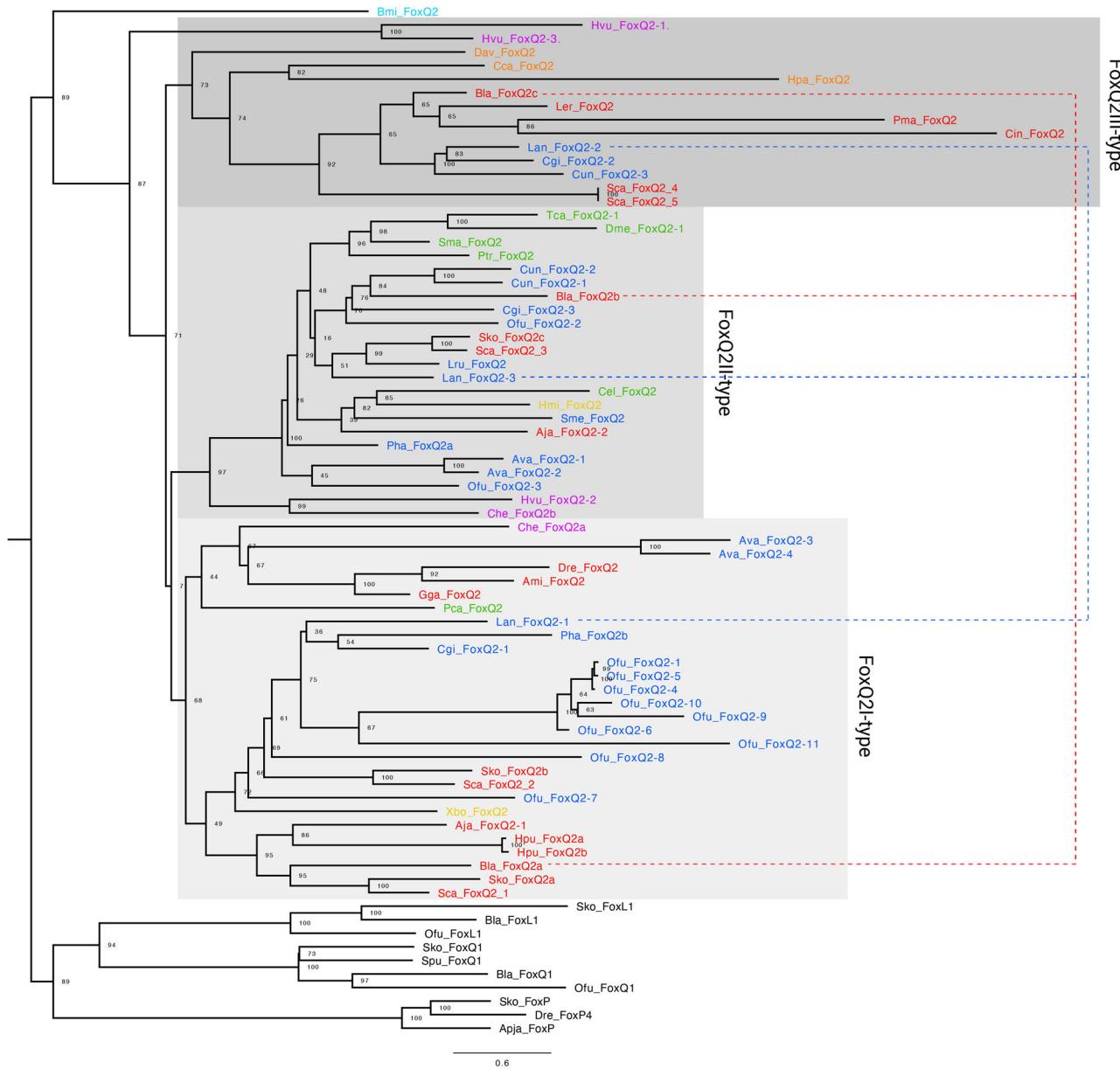


Fig. 1 | Phylogenetic analysis of metazoan FoxQ2 genes. Maximum likelihood tree topology based on *FoxQ2* complete gene sequences from 33 species belonging to 15 animal phyla. Individual genes are colored based on species taxonomy into deuterostomes (red), spiralian (blue) and ecdysozoan (green) protostomes, xenacoelomorphs (yellow), cnidarians (purple), poriferans (orange) or ctenoforans (cyan). Gray-shaded boxes demarcate the three *FoxQ2* types, *FoxQ2I*, *FoxQ2II* and

FoxQ2III. Dotted lines highlight conservation of all three types in representative deuterostome (amphioxus, red) and protostome (brachiopod, blue) species. *FoxQ1* and *FoxP* genes are used as outgroup. Full species names are listed in Supplementary Data 1. Scale bar indicates the number of amino acids substitutions per site.

bootstrap values in most methods, suggesting fast sequence evolution and high divergence as proposed by Seudre et al.¹¹. The family was identified in 12 metazoan phyla, including protostomes, deuterostomes, xenacoelomorphs, cnidarians and sponges (Fig. 1, Supplementary Fig. 3A). *FoxQ2I* was present in most spiralians, including mollusks, annelids, brachiopods, phoronids and rotifers, often in high copy number, but was not recovered from flatworm, nemertean, orthonectida and bryozoan genomes. Moreover, its presence was variable even within spiralian phyla; for example, *FoxQ2I* was present in bivalve but not gastropod mollusks, and in several marine annelids but not in clitellates (Supplementary Data 1). Among ecdysozoans, *FoxQ2I* was found only in priapulids and was absent from all other lineages considered. However, the robust position of the priapulid sequence in all analyses supports *FoxQ2I* conservation in both protostome branches, followed by loss in most ecdysozoan lineages (Fig. 1,

Supplementary Figs. 3, 4). *FoxQ2I* orthologs were identified in all three deuterostome phyla, with independent duplications in hemichordates and sea urchin. Within chordates, the first *FoxQ2* gene discovered in amphioxus belonged to this clade, together with vertebrate *FoxQ2* orthologs in ray-finned fishes, reptiles and birds, while no paralog could be found in tunicsates, cyclostomes and cartilaginous fishes (Fig. 1).

FoxQ2II includes the highly-conserved *FoxQ2-C/FoxQD* genes that were recovered as a monophyletic clade in accordance with all previous analyses^{11–13} (Fig. 1). We identified *FoxQ2II* orthologs in 18 out of the 21 bilaterian phyla considered - with the exception of priapulids, ctenophores and poriferans - indicating a higher level of conservation of this gene compared to the other two *FoxQ2* families (Supplementary Data 1). Compared to *FoxQ2I*, *FoxQ2II* genes are generally present in lower copy number, with a single gene in each species and rare duplications in some

annelid, mollusk and rotifer species. Despite the high conservation across bilaterians, it is also worth noting that within deuterostome phyla the gene has been lost in specific lineages, such as eleutherozoan echinoderms (starfish, sea urchins, sea cucumbers, brittlestars) and Olfactores (tunicates + vertebrates) chordates.

Surprisingly, our analysis identified a third, previously undescribed *FoxQ2* family: *FoxQ2III*. This family consistently branched separately from the other two in all phylogenetic analyses performed in this study (Fig. 1, Supplementary Figs. S3, S4). When considering full sequences, both ML and NJ trees recover *FoxQ2III* as a sister clade to the rest of the class (*FoxQ2I* + *FoxQ2II*) with high bootstrap values (Fig. 1). The forkhead domain analysis instead showed the *FoxQ2III* family nested within the tree as a sister clade to *FoxQ2II*, although with lower support (Supplementary Fig. 3A). Moreover, cnidarian *FoxQ2III*-type genes branched separately in the analysis with full sequences but were together with bilaterian and poriferan *FoxQ2III* genes when considering only the forkhead domain, a discrepancy likely due to high sequence divergence (Fig. 1, Supplementary Fig. 3A). While orthologs of the other two *FoxQ2* families were found in most metazoan phyla analyzed, *FoxQ2III* was recovered only from specific lineages: brachiopods and mollusks among protostomes; chordates and hemichordates among deuterostomes; hydrozoan and scyphozoan cnidarians; and homoscleromorph sponges (Fig. 1; Supplementary Fig. 3). Supporting this subdivision, a branch including *FoxQ2III*-type sequences from amphioxus and bivalve mollusks is also visible in a previous phylogeny of *FoxQ2* published by Seudre et al.¹¹. However, the low number of sequences in that study likely prevented its identification as a separate clade. Using additional sequences from cnidarians, mollusks, brachiopods, hemichordates, tunicates and vertebrates allowed this clade to be clearly visualized in our study. Among vertebrates, the *FoxQ2III* branch included only the cyclostome and cartilaginous fish sequences identified in this study, which branched separately from the *FoxQ2I* sequences found in bony fishes (Fig. 1).

To confirm the presence of two separate *FoxQ2* paralogs in vertebrates (*FoxQ2I* and *FoxQ2III*), we performed synteny analysis of *FoxQ2* genes in four chordate species: amphioxus (*Branchiostoma lanceolatum*) (which possesses *FoxQ2* paralogs from all three families); zebrafish (*Danio rerio*) that has a *FoxQ2I*-type gene according to our analysis; lamprey (*Petromyzon marinus*) and skate (*Leucoraja erinacea*), which instead only have *FoxQ2III* genes (Fig. 2A). Strikingly, we found that *FoxQ2III*-type paralogs are located in a similar genomic region in amphioxus, lamprey and skate, characterized by the close proximity to *FoxP* genes. Conversely, the conserved gene synteny of *FoxQ2I* previously identified in bony fishes⁴⁹ cannot be traced back to amphioxus (see below), further supporting the high divergence of this paralog.

In the three non-bilaterian phyla considered (Cnidaria, Ctenophora, Porifera) the assignment of *FoxQ2* sequences to *FoxQ2I*-, *FoxQ2II*- and *FoxQ2III*-types proved more challenging than in bilaterians, likely due to a high level of sequence divergence compared to bilaterian *FoxQ2* orthologs. Cnidarian sequences were found in all three *FoxQ2* branches, but in certain tree configurations *FoxQ2III*-types in *Hydra vulgaris* branched as sisters to all other *FoxQ2* groups (Fig. 1, Supplementary Fig. 3). Similarly, in all analyses the ctenophore *FoxQ2* sequence branched separately at the base of the *FoxQ2* tree, indicating a particularly high sequence divergence from other *FoxQ2* sequences (Fig. 1, Supplementary Fig. 3). For poriferans, the position of *FoxQ2* sequences varied based on the type of analysis: for example, *FoxQ2* from *Halichondria panicea* branched with *FoxQ2III*-type when considering full sequences (Fig. 1), but branched with *FoxQ2I* when considering only the forkhead domain (Supplementary Fig. 3A).

Evolutionary origin and duplication of *FoxQ2* paralogs revealed by macro-synteny

Recent comparative analyses of chromosome-level genomes from diverse bilaterians, cnidarians, sponges and ctenophores have revealed ancient chromosome-scale syntenies conserved across metazoan animals^{53–55}. We reasoned that these conserved macro-synteny patterns could provide

valuable information for tracing the evolutionary history of *FoxQ2* genes, including the more elusive orthologs in basally-branching non-bilaterian metazoans, further strengthening and enhancing the results of our phylogenetic reconstruction. Given this, we performed macro-synteny analysis to compare the correspondence of ancestral linkage groups (ALGs) of *FoxQ2*-bearing chromosomes across 25 metazoan species spanning bilaterian, cnidarian, poriferan and ctenophore phyla (Fig. 2B, Supplementary Data 2)⁵⁴. We found that *FoxQ2I*- and *FoxQ2II*-type genes were usually located in chromosomes originated from ALG_C1 (Fig. 2B, chromosomes connected by blue ribbons), while *FoxQ2III*-type genes were located on chromosomes corresponding to ALG_E/Eb (Fig. 2B, chromosomes connected by green ribbons) in chordates, hemichordates, mollusks, and cnidarians. Within the chordate lineage, vertebrates experienced dynamic gene loss of *FoxQ2* genes. For instance, we observed that the remaining *FoxQ2I* gene in spotted gar was still residing in ALG_C1-derived chromosome, while the surviving *FoxQ2III*-type genes in lamprey and skate are located on chromosome segments derived from ALG_E (Fig. 2B, Supplementary Data 2). This is consistent with the notion that individual *FoxQ2* genes associated with different ALGs seem to have distinct evolutionary trajectories.

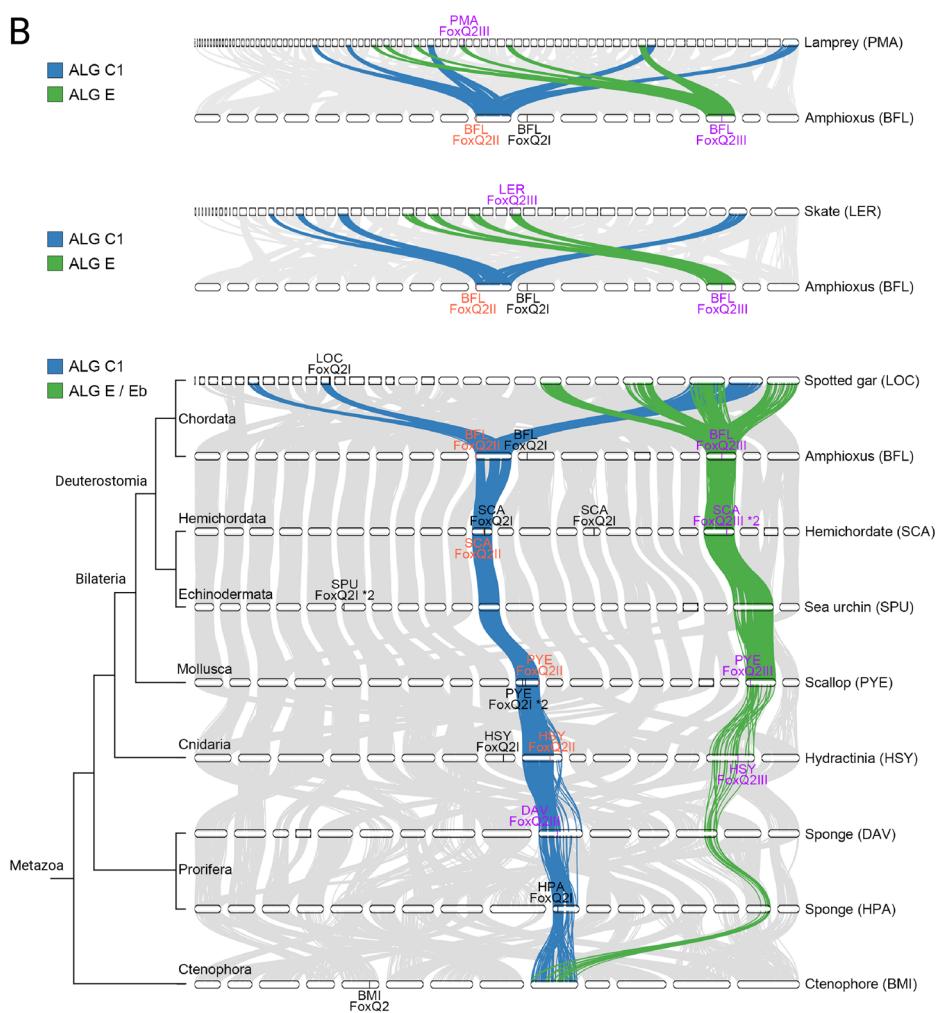
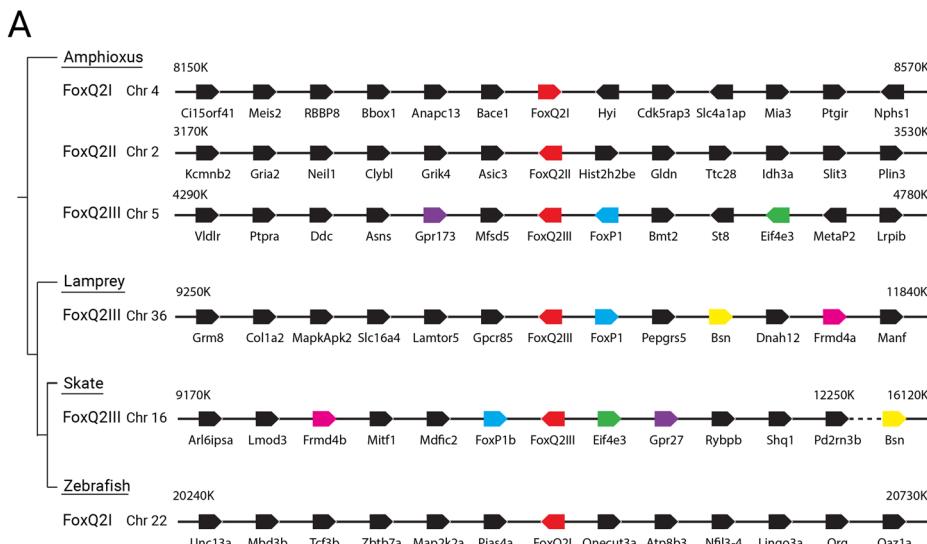
In poriferans, the *FoxQ2* sequences in the demosponges *Halichondria panicea* (HPA) and *Dysidea avara* (DAV) were found in ALG_C1-derived chromosomes (Fig. 2B, Supplementary Data 2), supporting the hypothesis that ancestral *FoxQ2* genes were already associated with ALG_C1 in the common ancestor of Porifera+Cnidaria+Bilateria. However, we found that *FoxQ2I*-type sequences in homoscleromorphs *Corticium candelabrum* (CCA) and *Oscarella lobularis* (OLO) were located on ALG_P-derived chromosomes (Supplementary Data 2), possibly due to a lineage-specific gene translocation event. In addition, the ctenophore *FoxQ2* genes are located on a chromosome formed by the fusion of ALG_L and ALG_M. This genomic position was unique compared to those in all the metazoan genomes that we have examined, and thus could not allow us to distinguish between various possibilities regarding the position of *FoxQ2* genes in the metazoan common ancestor (see “Discussion”). Therefore, although *FoxQ2* is present in this phylum, its relationship to the paralogs in other metazoans remains unclear.

The presence of *FoxQ2I* and *FoxQ2II* in the same ALG and the absence of a clear *FoxQ2II* paralog in early-branching metazoans (poriferans and ctenophores) also suggest that *FoxQ2I* and *FoxQ2II* share a closer evolutionary relationship, and they might have evolved by a tandem duplication event on ALG_C1 in the ancestor of Parahoxozoa (Cnidaria and Bilateria). The complementary distribution of *FoxQ2I* and *FoxQ2II* paralogs in ALG_C1-derived chromosomes in different species (Supplementary Data 2) may reflect an evolutionary process predicted by the duplication-degeneration-complementation model⁵⁶. Moreover, from this ancestral condition, *FoxQ2* paralogs have been considerably re-shuffled in selected lineages (Supplementary Data 2). Interestingly, we observed that *FoxQ2I*-type genes appeared to have undergone more frequent changes in their genomic position in various lineages, coinciding with their fast sequence evolution and high divergence in copy numbers among animals. For example, we identified duplicated *FoxQ2I*-type genes in both the hemichordate *Schizocardium californicum* (SCA) and the mollusk *Patinopecten yessoensis* (PYE) genomes. While in the scallop PYE, both *FoxQ2I*-type paralogs remained on the ALG_C1-derived chromosome, the *FoxQ2I*-type paralogs in the hemichordate SCA translocated to an ALG_O1-derived chromosome (Fig. 2B, Supplementary Data 2). In echinoderms, *FoxQ2I* was translocated to ALG_H and ALG_O1 in sea urchin and sea star, respectively, possibly indicating independent events in different echinoderm lineages. In amphioxus, *FoxQ2I* was translocated to a different ALG, ALG_A, which explains its different genomic location identified with microsynteny analysis.

Taken together, phylogenetic, microsynteny and macro-synteny analyses all support the presence of three ancient *FoxQ2* paralogs in metazoans, with two distinct groups (*FoxQ2I* + *FoxQ2II* and *FoxQ2III*) dating back to the ancestor of poriferans, cnidarians and bilaterians.

Fig. 2 | Synteny of FoxQ2 genes. A Microsynteny analysis of the genomic environment around *FoxQ2* genes in chordates. Orthologous genes across species are indicated by color-coding the corresponding box. Chromosomal number and location of the area analyzed are indicated for each gene.

B Macrosyntenic orthology relationships of metazoan chromosomes in representative species of chordates, hemichordates, echinoderms, mollusks, cnidarians, ctenophora and porifera, highlighting two ancestral linkage groups (ALGs) where *FoxQ2I* and *FoxQ2II* (ALG C1) and *FoxQ2III* (ALG E/Eb) likely resided ancestrally. Full species names are listed in Supplementary Data 1.



Expression of *FoxQ2* paralogs across bilaterians: a conserved role in anterior development?

Following the first description of *FoxQ2* expression in the cephalochordate amphioxus⁸, its spatial distribution during development has been investigated in 12 phyla (Supplementary Fig. 5). The discovery of three *FoxQ2* paralogs allowed us to re-evaluate the expression pattern of

FoxQ2 genes previously reported in the literature for bilaterians. To the best of our knowledge, information on the localization of *FoxQ2I* is available for all deuterostome groups (chordates, echinoderms, hemichordates)^{8,13,24,31,32,46,49,57} as well as annelids, mollusks and phoronids^{11,19,39}. *FoxQ2II* expression has been investigated in hemichordates, annelids, mollusks, nemerteans, brachiopods, platyhelminths,

arthropods and onychophorans^{11,13,16–18,20,28,29,39,42,43,58,59}. Although these studies have rarely considered the phylogenetic relationships among *FoxQ2* genes, they generally suggest a conserved expression of *FoxQ2I*- and *FoxQ2II*-type genes in the anterior ectoderm. Conversely, the spatial expression of *FoxQ2III*-type genes has not been investigated in any species. By combining bulk RNA-seq, scRNAseq and in situ hybridization, here we characterize the expression of *FoxQ2I*, *FoxQ2II* and *FoxQ2III* paralogs in multiple bilaterians.

As we hypothesized a high level of conservation in the anterior localization of *FoxQ2I* and *FoxQ2II* in bilaterians, we first queried publicly available scRNAseq datasets to detect the expression of *FoxQ2I* and *FoxQ2II* paralogs in two species for which expression data is not available: the zebra mussel *Dreissena polymorpha* (bivalve mollusk)⁶⁰ and the acoel *Hofstenia miamia* (xenacoelomorph)⁶¹ (Supplementary Fig. 6). *D. polymorpha* has 6 *FoxQ2* paralogs, including 4 *FoxQ2I*, 1 *FoxQ2II* and 1 *FoxQ2III*. We re-analyzed the data available for trophophore larvae, using the marker genes provided in the original publication to re-annotate the dataset⁶⁰, then plotted the expression of all six paralogs (Supplementary Fig. 6A). Of these, only one *FoxQ2I* paralog was expressed at significant levels, and we found that it labeled neuronal cells. The other *FoxQ2I* paralogs and *FoxQ2II* were instead detected only in scattered cells of the neural and ciliary ectoderm clusters. *FoxQ2* has not been investigated in the phylum Xenacoelomorpha, but the (debated) phylogenetic position of this group as the sister group to the other bilaterians makes it interesting to evaluate the evolutionary history of this gene⁶². We found that the genome of *H. miamia* contains a single *FoxQ2* sequence belonging to the *FoxQ2II* clade (Fig. 1). scRNAseq data of *H. miamia* hatchling juveniles is available as an interactive dataset. We therefore plotted the expression of *FoxQ2* (annotated as 98012160_foxb1 in the genome) and found expression in scattered cells within neoblast and neural clusters (Supplementary Fig. 6B). By referring to the spatial mapping described in the original paper^{61,63}, we observed that the neural clusters containing *FoxQ2*-positive cells were primarily located on the anterior portion of the acoel's body.

We next turned our attention to chordates, where data on *FoxQ2* expression is scarce. In amphioxus, previous studies have shown that *FoxQ2I* starts to be expressed at the blastula stage, immediately following the maternal to zygotic transition, across the entire animal side of the embryo. Its expression then progressively restricts to the antero-dorsal side throughout gastrulation and neurulation^{8,32,64}. From late gastrula (G4) to early neurula (3-4ss) stages, *FoxQ2* is expressed in both neural and non-neural ectoderm, but by the mid neurula (7ss) stage its expression becomes restricted to the anterior epidermis, and at early larval stages it remains expressed at the tip of the rostrum and in a small portion of the mouth³². Conversely, the expression of *FoxQ2II* and *FoxQ2III* has not been reported in the literature. By combining transcriptomic approaches and in situ hybridization chain reaction (HCR), here we show that in the European amphioxus *B. lanceolatum*, *FoxQ2II* starts to be expressed during the early phases of neurulation within the domain of *FoxQ2I* (Fig. 3Ai, Supplementary Fig. 7A). Specifically, in the early neurula *FoxQ2II* is expressed in the most anterior portion of the neural plate and at the border between neural and non-neural ectoderm. As *FoxQ2I* restricts outside of the neural plate at the 7ss stage, *FoxQ2II* is also found only in the anterior epidermis, and in the larva it is co-expressed with *FoxQ2I* in the anterior rostrum (Fig. 3Aii–iii). The expression of both *FoxQ2I* and *FoxQ2II* decreases during late development, as shown by RNAseq analysis⁶⁵ (Supplementary Fig. 7A, D). This data is corroborated by our analysis of a published developmental scRNAseq dataset of a different amphioxus species, *B. floridae*⁶⁶, which show early and widespread expression of *FoxQ2I* in neural and non-neural ectoderm, and a much more sparse and restricted expression of *FoxQ2II* (Supplementary Fig. 7B).

Until recently, no information on the expression of *FoxQ2* was available for any vertebrate, and the gene was thought to be lost in several lineages. Although *FoxQ2I* orthologs had been detected in several species of ray-finned fishes, they were not found to be expressed during early zebrafish

development, in contrast with its early distribution in many invertebrate deuterostomes. However, analysis of bulk RNA-seq and scRNAseq datasets that include late developmental stages shows that zebrafish *FoxQ2I*-type starts to be expressed after hatching in photoreceptor precursor and mature cells (Supplementary Fig. 7F)^{67,68}. A recent paper provided a detailed description of the expression and function of *FoxQ2* in zebrafish early larva (3–5 dpf) and showed that it is localized in blue cone cells, where it is essential to establish their identity⁴⁹. To date, this remains the only description of *FoxQ2I* expression in a vertebrate. By using in situ HCR, here we detected and compared the expression of *FoxQ2I*-type in later larval (8 dpf) and adult zebrafish as well as in chicken embryos (E8 and E16 stages) (Fig. 3B, C). The analysis of zebrafish *FoxQ2I* confirmed its expression in the larval photoreceptor layer of the retina (Fig. 3Bi), showing that it persists into adulthood and is thus not limited to development, while transcripts are absent in the brain (Fig. 3Bii, Supplementary Fig. 8A). Turning to the chicken embryo, we found a very similar distribution of *FoxQ2I* in photoreceptor progenitors at E8 (Fig. 3Ci), and later in the mature photoreceptor layer at E16 (Fig. 3Cii), indicating a conservation in retinal expression of *FoxQ2I*-type genes across bony fishes.

An endodermal domain of *FoxQ2III* expression

As indicated above, although at least four bilaterian phyla (mollusks, brachiopods, hemichordates, chordates) possess *FoxQ2III*-type genes, to the best of our knowledge no previous study has investigated their spatial expression. RNA-seq data across the development of bivalve mollusks and cephalochordates suggest that these genes are expressed between gastrula and larval phases^{11,40,41,65}; however, this does not provide information on the identity and location of *FoxQ2III*-positive cells. We therefore analyzed the localization of *FoxQ2III* in amphioxus (*B. lanceolatum*) using in situ HCR. Surprisingly, we found expression in a restricted domain within the endoderm of the early larva (12–14ss) which persisted at the one gill slit larval stage (Fig. 4Ai–ii). In particular, *FoxQ2III*-positive cells are located within the midgut, just anterior to the intestinal thickening. This result is corroborated by recent scRNAseq data during development of *B. floridae*, showing *FoxQ2III* in the endoderm, and particularly in the midgut at the 14ss (T1) stage (Supplementary Fig. 8A)^{66,69}. While *FoxQ2I* and *FoxQ2II* expression decreases during late development, *FoxQ2III* remains active at high level in the endoderm even in the adult, where bulk RNA-seq shows it remains localized in gut tissues (Supplementary Fig. 9A)⁶⁵. By investigating *FoxQ2III* expression in scRNAseq of the adult *B. floridae* digestive tract⁷⁰, we found that *FoxQ2III* is localized in gut epithelial cells enriched in the midgut (Fig. 4D).

Given the surprising localization of *FoxQ2III* in the amphioxus endoderm, we next sought to test whether this is an amphioxus-specific trait, or if *FoxQ2III* is also expressed in the endoderm of other chordates. To this aim, we investigated the expression of *FoxQ2III* in ammocoete larvae (40 days old) of the lamprey *P. marinus* and late embryos (S29) of the little skate *L. erinacea* (Fig. 4B, C). In situ HCR on whole samples showed that *FoxQ2III* is indeed expressed in specific locations within the developing digestive system of basally-branching vertebrates: lamprey larvae had widespread expression across the middle and posterior portion of the gut tube, concentrated in scattered, strongly labeled cells (Fig. 4B, Supplementary Fig. 10). In skates, strong expression was found in the midgut, between the esophagus and the intestine, while more scattered *FoxQ2III*-positive cells are present at the base of the yolk stalk (Fig. 4C, Supplementary Fig. 10). Overall, these results strongly indicate a conserved midgut expression of *FoxQ2III* across chordates, defining a previously undescribed domain of *FoxQ2* family expression.

In order to characterize the molecular identity of *FoxQ2III*-expressing cells, we integrated amphioxus and vertebrate scRNA data. We first used SAMap⁷¹ to co-embed endodermal subsets of cells from *B. floridae* T1 larvae⁶⁹, 3–120 hpf zebrafish larvae⁷², and mouse E8.5–9 somite stage larvae⁷³. Amphioxus *FoxQ2III*-positive clusters – annotated as midgut 5, 8, and 67 – had the highest mapping scores (fraction of cross-species mutual nearest neighbors) to zebrafish hepatocytes, particularly cluster 5 (mapping

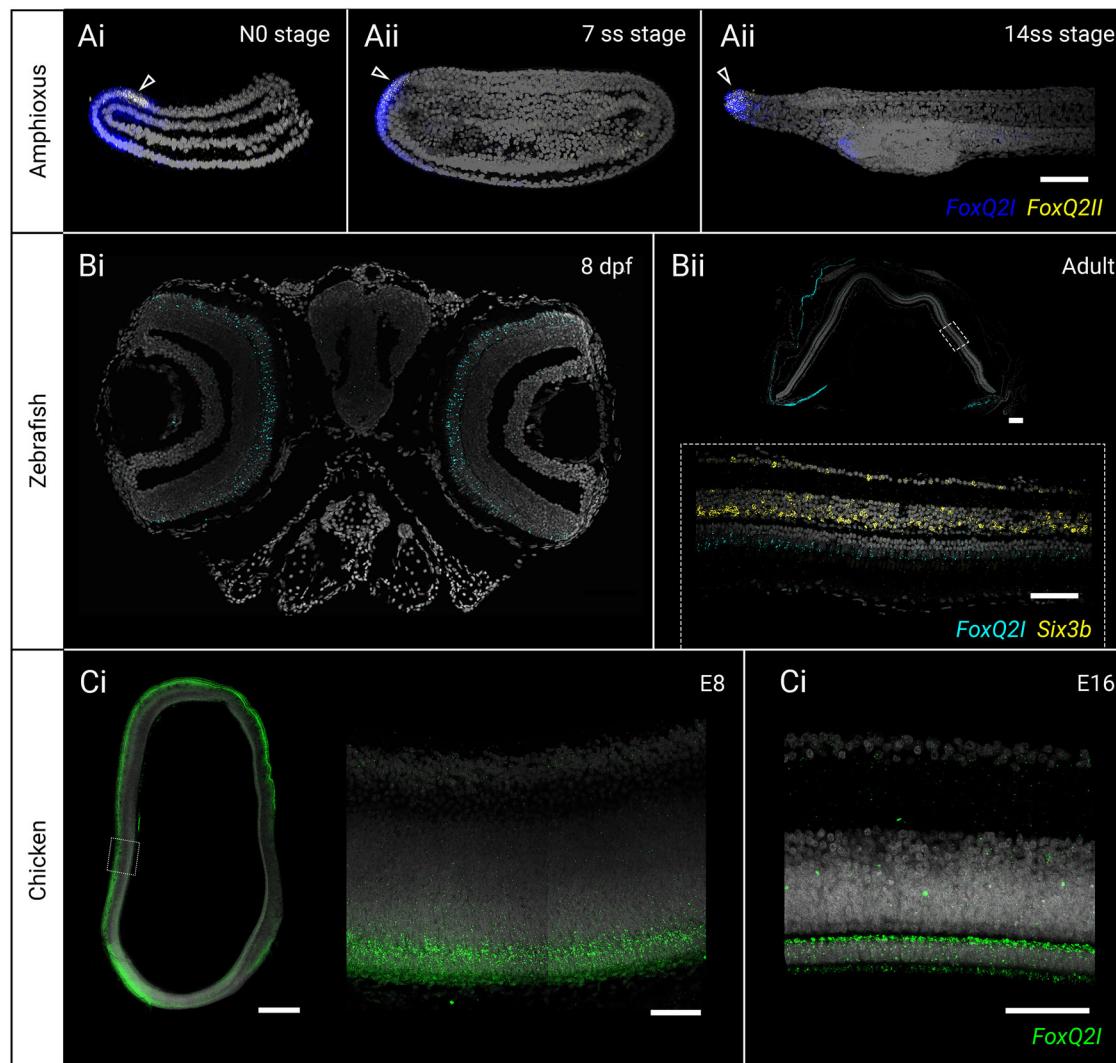


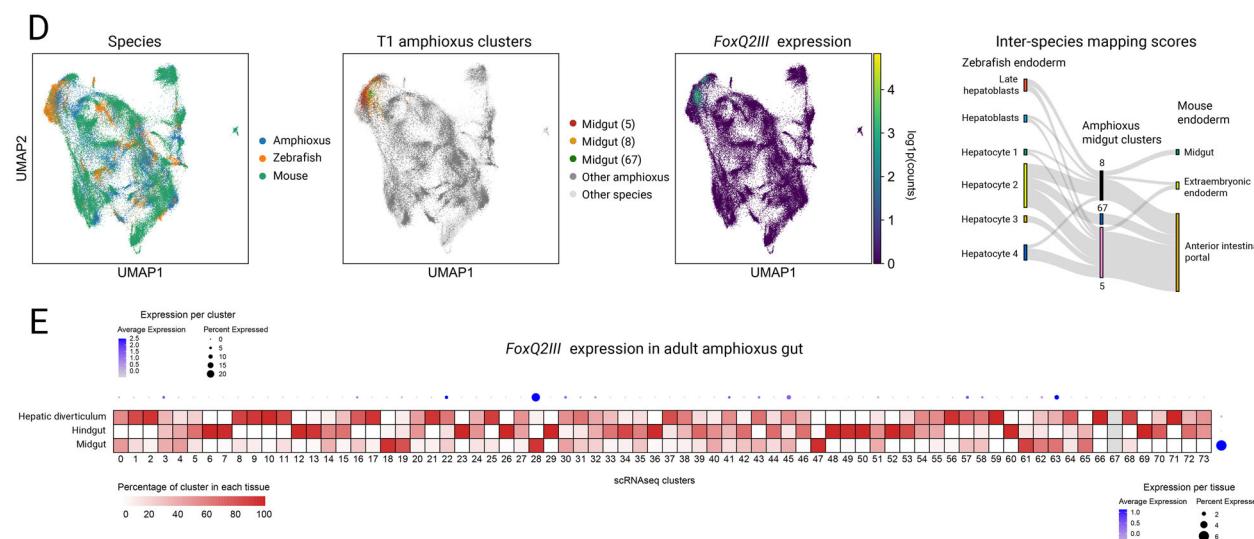
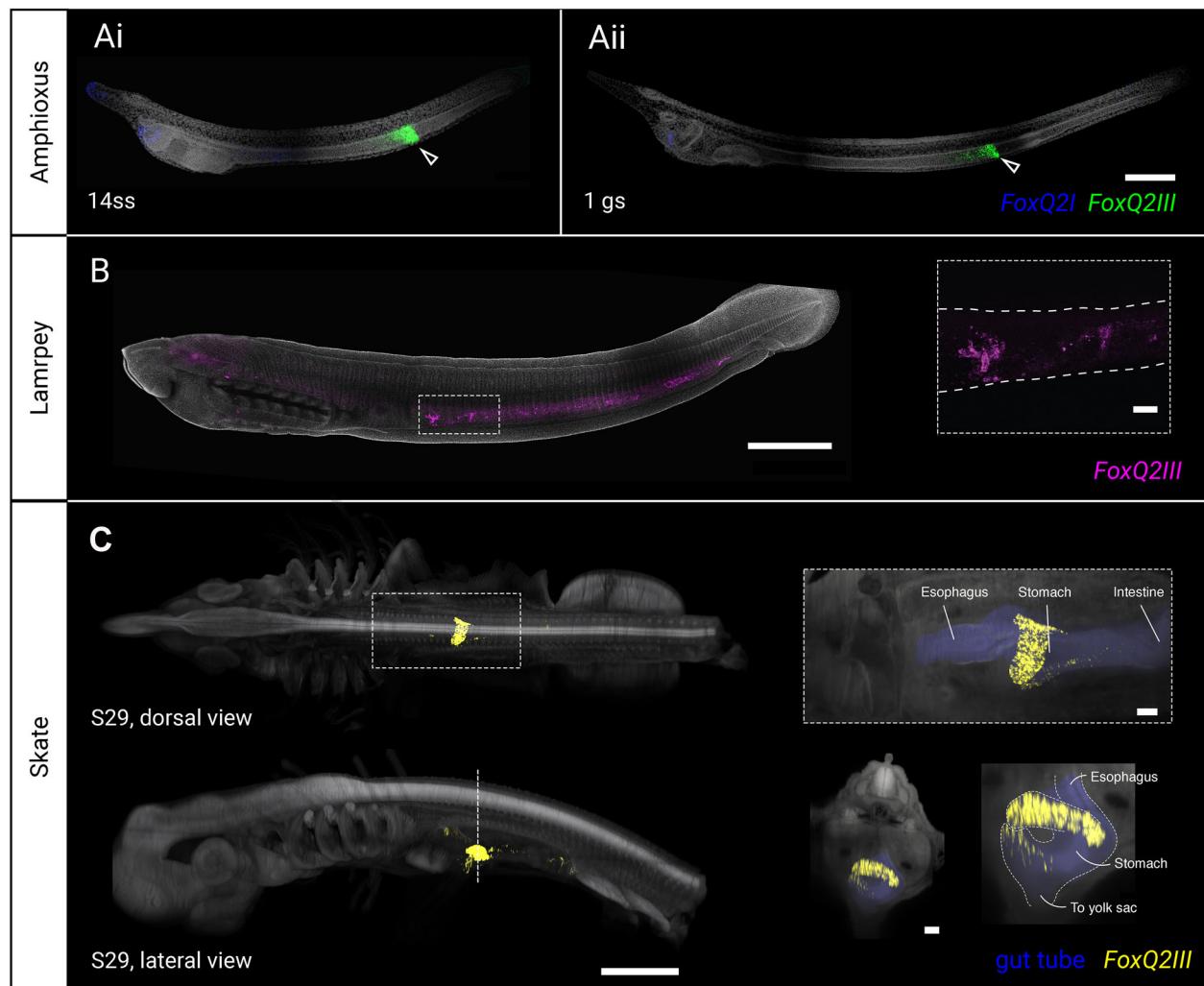
Fig. 3 | Expression pattern of *FoxQ2I*- and *FoxQ2II*-type paralogs in chordates by *in situ* hybridization chain reaction. A Co-localization of *FoxQ2I* (blue) and *FoxQ2II* (yellow) in the anterior ectoderm of amphioxus whole-mount embryos spanning the key stages of neural tube development: N0—early neurula (Ai); 7ss—mid neurula (Aii); 14ss—early larva (Aiii). Arrowhead indicates the area of co-expression of the two *FoxQ2* paralogs. B Expression of *FoxQ2I* (cyan) in the retina photoreceptors of 8-day larvae (Bi) and adult (Bii) zebrafish, detected on paraffin sections of the larval head and adult eye, respectively. Dashed square in Bii highlights

the magnified section of the adult retina showing differential expression of *FoxQ2I* in the photoreceptor layer and *Six3b* (yellow) in nuclear and ganglionic layers. C Distribution of *FoxQ2I* transcripts (green) in vibratome sections of the developing chicken eye at early (E8—embryonic day 8, Ci) and late (E16—embryonic day 16, Cii) stages, showing expression in the developing and mature photoreceptor layers of the retina. Dotted box in Ci indicates the position of the magnified section of the retina at E8 stage. Scalebars are 50 μ m for (Ai-iii), (Bi, Cii-iii), 100 μ m for (Bii), 500 μ m for (Ci).

score 0.48; Fig. 4D; Supplementary Fig. 8B–D). The genes that most strongly contribute to this similarity included amphioxus ficolins (amphioxus bf1620-1396 and 1397), a plasminogen (bf160-233), and a gene with partial similarity to HGF and SERPINB proteins (bf7-240) (Supplementary Fig. 8E). With respect to mouse, amphioxus *FoxQ2III*-positive clusters mapped best to the anterior intestinal portal, which gives rise to the foregut, driven by similar markers (*Fgb*, *Fgg*, *F10*, *Serpina1e*, *Serpinf2*, *Serpina1b*, *Serpina1a*) and other liver-expressed genes (*Bhmt* and *Foxa3*). Vertebrate genes matched by sequence similarity and expression correlation included zebrafish fibrinogens (*fga*, *fgb*, *fgg*), coagulation factors (*f2*, *f7*, *f7i*, *f9b*, *f2*) and SERPIN genes (*serpina1l*, *serpinc1*) involved in blood clot regulation, which are liver-specific in humans (Supplementary Data 3). Considering only the amphioxus data, the strongest marker of *FoxQ2III*-expressing clusters was an additional plasminogen-like gene (bf160-234), and the adjacent plasminogen (bf160-233) even more closely matched *FoxQ2III* expression (Supplementary Fig. 8E). *FoxQ2III* also overlapped expression domains of transcription factors that drive vertebrate liver development, though these are not exclusively expressed in the liver (*HNF4*, *HNF1B*, *Gata4*, *Hhex*).

(Supplementary Fig. 8F). Notably, the hepatic diverticulum has not yet developed by T1, and *in situ* probes definitively localize *FoxQ2III* to a more posterior position at this stage.

In the adult amphioxus gut, scRNA-seq⁷⁰ also indicates *FoxQ2III* is expressed in the midgut, while it is only sparsely detected in the hepatic diverticulum (Fig. 4E). Compared to T1 larvae, *FoxQ2III* in adults is detected in a lower proportion of cells in the positive clusters (clusters 28 and 63, annotated as gut epithelium, hepatic/midgut-enriched), and none of the above liver-related genes were detected. We performed another multispecies integration using the endoderm samples from the same zebrafish data and the mouse Tabula Muris dataset⁷⁴, and found that cells in cluster 28 matched best to zebrafish enterocytes (mapping score 0.24) and mouse large intestine (mapping score 0.21), with marker gene pairs corresponding to the intestine and colon (zebrafish *ace2*, *cdx1*, *slc6a19*, *slc15a1*, *clca1*; mouse *Tspan8*, *Tmem45b*). Cluster 63 was a stronger match to mouse large intestine only (mapping score 0.37) (Supplementary Fig. 9A–C, Supplementary Data 3). Thus, in adult amphioxus, *FoxQ2III* appears to be expressed in the gut epithelium.



Finally, to check whether *FoxQ2III* is expressed in the endoderm beyond chordates, we also looked at the expression of the *FoxQ2III*-type gene in the scRNAseq of *D. polymorpha*⁶⁰: at the stage analyzed we found only three positive cells, that were nonetheless located in the endoderm (Supplementary Fig. 6A). This could indicate that the gene is not expressed in the bivalve endoderm, or that it turns on at later

developmental stages. In support of the second hypothesis, a previous RNA-seq analysis of adult *C. gigas* showed that *FoxQ2III*-type in this bivalve mollusk can be detected in the digestive gland⁴⁰, suggesting a conserved expression within endodermally-derived tissues. However, spatial analysis of expression in more mollusks and brachiopods is needed to test this hypothesis.

Fig. 4 | Expression of *FoxQ2III*-type genes in chordates. A Co-detection of *FoxQ2I* (blue) and *FoxQ2III* (green) in amphioxus at early larva (14ss, A_i) and feeding 1 gill slit larva (1gs, A_{ii}) stages. In the amphioxus larva *FoxQ2III* expression appears in a restricted portion of the midgut endoderm. B Localization of *FoxQ2III* (magenta) in ammocoete larvae of the lamprey, showing expression in the gut (magnification in dashed box). C Expression of *FoxQ2III* (yellow) in skate embryos at stage 29, showing restricted expression in a diverticulum of the digestive tube. D UMAP embeddings of SAMap multispecies integration

with cells colored by species (left), amphioxus *FoxQ2III*-expressing clusters (middle), and *FoxQ2III* expression (right). The Sankey diagram on the right depicts the three *FoxQ2III*-clusters and all zebrafish or mouse cell groups with a mapping score of at least 0.05. Lines are scaled to the maximum sum of mapping scores in either vertebrate, so that their weights represent relative similarly for each cell type pair. All pairwise, unscaled mapping scores are shown in Supplementary Fig. 8B, C.

E Heatmap visualizing the tissue composition of all clusters in adult amphioxus gut scRNA-seq (Dai et al.⁶⁹). *FoxQIII* expression is visualized on the margins, with dots scaled to the fraction of expressing cells per group and colored by mean expression. Scalebars are 100 μ m for A and magnification insets in (B, C), and 1 mm for (B, C).

Conserved candidate regulatory sequences across deuterostomes

Both *FoxQ2I* and *FoxQ2II* have been shown to be involved in the determination of anterior identity during embryonic development. Loss of *FoxQ2I* results in the loss of apical organ neurons in echinoderm larvae, of blue photoreceptors in zebrafish, and of anterior neural identity in the amphioxus cerebral vesicle, while loss of *FoxQ2II* in arthropod embryos causes defects in anterior brain development and labrum formation^{22,25–27,29,31,69}. Several studies have also shown that *FoxQ2I* and *FoxQ2II* genes are part of a highly conserved aGRN that is involved in the specification of anterior fate^{16,32,35,75}. However, little is known about the mechanisms that control *FoxQ2* expression. Functional studies and the analysis of cis-regulatory regions in echinoderms showed that *Meis* and *BicC* regulate *FoxQ2* maintenance, and *Six3/6* is required for *FoxQ2I* expression in sea urchin but not starfish^{21,22,31,76–78}. Similarly, *Six3/6* promotes *FoxQ2II* expression in arthropods^{27,29}.

Here, we aimed to compile a comparative list of candidate factors that regulate *FoxQ2* genes in cephalochordates, given that they are one of the few taxa to possess all three *FoxQ2* paralogs (Fig. 5 shows the exemplified pipeline for *FoxQ2I*). To reconstruct conserved regulatory sequences across cephalochordates, we first identified *FoxQ2I*, *FoxQ2II* and *FoxQ2III* orthologs in five amphioxus genomes representing all three extant cephalochordate genera (*B. lanceolatum*, *B. floridae*, *B. belcheri*, *Asymmetron*, *Epigonichthys*). We then selected a region of ~5000 bp upstream of the start codon in all five species, and compared them for each gene using mVISTA⁷⁹ (Fig. 5A). This approach identified conserved non-coding sequences (CNCSs) upstream of *FoxQ2* paralogs across cephalochordates: three CNCSs for *FoxQ2I* (Fig. 5A), one for *FoxQ2II* and one for *FoxQ2III* (Supplementary Fig. 11, Supplementary Data 3). In parallel, we used a published ATACseq datasets of *B. lanceolatum* at four stages of development, spanning early gastrula to larva stages (8hpf, 15hpf, 36hpf, 60hpf)⁶⁵, to identify open chromatin sequences. Strikingly, a majority of ATACseq peaks corresponded precisely to CNCS (Supplementary Fig. 11). These two lines of evidence supported their identification as conserved regulatory sequences. We therefore used CiiIDER⁸⁰ to predict transcription factor binding sites (TFBSs) within each CNCS for all three paralogs across species. To select only candidate conserved TFBSs, we normalized the length of each CNCS across species, and then selected only TFBSs that could be identified in all five species and located in similar positions (10% margin in each direction) (Fig. 5B, Supplementary Data 3).

This analysis resulted in a list of candidate *FoxQ2* TFBSs in cephalochordates (Supplementary Data 3). We next refined this selection further by interrogating a published RNAseq datasets of *B. lanceolatum* development⁶⁵ and a recently published scRNAseq dataset of *B. floridae* development⁶⁹ (see “Materials and Methods” for details). This allowed us to select only candidate TFBSs for transcription factors that are active around the developmental time when each paralog is expressed, and in the correct cell type and embryonic location (Fig. 5C, Supplementary Data 3). In *FoxQ2I* CNCSs we found, among others, multiple candidate TFBSs for transcription factors of the *Meis*, *Pou6* and *SoxB* families, which were previously found in regulatory sequences of sea urchin and which are expressed maternally in both echinoids and amphioxus⁷⁶ (Fig. 5B–D). As *Meis* is known to regulate *FoxQ2I* maintenance in sea urchin, these results suggest a possible conserved role of *Meis* in controlling *FoxQ2I* expression

across deuterostomes. Putative TFBSs for *Meis* but not for *Pou6* and *SoxB* genes were found in the *FoxQ2II* CNCS, possibly suggesting regulatory differences that might underlie the difference in activation timing (Supplementary Data 3). Strikingly, the *FoxQ2III* CNCSs possessed TFBSs for endodermal markers, such as *FoxA* and *Pdx*. Overall, this analysis provides a framework for the identification of putative conserved and cell-type specific TFBSs, which can then be tested functionally, and suggests that *FoxQ2I* activation and maintenance in amphioxus might be under the control of early maternal signals that are conserved across deuterostomes.

Discussion

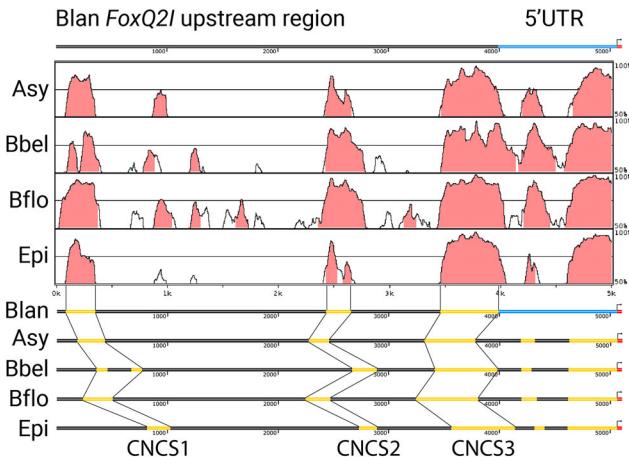
Evolution of the three *FoxQ2* genes across metazoans

FoxQ2 is one of the most conserved classes of transcription factors in metazoans and has a widespread role in the specification of anterior ectodermal identity^{28,32,33,35}. At the same time, the *FoxQ2* family has undergone a dynamic evolutionary history, characterized by fast evolutionary rates and a high number of lineage-specific duplications and losses^{11,12}. Here, we specifically addressed this apparent discrepancy between conservation and divergence by analyzing the evolution of *FoxQ2* genes across all major metazoan lineages.

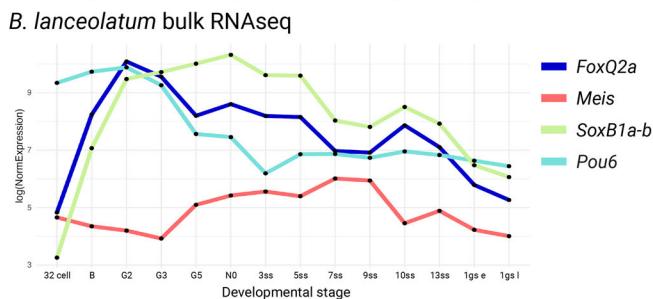
Our phylogenetic and synteny analyses, which included sequences from 21 animal phyla, revealed that the *FoxQ2* class originated in the common ancestor of metazoans, and identified three ancient *FoxQ2* paralogs. To avoid confusion with previous nomenclatures in different species, we named these three paralogs *FoxQ2I*, *FoxQ2II* and *FoxQ2III* and proposed to rename genes in each species accordingly (Supplementary Data 1). For example, as the amphioxus *FoxQ2II*-type has been previously annotated as *FoxQ2c* and *FoxQ2III*-type was annotated as *FoxQ2b*, we renamed both genes. For those species in which multiple copies of a paralog are present, the *FoxQ2* type would be followed by a number. For example, *Saccoglossus kowalevskii* has two *FoxQ2I*-type copies and one *FoxQ2II* copy originally called *FoxQ2-1*, *FoxQ2-2* and *FoxQ2-3* respectively. In the new nomenclature, these would be named *FoxQ2I-1*, *FoxQ2I-2* and *FoxQ2II*. Species abbreviation could additionally be added in front of the gene name to help distinguish between lineage-specific duplications.

The evolutionary history of *FoxQ2* genes is summarized in Fig. 6. Through a survey of macro-synteny patterns across extant metazoan genomes, we found that *FoxQ2I*- and *FoxQ2II*-type genes are often located in chromosomes derived from ALG_C1, while *FoxQ2III*-type genes are mostly located on chromosomes corresponding to ALG_E/Eb. This trend is more apparent among bilaterian and cnidarian genomes, suggesting the existence of these two ancient *FoxQ2* paralogs in the common ancestor of these animals. In addition, the presence of poriferan *FoxQ2I*-type and *FoxQ2III*-type genes in ALG_C1-derived chromosome also provides evidence to support the idea that in the common ancestor of poriferans, cnidarians and bilaterians, the original *FoxQ2* gene was likely located in ALG_C1. The initial gene duplication event generating the *FoxQ2III* paralog might have occurred in ALG_C1 as well. In contrast, the basal branching position of ctenophore *FoxQ2* sequence in the phylogenetic tree and its unique genomic position within the ALG_L/M group (Figs. 1, 2B) make it difficult to infer the ancestral position of *FoxQ2* gene in the common ancestor of all metazoans when we consider the ctenophore-sister hypothesis^{55,81,82}. If we consider the alternative sponge-sister hypothesis^{83,84}, the macro-synteny patterns would then suggest that the ancestral location of *FoxQ2* gene in the

A Identification of CNCSs upstream of *FoxQ2l* in five amphioxus species combining mVISTA and ATACseq



C Selection of stage- and tissue-compatible TFBSs using bulk and single cell RNA sequencing



B. floridae B stage scRNASeq

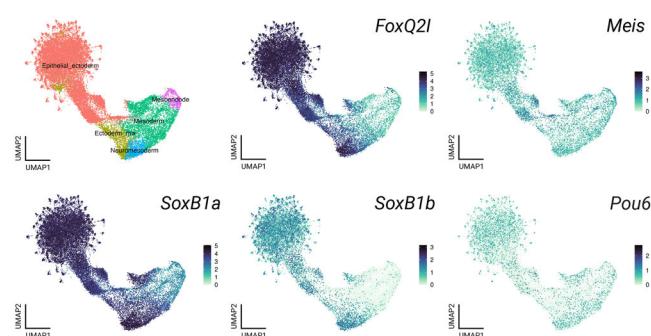
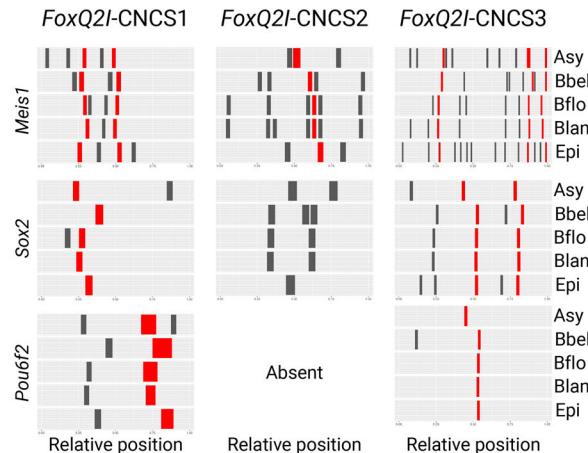


Fig. 5 | Prediction of conserved *FoxQ2* transcription factor binding sites with spatial and temporal resolution in amphioxus. Schematic workflow for the identification of transcription factor binding sites (TFBSs) for amphioxus *FoxQ2I*. **A** Graphical representation of comparisons of the 5Kb upstream of *FoxQ2I* between *Branchiostoma lanceolatum* (Blan) and four other amphioxus species: *Asymmetron* (Asy); *Branchiostoma belcheri* (Bbel); *Branchiostoma floridae* (Bflo); *Epigonichthys* (Epi) using mVISTA (top), and schematic map of the position of the conserved non-coding sequences (CNCs) in each species highlighted in yellow (bottom). mVISTA graph for Blan comparison is shown as a representative, the entire list of comparisons can be found in Supplementary Fig. 11. **B** Barplot showing the position of three

common ancestor of all metazoans was in ALG_C1, and that ctenophore *FoxQ2* was translocated to other ALGs and subsequently underwent drastic sequence divergence. Regardless these two competing phylogenetic hypotheses on whether sponges or ctenophores represent the sister group to all other animals, our results firmly support the existence of two distinctive groups of *FoxQ2* genes, namely the *FoxQ2I*/*FoxQ2II* and *FoxQ2III*, which

B Discovery of candidate conserved TFBSs for *FoxQ2L* across cephalochordates



D Final list of candidate TFBSS binding to conserved regulatory sequences of *FoxQ2l*

TF Class	TF Family	Predicted TFs	CNCSS	Shared with sea urchin <i>FoxQ2</i>
bHLH	MyoD-ASC-related	<i>Ascl1</i>	CNC53	-
	PAS	<i>Hif1α, Npas2</i>	CNC51, CNC53	-
	Tal-related	<i>Fif1a</i>	CNC53	-
bZIP	CEBP-related	<i>Hlf</i>	CNC51, CNC52, CNC53	+
	Jun-related	<i>Jun</i>	CNC51, CNC53	-
C2H2 zinc finger	3ZF Kruppel-related	<i>Klf4, Klf5, Klf6, Klf11, Klf12, Klf15, Sp4</i>	CNC51, CNC53	-
	multiple zf	<i>Prdm4</i>	CNC53	+
C4 zinc finger	C4-GATA	<i>Gata2</i>	CNC51, CNC53	-
Forkhead	E2F	<i>E2f6</i>	CNC51, CNC53	-
	FOX	<i>FoxK1, FoxO6</i>	CNC51, CNC53	+
HMG	SOX	<i>Sox2, Sox4, Sox6, Sox9, Sox14</i>	CNC51, CNC53	+
Homeodomain	HOX	<i>Evx1</i>	CNC51, CNC53	-
	LIM	<i>Lhx8</i>	CNC51, CNC53	-
	NK	<i>Barhl2, Dlx1, Msx2</i>	CNC51, CNC53	+
	Paired-related	<i>Drgx, Gsc2, Otx2, Pitx3, Vsx1</i>	CNC51, CNC53	-
	POU	<i>Pou6f2</i>	CNC51, CNC53	+
	TALE-type	<i>Meis1</i>	CNC51, CNC52, CNC53	+
RHR	NFAT-related	<i>Nfat5</i>	CNC52, CNC53	-
SAND	SAND-domain	<i>Gmeb1</i>	CNC51, CNC53	-
SMAD/NF-1	NF-1	<i>Nfix</i>	CNC52, CNC53	-
T-box	T-box	<i>Eomes, Mga, Tbx2, Tbx3, Tbx6</i>	CNC51, CNC52, CNC53	-

predicted TFBSS (*Meis*, *SoxB* and *Pou6*) along the three CNCSs (x axis) found in the five amphioxus species (y axis). Sites that are present in all five species in the same position are marked in red. **C** Developmental expression by bulk RNAseq allows to select among the list of conserved TFBSS those for transcription factors that are active when *FoxQ2* is expressed; further filtering with scRNAseq at the blastula stage, when *FoxQ2I* is first activated, retains only candidate TFBSS with the correct spatiotemporal resolution. **D** Filtered list of conserved candidate *FoxQ2I* TFBSS in cephalochordates, highlighting the TF class and family, the presence in each CNCS and whether candidate TFBSS are shared with sea urchin *FoxQ2I* regulatory regions.

can be traced back to the common ancestor of cnidarians and bilaterians. Following previous interpretations^{11–13}, we hypothesize that the ancestral *FoxQ2I*/*FoxQ2II* paralog further underwent additional duplication at the base of Parahoxozoa to give rise to *FoxQ2I* and *FoxQ2II*. These paralogs diverged extensively in the cnidarian and bilaterian lineages and their evolutionary relationships are therefore difficult to trace. Two less

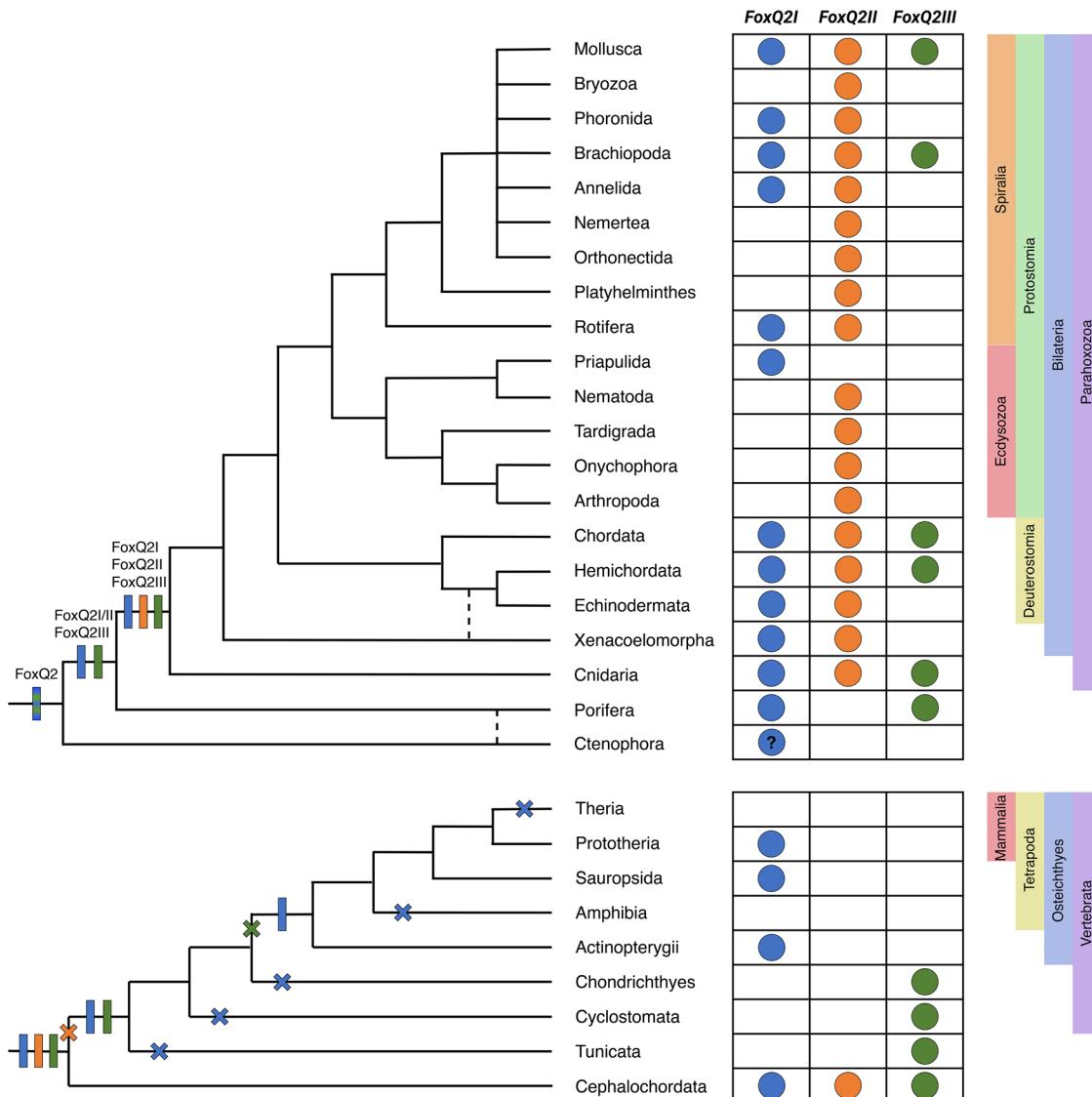


Fig. 6 | Schematic summary of the evolutionary history of *FoxQ2* genes in the main metazoan taxa and in chordates. The appearance of *FoxQ2I* (blue) *FoxQ2II* (orange) and *FoxQ2III* (green) genes during metazoan evolution is indicated by bars,

and the *FoxQ2* repertoire for each metazoan phylum and chordate taxa is represented next to the phylogenetic trees. Subsequent losses of *FoxQ2* genes are detailed for chordates with an X.

parsimonious alternative scenarios imply that only one *FoxQ2* gene was present in the ancestor of Parahoxozoa, and it independently duplicated in cnidarians and bilaterians, or that both paralogs were already present in the metazoan ancestor and *FoxQ2II* was lost in sponges and ctenophores. *FoxQ2I* and *FoxQ2II* paralogs were maintained in bilaterians, although subsequent independent events of gene loss resulted in different combinations of *FoxQ2* genes in each phylum (Fig. 6). In spirilians, both paralogs were still present in most early lineages, but then underwent several taxon-specific expansions and reductions¹. In contrast, in ecdysozoans, *FoxQ2I* orthologs were lost in most lineages, leaving *FoxQ2II* as the only remaining member in arthropods, onychophorans, tardigrades, and nematodes. In deuterostomes, *FoxQ2I* was retained in all three lineages, while *FoxQ2II* was lost in eleutherocoan echinoderms and vertebrates, and both genes were lost in tunicates.

The *FoxQ2III* paralog identified here was lost in several lineages, including all ecdysozoans, the lineages leading to platyhelminths and rotifers within spirilians, and echinoderms among deuterostomes. However, it is still present in lophotrochozoans, hemichordates, and chordates. As such, cephalochordates, enteropneust hemichordates, bivalve mollusks and brachiopods are the only taxa in which all three paralogs have been identified in

the same species to date. Focusing on chordates, the presence of *FoxQ2I*, *FoxQ2II* and *FoxQ2III* genes in amphioxus and of *FoxQ2I* and *FoxQ2III* genes in different vertebrate lineages, together with the conserved gene synteny of chordate *FoxQ2III*, indicates that vertebrates ancestrally possessed two *FoxQ2* genes. Cyclostomes and cartilaginous fishes then independently lost *FoxQ2I*, while bony fishes lost *FoxQ2III*. This, in turn, suggests that the vertebrate *FoxQ2* repertoire is richer than previously estimated and encourages further research on *FoxQ2* genes in this group.

Expression and function of *FoxQ2I* and *FoxQ2II* in bilaterians

In contrast with their ancient origin and high evolutionary divergence, *FoxQ2I* and *FoxQ2II* expression remained exceptionally conserved across Parahoxozoa. Orthologs of both genes are expressed in and pattern the aboral ectoderm of cnidarians and the anterior ectoderm of bilaterians, with both neural and non-neural derivatives. Structures of the anterior neuroectoderm expressing *FoxQ2I* or *FoxQ2II* during development include: the apical organs present in the larvae of at least 8 phyla^{14,16,17,19,39,46,47,57,85}; the anterior portions of arthropod and onychoporan brains^{18,27,28,43}, planarian cerebral ganglia⁴², amphioxus early neural plate³², and hemichordate nerve plexi¹³; and photoreceptor cells including the eyes of flatworms and

chordates^{39,49,58}. Anterior non-neuroectodermal derivatives have also been described, such as the anterior epidermis and rostrum of amphioxus, the labrum of arthropods and the apical tuft cells of ciliated larvae^{27,29,32,86}. Interestingly, the expression dynamics of *FoxQ2I* and *FoxQ2II* genes in the anterior ectoderm appear to differ while remaining conserved across phyla. Indeed, *FoxQ2I* is generally expressed from very early stages, often at the beginning of zygotic transcription, in a broad animal/anterior domain, and then restricts towards the animal tip of the embryo. This pattern appears particularly conserved in deuterostomes, for which developmental data has been obtained for crinoid, echinoid and asteroid echinoderms, enteropneust hemichordates and cephalochordates^{23,24,31,32,46,57}. After restriction, *FoxQ2I* remains expressed in many marine planktonic larvae and contributes to the specification of apical organ or apical tuft cells^{22,23,26,57,75}. Curiously, species that have lost their marine larva stage (vertebrate chordates, arthropods, panpulmonate mollusks, clitellate annelids) also appear to have lost either the early expression of *FoxQ2I* or the paralog altogether. There are however exceptions, such as polyclad flatworms, which have lost *FoxQ2I* but maintain a marine larva stage possessing an apical organ that expresses *FoxQ2II*⁸⁵. For species in which we have expression data of both *FoxQ2I* and *FoxQ2II*, such as annelids, hemichordates and cephalochordates, *FoxQ2II* appears after *FoxQ2I*, and generally starts to be expressed within the *FoxQ2I* domain^{11,13}.

The conservation in expression dynamics of *FoxQ2I* and *FoxQ2II* genes is reflected in their similar function within the aGRN that controls the specification of anterior neural identities. In fact, *FoxQ2I* mediates the formation of apical organ neurons and apical tuft cells in echinoderms, as well as the anterior brain in cephalochordates^{22,32,49}. Notably, a recent paper showed that *FoxQ2I* knock-out in amphioxus leads to the complete loss of anterior retina- and hypothalamus-like regions of the larval brain (cerebral vesicle)⁶⁹, supporting our previous hypothesis on the presence of a conserved anterior brain region, which forms in a Wnt-free area of the neuroectoderm through a conserved aGRN³². Strikingly, loss of *FoxQ2I* in zebrafish leads to the disappearance of blue cones from the retina, which develops from the embryonic secondary prosencephalon (the portion of the central nervous system that includes telencephalon, hypothalamus and retina)⁴⁹. The similar expression of *FoxQ2I* in chicken embryos found here suggests that this is a conserved mechanism in multiple bony fish lineages. Another line of evidence comes from the fact that *FoxQ2I* was lost in placental mammals, concomitantly with the loss of blue photoreceptor types. However, the fact that amphibians, cyclostomes and cartilaginous fishes have lost *FoxQ2I* while still maintaining blue cones raises an intriguing hypothesis: that additional and partially redundant mechanisms may regulate blue cone formation in different vertebrate lineages.

Similar to *FoxQ2I*, the single arthropod *FoxQ2II* has been shown to control labrum and anterior brain formation, including central complex²⁷⁻²⁹. Moreover, both paralogs appear to be negatively regulated by Wnt signaling, which originates from the vegetal pole and subsequently from the posterior side of the embryo during the development of most bilaterians. Wnt overactivation has been shown to downregulate expression of *FoxQ2I* in echinoderms, hemichordates and chordates, and of *FoxQ2II* in annelids and arthropods^{16,23,28,32,75}. This downregulation is followed by the loss of aGRN markers and by severe defects in the formation of anterior neuroectodermal structures, including larval apical organs in annelids and the anterior cerebral vesicle in cephalochordates. Strikingly, in cephalochordates, Wnt overactivation leads to a loss of both *FoxQ2I* and *FoxQ2II* (Supplementary Fig. 7E). These results suggest a similar and possibly redundant function for these two paralogs within the aGRN of bilaterians. However, it also indicates that functional considerations across bilaterians should only be made after careful examination, as the two genes have an ancient origin and have undergone independent evolution. This is further shown by the differences in the predicted TFBSS found in the proximal cis-regulatory regions of *FoxQ2I* and *FoxQ2II* across cephalochordates. By leveraging the availability of five amphioxus genomes from all three extant genera, as well as the genomics and transcriptomics resources built by the amphioxus community in the last decade, we devised a method for the predicting candidate

conserved TFBSS with both developmental timing- and cell type-specificity. This resource reveals that while *FoxQ2I* regulatory regions contain candidate TFBSS that appear conserved with echinoderms⁷⁶, *FoxQ2II* is regulated by partially different transcription factors. Future analyses aimed at functionally testing the activity of these putative TFBSS could provide insights into the differences in the temporal activation of these two anterior *FoxQ2* paralogs in amphioxus.

Discovery of a *FoxQ2* paralog group active in the endoderm

Finally, here we report the first expression pattern of the third *FoxQ2* paralog, *FoxQ2III*, in multiple chordate species. While the localization of the other two *FoxQ2* genes has always been associated with the anterior ectoderm, we find that this gene family is expressed in endodermal tissues during development. scRNAseq and *in situ* HCR in amphioxus, lamprey and skate show that *FoxQ2III* transcripts can be detected in the gut during late development. The timing of their activation suggests that this gene is not involved in early endodermal specification, but more likely in the differentiation of specific midgut cell types. In amphioxus, for example, *FoxQ2III* is activated at early larval stages, prior to the opening of the mouth and the start of feeding, in cells that express effector genes and secreted molecules similar to those in vertebrate hepatocytes. Based on their transcriptomic identity, we speculate that larval *FoxQ2III*-positive midgut cells in amphioxus might play a multifaceted role related to coagulation and innate immune response, similar to those described for vertebrate hepatocytes. Our comparative results also suggest that chordates ancestrally possessed *FoxQ2III*-positive cell types in the midgut. Similar to the expression of *FoxQ2I* in the retina of only selected vertebrate lineages, the presence of *FoxQ2III* in the gut of amphioxus, lampreys and skates but not of other vertebrate lineages raises intriguing questions on the evolution of the endoderm. Were these cell-types lost in bony fishes? Or did their GRN modified extensively and lost *FoxQ2III* expression while maintaining the differentiated cell identity? The increasing availability of scRNAseq datasets from species across the chordate lineage will be critical to solve these questions and will help us gain a deeper understanding on the evolution of cell types.

Conclusions

We have characterized the evolution of *FoxQ2* genes and identified three distinct paralogs. Two of these, *FoxQ2I* and *FoxQ2III*, are shared between poriferans, cnidarians and bilaterians, suggesting a more complex repertoire of forkhead genes in the metazoan ancestor than previously thought. Despite their high similarity in protein structure, these two ancient paralogs are expressed in distinct embryonic domains in bilaterians, i.e., the anterior ectoderm and the digestive endoderm respectively.

The third paralog *FoxQ2II* likely duplicated from an ancestral *FoxQ2I/II* gene in the cnidarian-bilaterian ancestor, and *FoxQ2I* and *FoxQ2II* genes are generally expressed in a similar domain in modern bilaterians. How does the fast molecular evolution of *FoxQ2I* and *FoxQ2II*, highlighted by phylogenetic analysis, reconcile with the high conservation in their expression pattern? We propose that the ancient duplication of anterior genes *FoxQ2I* and *FoxQ2II*, which likely had a redundant function, might have provided an ideal background for subfunctionalization or specialization⁶⁵, so that new copies could be duplicated and others lost without large consequences for the organism. To test this hypothesis, it would be interesting to direct future studies at analyzing the function and possible redundancy of *FoxQ2I* and *FoxQ2II* paralogs in species that still possess both, such as annelids, mollusks, hemichordates or cephalochordates.

Methods

Phylogenetic analysis

Information on all sequences used in this study, including original gene name, new proposed name, species, taxon, paralog type, completeness, protein sequence, reference genome and reference code (when available), is provided in Supplementary Data 1. We first selected sequences from 17 phyla already available in the literature; incomplete sequences for *Gallus gallus* *FoxQ2* and *Branchiostoma lanceolatum* *FoxQ2III* were manually

expanded from the genome. We then used reciprocal BLAST best hits (default parameters) to recover *FoxQ2* sequences from the genomes of *Petromyzon marinus*, *Leucoraja erinacea*, *Pleurodeles waltli* (Chordata), *Schizocardium californicum* (Hemichordata), *Candidula unifasciata*, *Mercenaria mercenaria* (Mollusca), *Lombricus rubellus*, *Hirudo medicinalis* (Annelida), *Membranipora membranacea* (Bryozoa), *Adineta vaga* (Rotifera), *Hypsibius exemplaris* (Tardigrada), *Priapulis caudatus* (Priapulida), *Hofstenia miamia*, *Symsagittifera roscoffensis*, *Xenoturbella bocki* (Xenacoelomorpha), *Hydractinia symbiolongicarpus*, *Hydra vulgaris*, *Acropora millepora*, *Pocillopora verrucosa*, *Rhopilema esculentum* (Cnidaria), *Dysidea avara*, *Ephydatia muelleri*, *Halichondria panicea*, *Corticium candelabrum*, *Oscarella lobularis* (Porifera), *Bolinopsis microptera*, *Horniphora californensis* (Ctenophora). For all sequences used in the analysis, membership to *FoxQ2* family was assessed by confirming the presence of a *FoxQ2* domain using NCBI conserved domain search based on Conserved Domain Database (CDD) v3.21⁸⁷. Following this confirmation, two groups of sequences were used for phylogenetic analysis:

- Group 1: complete *FoxQ2* sequences from 33 species belonging to 17 phyla,
- Group 2: sequences in which the *FoxQ2* domain was specifically isolated from 47 species belonging to 21 phyla.

For both groups, amino acid sequences were aligned using MAFFT v.7.526 under default settings⁸⁸. Phylogenetic analysis was then performed for both groups with two methods: Maximum Likelihood, using IQ-TREE web server with default parameters⁸⁹, and Neighbor Joining using Seaview⁹⁰. A thousand ultrafast bootstraps were used to extract branch support values with each method. The resulting trees were then visualized with FigTree v.1.4.4. Uniform manifold approximation and projection (UMAP) dimensionality reduction and visualization was performed on the web version of Alignmentviewer (<https://alignmentviewer.org/>). Protein secondary structures were predicted using ColabFold⁹¹.

Microsynteny analysis was performed by manually comparing the genomic loci around *FoxQ2* genes in *Branchiostoma lanceolatum* (v.klBraLanc5.hap2, GCF_035083965.1), *Petromyzon marinus* (v.kPet-Mar1.pri, GCF_010993605.1)⁵⁰, *Leucoraja erinacea* (v.Leri_hhj_1, GCF_028641065.1)⁵¹ and *Danio rerio* (v.GRCz11, GCF_000002035.6) in NCBI Datasets.

Macro-synteny

Protein sequences and gene models were obtained from publicly available sources, with genome data details provided in Supplementary Data 2. Candidate *FoxQ2* genes were identified using OrthoFinder v2.5.4⁹², and their classification was based on the results of phylogenetic analysis. Pairwise macrosyntenic comparisons between species were conducted using MCscan (Python version) implemented in JCVI v1.2.7^{93,94}, as previously described^{95,96}. Briefly, Orthologous gene pairs between species were identified using the LAST aligner integrated in MCscan. A C-score threshold of 0.99 was applied to retrieve reciprocal best hits. For comparisons involving amphioxus and species that have undergone whole genome duplication (WGD), including lamprey, skate, and spotted gar, a relaxed C-score threshold of 0.7 was applied. Corresponding chromosome pairs were determined using Fisher's exact test with Bonferroni correction (adjusted $p < 0.05$). To enhance the sensitivity in detecting syntenic relationships between sponge and ctenophore, a more lenient adjusted p-value threshold of 0.2 was used. Gene pairs located outside the identified corresponding chromosome pairs, as well as those involved in a small-scale chromosomal rearrangement event in hemichordate⁹⁶, were excluded from macro-synteny visualizations.

Animal collection

Amphioxus. Adult *Branchiostoma lanceolatum* were collected in Banyuls-sur-Mer (France) and maintained in a custom-made facility at the Department of Zoology, University of Cambridge (UK). Spawning and fertilization was performed following⁹⁷, and embryos were raised in

Petri dishes in filtered artificial salt water at 21 °C. For embryos at 4 dpf, from the 48 h stage embryos were fed with a mix of algae. At the desired stage embryos were collected and fixed in ice-cold 3.7% Paraformaldehyde (PFA) + 3-(Nmorpholino) propane sulfonic acid (MOPS) buffer for 12 h, then washed in sodium phosphate buffer saline (NPBS), dehydrated and stored in 100% methanol at -20 °C.

Zebrafish. Zebrafish (*Danio rerio*) embryos were raised to 8 dpf and fixed overnight at 4 °C in 4% paraformaldehyde in phosphate-buffered saline (PBS), then rinsed in PBS, dehydrated into 100% methanol and stored at -20 °C. Experiments using larval and adult zebrafish were conducted according to protocols approved by the Institutional Animal Care and Use Committees in facilities accredited by the Association for Assessment and Accreditation of Laboratory Animal Care International (AAALAC). The use of tissue samples for this study was reviewed and ethically approved by the University of Cambridge Animal Welfare and Ethical Review Body (AWERB) Committee. We have complied with all relevant ethical regulations for animal use.

Chicken. Fertilized White Leghorn Chicken (*Gallus gallus*) eggs (Charles River) were incubated in a 38 °C humidified chamber with embryos staged according to ref. 98. At E8 and E16 stages, eyes were dissected from the embryo in cold PBS and immediately fixed overnight at 4 °C in 4% paraformaldehyde in PBS, then dehydrated to 100% methanol and stored at -20 °C.

Lamprey. Adult sea lamprey (*Petromyzon marinus*) were collected from the Hammond Bay Biological Station, Millersburg, MI, and shipped to Northwestern University. Embryos and larvae were obtained by in vitro fertilization, fixed for 2 h at room temperature in MEMFA at desired stages, and then dehydrated and stored in 100% methanol at -20 °C prior to analysis. All procedures were approved by Northwestern University's Institutional Animal Care and Use Committee (IACUC A3283-01), and we have complied with all relevant ethical regulations for animal use.

Skate. Little skate (*Leucoraja erinacea*) embryos were obtained from the Marine Resources Center at the Marine Biological Laboratory (MBL) in Woods Hole, MA, U.S.A., and reared to stage 29 as described in ref. 99. All skate experiments were conducted according to protocols approved by the Institutional Animal Care and Use Committee of the MBL. We have complied with all relevant ethical regulations for animal use. Skate embryos were euthanized with an overdose of MS-222 (1 g/L in sea-water), and all embryos were fixed overnight at 4 °C in 4% paraformaldehyde in phosphate-buffered saline PBS, and then rinsed in PBS and dehydrated into methanol prior to analysis.

In situ hybridization chain reaction

For *in situ* HCR v.3¹⁰⁰ on amphioxus, zebrafish, chicken and skate specimens, probes were ordered through Molecular Instruments Inc.

Amphioxus. Reactions were performed as described in ref. 101. Amphioxus embryos and larvae were rehydrated in NPBS + 0.1% Triton X, (NPBT), bleached for 30 min (5% Deionized formamide, 1.5% H2O2, 0.2% SSC in nuclease-free water) and permeabilized for 3 h (NPBS + 1% DMSO + 1% TritonX). The embryos were incubated in Hybridization Buffer (Molecular Instruments) for 2 h and then probes were added overnight at 37 °C. The following day, embryos were rinsed in Wash Buffer (Molecular Instrument) followed by 5X-SSC + 0.1% Triton X. The samples were then incubated in Amplification Buffer (Molecular Instruments) for 30 min and left overnight in the dark at room temperature in Amplification Buffer + 0.03 µM of each hairpin (Molecular Instruments). Embryos were washed in the dark in 5X-SSC + 0.1% Triton X and incubated overnight with 1 µg/mL DAPI in NPBT, then washed in NPBT, transferred in a glass-bottomed dish in 100% glycerol and imaged with an Olympus V3000 inverted laser scanning confocal microscope.

Zebrafish. Zebrafish embryos were processed as described in ref. 102. Adult zebrafish were decalcified in Morse solution (10% w/v sodium citrate dihydrate and 25% v/v formic acid in DEPC water) prior to embedding for 24 h at room temp. Adult and 8 dpf larval zebrafish samples were then cleared with Histosol (National Diagnostics) for 3 × 20 min at room temperature, incubated in 1:1 Histosol:Paraffin for 2 × 30 min at 60 °C, then infiltrated with molten paraffin overnight at 60 °C. After additional 4 × 1 h paraffin changes, samples were embedded in peel-a-way molds (Sigma), left to set for 24 h and then sectioned at 7 μm on a Leica RM2125 rotary microtome. Sections were mounted on SuperFrost Plus charged glass slides, and processed for *in situ* HCR¹⁰². Imaging was performed on an Olympus V3000 inverted laser scanning confocal microscope. Images were analyzed using Imaris v.10.0.0 (Oxford Instruments). Due to the high level of autofluorescence from blood and body cavities in whole late embryos, the endoderm was manually segmented using the nuclear staining, and the HCR signal for *FoxQ2III* was then masked within the endoderm to improve visualization (see original image in Supplementary Fig. 10).

Chicken. Dissected and fixed chicken eyes stored in 100% methanol were rehydrated in PBS and embedded in 4% agarose in PBS. 150 μm-thick sections were then obtained using a Leica VT1200S vibratome, and *in situ* HCR on floating sections was performed in a 12-well plate as described in ref. 103. Briefly, sections collected in 12-well plates were bleached for 40 mins (5% Deionized formamide, 1.5% H2O2, 0.2% SSC in nuclease-free water), permeabilized for 1 h (NPBS + 1% DMSO + 1% TritonX), and incubated with *in situ* HCR probes and hairpins (Molecular Instruments) as detailed above for amphioxus, with the only difference that DAPI (1 μg/mL) was added to Amplification Buffer together with hairpins. After washes in 5X-SSC + 0.1% Triton X, sections were mounted on SuperFrost slides with Fluoromount-G (ThermoFisher Scientific 00-4958-02) and imaged using a Zeiss LSM800 inverted laser scanning confocal microscope.

Lamprey. For hybridized chain reaction–fluorescence *in situ* hybridization (HCR–FISH), we adopted the third-generation HCRv3–FISH protocol¹⁰⁰. HCR–FISH probe sets were custom-designed by Molecular Instruments. Following HCR–FISH, embryos and larvae were incubated with SYTOX Green nucleic acid stain (Thermo Fisher Scientific, catalog no. S7020), followed by brief washes in PBS. Samples were mounted in PBS and then imaged using a Nikon C2 confocal microscope.

Skate. Skate S29 embryos were processed for *in situ* HCR as described in ref. 104, with modifications. First, embryos were delipidated by overnight incubation in dichloromethane (DCM). After washes in 100% methanol, the embryonic tail was cut out at the level of the cloaca using razor blades, and samples were rehydrated in PBS and bleached for 30 min (5% Deionized formamide, 1.5% H2O2, 0.2% SSC in nuclease-free water). Embryos were incubated for 2 h with hybridization buffer and then for 3 days with hybridization buffer and 12 μM of probes at 37 °C. Samples were washed with 5X-SSC + 0.1% Triton X, then amplification buffer, and then left for 3 days at 4 °C with 0.15 μM of hairpin and 1 μg/mL of DAPI. After incubation, samples were rinsed with 5X-SSC + 0.1% Triton X and then left in TrisHCl 0.5 M pH7 overnight. The following day, embryos were embedded in 4% agarose in TrisHCl 0.5 M pH7 and progressively dehydrated in methanol. For clearing, after overnight incubation in 100% methanol, samples were treated with 66% DCM and 33% methanol for 3 h, then with 100% DCM for 1 h and then transferred in Dibenzyl ether (DBE) and left overnight. Cleared samples were imaged using an UltraMicroscope II Light-sheet microscope (Miltenyi biotec).

Transcriptomic data analysis

The following published bulk RNAseq and scRNAseq datasets were used to investigate the expression of *FoxQ2* orthologs across bilaterians:

Dreissena polymorpha (Mollusca, Bivalvia): dataset of trochophore larva obtained from ref. 60 available through GEO accession number GSE192624. Normalization (SCTransform), dimensionality reduction and clustering of the raw dataset was carried out using Seurat¹⁰⁵. Marker genes listed in the original publication were used to annotate the clusters.

Hofstenia miamia (Xenacoelomorpha): dataset of hatchling juveniles from ref. 61 available in NCBI BioProject database under accession codes PRJNA88438. *FoxQ2* expression was visualized using the online resource generated by the authors <https://n2t.net/ark:/84478/d/q6xfc7jj>.

Branchiostoma lanceolatum (Chordata, Cephalochordata) bulk RNA-seq datasets of embryonic and larval development and of adult tissues from ref. 65 were available from GEO with accession number GSE106430. Counts were normalized for library size and composition bias using DESeq2¹⁰⁶.

Branchiostoma floridae (Chordata, Cephalochordata). The single nuclei RNAseq dataset of amphioxus development from ref. 66 was available from CNSA of CNGBdb under accession number CNP0000891. Analysis was performed following the code prepared by the authors and deposited at <https://github.com/XingyanLiu/AmphioxusAnalysis>.

Seurat objects of processed scRNAseq datasets of specific embryonic stages (B, G0, N0, T1), corresponding to the moment of activation of *FoxQ2I*, *FoxQ2II* and *FoxQ2III* expression in amphioxus, obtained by ref. 69 were downloaded from the Science Data Bank (<https://doi.org/10.57760/sciencedb.08801>)¹⁰⁷.

Danio rerio: bulk RNAseq dataset of embryonic and larval development from ref. 68, stored in the EBI European Nucleotide Archive with accession numbers PRJEB12296, PRJEB7244 and PRJEB12982, was available to explore in Expression Atlas (<https://www.ebi.ac.uk/gxa/experiments/E-ERAD-475/Results>)¹⁰⁸. Single cell RNAseq of 8 dpf zebrafish larvae from ref. 67 was downloaded from GEO (accession number GSE158142).

All scRNAseq datasets were analyzed and all plots were generated using Seurat v.5.0.3.

Single-cell data integration and comparative analysis

The following amphioxus and vertebrate datasets were subsetted as described and integrated with SAMap using the default pipeline⁷¹: *B. floridae* larval (T1) endoderm subset from original dataset⁶⁹; *B. floridae* adult gut tissues (batch B and control samples only)⁷⁰; *D. rerio* endoderm subset from original dataset (all developmental stages)⁷²; *M. musculus* endoderm, embryonic tissues only⁷³; *M. musculus* endoderm samples (bladder, liver, thymus, pancreas, intestine, trachea, lung) from Tabula Muris Consortium⁷⁴.

To generate protein-protein similar inputs to initialize SAMap, we took the longest isoforms for each gene from the *B. floridae* (based on Dai 2024 genome and annotation), zebrafish (GRCz11), and mouse (GRCm39). A mapping score cutoff of 0.2 was used to select similar clusters for the identification of marker gene pairs.

Prediction of regulatory sequences

The genomes of *Branchiostoma lanceolatum* (v.BI71nemr, GCA_900088365.1), *B. floridae* (Version 2)¹⁰⁹ and *B. belcheri* (v.Haploidv18h27, GCF_001625305.1), as well as the genomes of *Asymmetron lucayanum* indo-pacific clade A and *Epigonichthys maldivensis* available in our labs¹¹⁰, were BLAST-searched for the three *FoxQ2* sequences (default parameters). Then, for *FoxQ2I*- and *FoxQ2II*-type genes, the first 5000 bp upstream of each gene's starting codon sequence were isolated, while for *FoxQ2III*-type genes the upstream 7000 bp were selected to account for the large 5'-UTR found in *B. lanceolatum*. The *FoxQ2I*, *FoxQ2II* and *FoxQ2III* sequences and upstream regions for *Asymmetron* and *Epigonichthys* have been deposited on GenBank with accession numbers PX516856–PX516861. For each gene, the five upstream sequences, one for each species, were searched for conserved regions using mVISTA online tool⁷⁹ (<https://genome.lbl.gov/vista/mvista/submit.shtml>), and the regions present in each species were manually annotated using SnapGene Viewer. Moreover, *B. lanceolatum* ATACseq data at four stages of development (8hpf, 15hpf,

36hpf, 60hpf)⁶⁵, available at GEO under accession number GSE106428, was mapped on the genome using IGV¹¹¹ and the sum of open peaks from all stages was also manually annotated using SnapGene Viewer (Supplementary Fig. 11). For each gene and each species, the sequences with overlapping sequence conservation for all five species (and, for *B. lanceolatum*, also the sequence contained within the ATACseq peaks) were labeled as conserved non-coding sequences (CNCSs). This resulted in three CNCSs for *FoxQ2I*, one CNCS for *FoxQ2II* and two CNCSs for *FoxQ2III* found in all five amphioxus species. Each CNCS from each species was then analyzed for the identification of putative TFBs using Ciiider⁸⁰, using JASPAR 2020 core vertebrate matrix of TFBs¹¹² and the default deficit threshold of 0.15. This resulted in an extensive list of putative TFBs, which were then filtered computationally to only include those found in all five species at the same position: the sequence length was normalized between 0 and 1 and TFB position was used to filter those within the same 10% region for each CNCS (Supplementary Data 3). This list was then further subset to account for the developmental timing of activation of each *FoxQ2* paralog as well as the cell types in which *FoxQ2* is expressed. We therefore used both the published developmental RNAseq dataset of *B. lanceolatum* from ref. 65 and the developmental scRNAseq of *B. floridae* from ref. 69 (see “Transcriptomic data analysis” section for details).

- For *FoxQ2I*, in the bulk-RNAseq we averaged the expression of genes at the 32-cells (maternal) and blastula (B, zygotic) stages, while for scRNAseq we selected clusters labeled as “animal pole” at the blastula (B) stage and “epithelial ectoderm” and “ectoderm mix” at the early gastrula (G0) stage, pseudo-bulked using Seurat’s AggregateExpression function, and calculated the average counts for each gene.
- For *FoxQ2II*, in bulk-RNAseq we averaged expression at 11 hpf (G5) and 15 hpf (N0-N2), while for scRNAseq we focused on the N0 scRNAseq dataset, selected clusters labeled as “epithelial ectoderm” and “neural ectoderm”, and calculated the average for each gene of the pseudo-bulked counts.
- For *FoxQ2III* we averaged expression at 36 hpf (T1/14ss) and 50 hpf (L0/1gs), and selected clusters within the endoderm where *FoxQ2III* was expressed in the T1 scRNAseq dataset (clusters 5, 8, 67), pseudo-bulked and calculated the average counts for each gene.

For each CNCS, we then subset the list of predicted TFBs to include only those expressed at the correct developmental time and cell type in which the corresponding gene is expressed. For RNAseq, we considered “significantly expressed” any gene whose expression was higher than the 25th percentile of normalized counts for the whole transcriptome at the stages considered (approximated at 17), while for scRNAseq, to account for the high number of zero counts in different cells, we considered “significantly expressed” genes for which expression was above the 50th percentile (approximated at >250). The final list for each gene is stored in Supplementary Data 3.

Statistics and reproducibility

Different phylogenetic analysis methods (Maximum Likelihood, Neighbor Joining) were compared to provide further support to tree structure. In each analysis, 1000 ultrafast bootstraps were used to extract branch support values. For *in situ* HCRs, the expression of every gene was investigated in at least 2 separate experimental replicates, each comprising 3 or more samples per species per stage. For the transcriptomic comparison of *FoxQ2III*-positive cells across chordate species, identification and ranking of marker gene pairs was performed with the GenePairFinder function in SAMap, which identifies genes that contribute most to positive cross-species correlation for a given cell type pair. Briefly, marker genes are designated in each species using a Wilcoxon rank-sum test with a p-value cutoff of 0.01. Then, all cross-species pairs of significant marker genes are ranked by the product of their expression levels (zero-truncated standardized expression) and SAMap weights (derived from protein sequence similarity and expression correlation) in each species.

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

All data required to evaluate the conclusions in the paper are present in the paper, Supplementary Figs. and Supplementary Data. SAMap mapping files and integrated datasets of embryonic/larval and adult amphioxus, mouse and zebrafish datasets can be downloaded from Zenodo at <https://doi.org/10.5281/zenodo.17143655>¹¹³. *FoxQ2I*, *FoxQ2II* and *FoxQ2III* sequences and upstream regions for *Asymmetron lucayanum* indo-pacific clade A and *Epigonichthys maldivensis* have been deposited on GenBank with accession numbers PX516856-PX516861. Publicly available datasets used in this study include: *Dreissena polymorpha* larval scRNAseq (Gene Expression Omnibus (GEO) accession number GSE192624)⁶⁰; *Hofstenia miamia* hatchling juveniles scRNAseq (NCBI BioProject PRJNA888438)⁶¹; *Branchiostoma lanceolatum* bulk RNAseq datasets of development and adult tissues (GEO accession number GSE106430)⁶⁵; *Branchiostoma floridae* snRNAseq developmental datasets (CNSA of CNGBdb accession number CNP0000891⁶⁶, and Science Data Bank <https://doi.org/10.57760/sciencedb.08801>⁶⁹), and snRNAseq adult gut tissues (Open Archive for Miscellaneous Data, National Genomics Data Center, China National Center for Bioinformation/Beijing Institute of Genomics, Chinese Academy of Sciences accession number OMIX006304)⁷⁰; *Danio rerio* embryonic bulk RNAseq dataset (EBI European Nucleotide Archive accession numbers PRJEB12296, PRJEB7244 and PRJEB12982)⁶⁸, and scRNAseq datasets of 8 dpf larvae (GEO accession number GSE158142)⁶⁷ and of developmental stages (GEO accession number GSE223922)⁷²; *Mus musculus* scRNAseq datasets of embryonic (GEO accession number GSE266977)⁷³ and adult (GEO accession number GSE109774)⁷⁴ endoderm. For more specific information consult the “Methods” section of the paper.

Code availability

All the code used to generate figures and data presented in this paper is available at https://github.com/eBGLab/FoxQ2_Evolution and on Zenodo at <https://doi.org/10.5281/zenodo.17143655>¹¹³.

Received: 19 August 2025; Accepted: 25 November 2025;

Published online: 21 December 2025

References

1. Shimeld, S. M., Degnan, B. & Luke, G. N. Evolutionary genomics of the Fox genes: origin of gene families and the ancestry of gene clusters. *Genomics* **95**, 256–260 (2010).
2. Yu, J.-K. et al. The Fox genes of *Branchiostoma floridae*. *Dev. Genes Evol.* **218**, 629–638 (2008).
3. Golson, M. L. & Kaestner, K. H. Fox transcription factors: from development to disease. *Development* **143**, 4558–4570 (2016).
4. Hannenhalli, S. & Kaestner, K. H. The evolution of Fox genes and their role in development and disease. *Nat. Rev. Genet.* **10**, 233–240 (2009).
5. Li, C. & Tucker, P. W. DNA-binding properties and secondary structural model of the hepatocyte nuclear factor 3/fork head domain. *Proc. Natl. Acad. Sci.* **90**, 11583–11587 (1993).
6. Gajiwala, K. S. & Burley, S. K. Winged helix proteins. *Curr. Opin. Struct. Biol.* **10**, 110–116 (2000).
7. Nakagawa, S., Gisselbrecht, S. S., Rogers, J. M., Hartl, D. L. & Bulyk, M. L. DNA-binding specificity changes in the evolution of forkhead transcription factors. *Proc. Natl. Acad. Sci.* **110**, 12349–12354 (2013).
8. Yu, J. K., Holland, N. D. & Holland, L. Z. AmphiFoxQ2, a novel winged helix/forkhead gene, exclusively marks the anterior end of the amphioxus embryo. *Dev. Genes Evol.* **213**, 102–105 (2003).

9. Mazet, F., Yu, J.-K., Liberles, D. A., Holland, L. Z. & Shimeld, S. M. Phylogenetic relationships of the Fox (Forkhead) gene family in the Bilateria. *Gene* **316**, 79–89 (2003).
10. Schomburg, C., Janssen, R. & Prpic, N.-M. Phylogenetic analysis of forkhead transcription factors in the Panarthropoda. *Dev. Genes Evol.* **232**, 39–48 (2022).
11. Seudre, O. et al. The Fox gene repertoire in the annelid *Owenia fusiformis* reveals multiple expansions of the foxQ2 class in Spiralia. *Genome Biol. Evol.* <https://doi.org/10.1093/gbe/evac139> (2022).
12. Pascual-Carreras, E. et al. Analysis of Fox genes in *Schmidtea mediterranea* reveals new families and a conserved role of Smed-foxO in controlling cell death. *Sci. Rep.* **11**, 2947 (2021).
13. Fritzenwanker, J. H., Gerhart, J., Freeman, R. M. & Lowe, C. J. The Fox/Forkhead transcription factor family of the hemichordate *Saccoglossus kowalevskii*. *EvoDevo* **5**, 17 (2014).
14. Sinigaglia, C., Busengdal, H., Leclère, L., Technau, U. & Rentzsch, F. The Bilaterian head patterning gene *six3/6* controls aboral domain development in a cnidarian. *PLoS Biol.* **11**, e1001488 (2013).
15. Chevalier, S., Martin, A., Leclère, L., Amiel, A. & Houlston, E. Polarised expression of FoxB and FoxQ2 genes during development of the hydrozoan *Clytia hemisphaerica*. *Dev. Genes Evol.* **216**, 709–720 (2006).
16. Marlow, H. et al. Larval body patterning and apical organs are conserved in animal evolution. *BMC Biol.* **12**, 7 (2014).
17. Santagata, S., Resh, C., Hejnol, A., Martindale, M. Q. & Passamaneck, Y. J. Development of the larval anterior neurogenic domains of *Terebratalia transversa* (Brachiopoda) provides insights into the diversification of larval apical organs and the spiralian nervous system. *EvoDevo* **3**, 3 (2012).
18. Hunnekuhl, V. S. & Akam, M. An anterior medial cell population with an apical-organ-like transcriptional profile that pioneers the central nervous system in the centipede *Strigamia maritima*. *Dev. Biol.* **396**, 136–149 (2014).
19. Gąsiorowski, L. & Hejnol, A. Hox gene expression during development of the phoronid *Phoronopsis harmeri*. *EvoDevo* **11**, 2 (2020).
20. Martín-Durán, J. M., Vellutini, B. C. & Hejnol, A. Evolution and development of the adelphophagous, intracapsular Schmidt's larva of the nemertean *Lineus ruber*. *EvoDevo* **6**, 28 (2015).
21. Range, R. Canonical and non-canonical Wnt signaling pathways define the expression domains of *Frizzled 5/8* and *Frizzled 1/2/7* along the early anterior-posterior axis in sea urchin embryos. *Dev. Biol.* **444**, 83–92 (2018).
22. Range, R. C. & Wei, Z. An anterior signaling center patterns and sizes the anterior neuroectoderm of the sea urchin embryo. *Development* **143**, 1523–1533 (2016).
23. Range, R., Angerer, R. C. & Angerer, L. M. Integration of canonical and noncanonical Wnt signaling pathways patterns the neuroectoderm along the anterior-posterior axis of sea urchin embryos. *PLoS Biol.* **11**, e1001467 (2013).
24. Yaguchi, S., Yaguchi, J., Angerer, R. C. & Angerer, L. M. A Wnt-FoxQ2-Nodal pathway links primary and secondary axis specification in sea urchin embryos. *Dev. Cell* **14**, 97–107 (2008).
25. Yaguchi, S. et al. Fez function is required to maintain the size of the animal plate in the sea urchin embryo. *Development* **138**, 4233–4243 (2011).
26. Yaguchi, J., Takeda, N., Inaba, K. & Yaguchi, S. Cooperative Wnt-Nodal signals regulate the patterning of anterior neuroectoderm. *PLOS Genet.* **12**, e1006001 (2016).
27. Kitzmann, P., Weißkopf, M., Schacht, M. I. & Bucher, G. A key role for *foxQ2* in anterior head and central brain patterning in insects. *Development* **144**, 2969–2981 (2017).
28. He, B. et al. An ancestral apical brain region contributes to the central complex under the control of *foxQ2* in the beetle *Tribolium*. *eLife* **8**, e49065 (2019).
29. Schacht, M. I., Schomburg, C. & Bucher, G. *six3* acts upstream of *foxQ2* in labrum and neural development in the spider *Parasteatoda tepidariorum*. *Dev. Genes Evol.* **230**, 95–104 (2020).
30. Yaguchi, J. & Yaguchi, S. Rx and its downstream factor, Musashi1, is required for establishment of the apical organ in sea urchin larvae. *Front. Cell Dev. Biol.* **11**, 1240767 (2023).
31. Cheatle Jarvela, A. M., Yankura, K. A. & Hinman, V. F. A gene regulatory network for apical organ neurogenesis and its spatial control in sea star embryos. *Development* **143**, 4214–4223 (2016).
32. Gattoni, G., Keitley, D., Sawle, A. & Benito-Gutiérrez, E. An ancient apical patterning system sets the position of the forebrain in chordates. *Sci. Adv.* **11**, eadq4731 (2025).
33. Range, R. Specification and positioning of the anterior neuroectoderm in deuterostome embryos. *Genesis* **52**, 222–234 (2014).
34. Arendt, D., Tosches, M. A. & Marlow, H. From nerve net to nerve ring, nerve cord and brain—evolution of the nervous system. *Nat. Rev. Neurosci.* **17**, 61–72 (2016).
35. Feuda, R. & Peter, I. S. Homologous gene regulatory networks control development of apical organs and brains in Bilateria. *Sci. Adv.* **8**, eab02416 (2022).
36. Posnien, N., Hunnekuhl, V. S. & Bucher, G. Gene expression mapping of the neuroectoderm across phyla—conservation and divergence of early brain anlagen between insects and vertebrates. *eLife* **12**, e92242 (2023).
37. Fenner, J. L., Newberry, C., Todd, C. & Range, R. C. Anterior-posterior Wnt signaling network conservation between indirect developing sea urchin and hemichordate embryos. *Integr. Comp. Biol.* **64**, 1214–1225 (2024).
38. Leclère, L., Bause, M., Sinigaglia, C., Steger, J. & Rentzsch, F. Development of the aboral domain in *Nematostella* requires β -catenin and the opposing activities of *six3/6* and *frizzled5/8*. *Development* **120931**. <https://doi.org/10.1242/dev.120931> (2016).
39. Vöcking, O., Kourtesis, I. & Hausen, H. Posterior eyespots in larval chitons have a molecular identity similar to anterior cerebral eyes in other bilaterians. *EvoDevo* **6**, 40 (2015).
40. Yang, M. et al. Phylogeny of forkhead genes in three spiralian and their expression in Pacific oyster *Crassostrea gigas*. *Chin. J. Oceanol. Limnol.* **32**, 1207–1223 (2014).
41. Wu, S. et al. Identification and expression profiles of Fox transcription factors in the Yesso scallop (*Patinopecten yessoensis*). *Gene* **733**, 144387 (2020).
42. Ong, T.-H. et al. Mass spectrometry imaging and identification of peptides associated with cephalic ganglia regeneration in *Schmidtea mediterranea*. *J. Biol. Chem.* **291**, 8109–8120 (2016).
43. Janssen, R., Schomburg, C., Prpic, N.-M. & Budd, G. E. A comprehensive study of arthropod and onychophoran Fox gene expression patterns. *PLoS ONE* **17**, e0270790 (2022).
44. Yankura, K. A., Martik, M. L., Jennings, C. K. & Hinman, V. F. Uncoupling of complex regulatory patterning during evolution of larval development in echinoderms. *BMC Biol.* **8**, 143 (2010).
45. Su, Y.-H. et al. BMP controls dorsoventral and neural patterning in indirect-developing hemichordates providing insight into a possible origin of chordates. *Proc. Natl. Acad. Sci.* **116**, 12925–12932 (2019).
46. Mercurio, S. et al. A feather star is born: embryonic development and nervous system organization in the crinoid *Antedon mediterranea*. *Open Biol.* **14**, 240115 (2024).
47. Paganos, P., Voronov, D., Musser, J., Arendt, D. & Arnone, M. I. Single cell RNA sequencing of the *Strongylocentrotus purpuratus* larva reveals the blueprint of major cell types and nervous system of a non-chordate deuterostome. *eLife* **10**, e70416 (2021).
48. Yuan, H., Hatleberg, W. L., Degnan, B. M. & Degnan, S. M. Gene activation of metazoan Fox transcription factors at the onset of metamorphosis in the marine demosponge *Amphimedon queenslandica*. *Dev. Growth Differ.* **64**, 455–468 (2022).

49. Ogawa, Y., Shiraki, T., Fukada, Y. & Kojima, D. Foxq2 determines blue cone identity in zebrafish. *Sci. Adv.* **7**, eabi9784 (2021).

50. Timoshevskaya, N. et al. An improved germline genome assembly for the sea lamprey *Petromyzon marinus* illuminates the evolution of germline-specific chromosomes. *Cell Rep.* **42**, 112263 (2023).

51. Marlétaz, F. et al. The little skate genome and the evolutionary emergence of wing-like fins. *Nature* **616**, 495–503 (2023).

52. Brown, T. et al. Chromosome-scale genome assembly reveals how repeat elements shape non-coding RNA landscapes active during newt limb regeneration. *Cell Genom.* **5**, 100761 (2025).

53. Simakov, O. et al. Deeply conserved synteny resolves early events in vertebrate evolution. *Nat. Ecol. Evol.* **4**, 820–830 (2020).

54. Simakov, O. et al. Deeply conserved synteny and the evolution of metazoan chromosomes. *Sci. Adv.* **8**, eabi5884 (2022).

55. Schultz, D. T. et al. Ancient gene linkages support ctenophores as sister to other animals. *Nature* **618**, 110–117 (2023).

56. Force, A. et al. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151**, 1531–1545 (1999).

57. Gonzalez, P., Uhlinger, K. R. & Lowe, C. J. The adult body plan of indirect developing hemichordates develops by adding a Hox-patterned trunk to an anterior larval territory. *Curr. Biol.* **27**, 87–95 (2017).

58. Lapan, S. W. & Reddien, P. W. Transcriptome analysis of the planarian eye identifies ovo as a specific regulator of eye regeneration. *Cell Rep.* **2**, 294–307 (2012).

59. Gąsiorowski, L. Evidence for multiple independent expansions of fox gene families within flatworms. *J. Mol. Evol.* <https://doi.org/10.1007/s00239-024-10226-4> (2025).

60. Salamanca-Díaz, D. A., Schulreich, S. M., Cole, A. G. & Wanninger, A. Single-cell RNA sequencing atlas from a bivalve larva enhances classical cell lineage studies. *Front. Ecol. Evol.* **9**, 783984 (2022).

61. Hulett, R. E. et al. Acoel single-cell atlas reveals expression dynamics and heterogeneity of adult pluripotent stem cells. *Nat. Commun.* **14**, 2612 (2023).

62. Jondelius, U., Raikova, O. I. & Martinez, P. Xenacoelomorpha, a key group to understand Bilaterian evolution: morphological and molecular perspectives. in *Evolution, Origin of Life, Concepts and Methods* (ed Pontarotti, P.) 287–315. https://doi.org/10.1007/978-3-030-30363-1_14. (Springer International Publishing, 2019).

63. Hulett, R. E., Potter, D. & Srivastava, M. Neural architecture and regeneration in the acoel *Hofstenia miamia*. *Proc. R. Soc. B Biol. Sci.* **287**, 20201198 (2020).

64. Onai, T. Canonical Wnt/β-catenin and Notch signaling regulate animal/vegetal axial patterning in the cephalochordate amphioxus. *Evol. Dev.* **21**, 31–43 (2019).

65. Marlétaz, F. et al. Amphioxus functional genomics and the origins of vertebrate gene regulation. *Nature* **564**, 64–70 (2018).

66. Ma, P. et al. Joint profiling of gene expression and chromatin accessibility during amphioxus development at single-cell resolution. *Cell Rep.* **39**, 110979 (2022).

67. Raj, B. et al. Emergence of neuronal diversity during vertebrate brain development. *Neuron* **108**, 1058–1074.e6 (2020).

68. White, R. J. et al. A high-resolution mRNA expression time course of embryonic development in zebrafish. *eLife* **6**, e30860 (2017).

69. Dai, Y. et al. Evolutionary origin of the chordate nervous system revealed by amphioxus developmental trajectories. *Nat. Ecol. Evol.* <https://doi.org/10.1038/s41559-024-02469-7> (2024).

70. Dai, Y. et al. Single-cell profiling of the amphioxus digestive tract reveals conservation of endocrine cells in chordates. *Sci. Adv.* **10**, eadq0702 (2024).

71. Tarashansky, A. J. et al. Mapping single-cell atlases throughout Metazoa unravels cell type evolution. *eLife* **10**, e66747 (2021).

72. Sur, A. et al. Single-cell analysis of shared signatures and transcriptional diversity during zebrafish development. *Dev. Cell* **58**, 3028–3047.e12 (2023).

73. Li, K.-R. et al. Spatiotemporal and genetic cell lineage tracing of endodermal organogenesis at single-cell resolution. *Cell* **188**, 796–813.e24 (2025).

74. The Tabula Muris Consortium et al. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367–372 (2018).

75. Darras, S. et al. Anteroposterior axis patterning by early canonical Wnt signaling during hemichordate development. *PLoS Biol.* **16**, e2003698 (2018).

76. Yamazaki, A., Yamamoto, A., Yaguchi, J. & Yaguchi, S. *cis*-Regulatory analysis for later phase of anterior neuroectoderm-specific foxQ2 expression in sea urchin embryos. *Genesis* **57**, e23302 (2019).

77. Yaguchi, S., Yaguchi, J. & Inaba, K. Bicaudal-C is required for the formation of anterior neurogenic ectoderm in the sea urchin embryo. *Sci. Rep.* **4**, 6852 (2014).

78. Yaguchi, J., Yamazaki, A. & Yaguchi, S. Meis transcription factor maintains the neurogenic ectoderm and regulates the anterior-posterior patterning in embryos of a sea urchin, *Hemicentrotus pulcherrimus*. *Dev. Biol.* **444**, 1–8 (2018).

79. Frazer, K. A., Pachter, L., Poliakov, A., Rubin, E. M. & Dubchak, I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res.* **32**, 273–279 (2004).

80. Gearing, L. J. et al. CiiIDER: a tool for predicting and analysing transcription factor binding sites. *PLoS ONE* **14**, e0215495 (2019).

81. Dunn, C. W. et al. Broad phylogenomic sampling improves resolution of the animal tree of life. *Nature* **452**, 745–749 (2008).

82. Whelan, N. V. et al. Ctenophore relationships and their placement as the sister group to all other animals. *Nat. Ecol. Evol.* **1**, 1737–1746 (2017).

83. Philippe, H. et al. Phylogenomics revives traditional views on deep animal relationships. *Curr. Biol.* **19**, 706–712 (2009).

84. Simion, P. et al. A large and consistent phylogenomic dataset supports sponges as the sister group to all other animals. *Curr. Biol.* **27**, 958–967 (2017).

85. Girstmair, J. Building a light-sheet microscope to study the early development of the polyclad flatworm *Maritigrella crozieri* (Hyman, 1939). Doctoral thesis, University College London, London, UK. (2017). <https://discovery.ucl.ac.uk/id/eprint/1573350/>.

86. Yaguchi, S. et al. AnkAT-1 is a novel gene mediating the apical tuft formation in the sea urchin embryo. *Dev. Biol.* **348**, 67–75 (2010).

87. Wang, J. et al. The conserved domain database in 2023. *Nucleic Acids Res.* **51**, D384–D388 (2023).

88. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).

89. Trifinopoulos, J., Nguyen, L. T., Haeseler, A. von & Minh, B. Q. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* **44**, W232–W235 (2016).

90. Gouy, M., Guindon, S. & Gascuel, O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Mol. Biol. Evol.* **27**, 221–224 (2010).

91. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).

92. Emms, D. M. & Kelly, S. OrthoFinder: Phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**, 238 (2019).

93. Tang, H. et al. JCVI: a versatile toolkit for comparative genomics analysis. *iMeta* **3**, e211 (2024).

94. Tang, H. et al. Synteny and collinearity in plant genomes. *Science* **320**, 486–488 (2008).

95. Tsai, F.-Y., Lin, C.-Y., Su, Y.-H., Yu, J.-K. & Kuo, D.-H. Evolutionary history of bilaterian FoxP genes: complex ancestral functions and evolutionary changes spanning 2R-WGD in the vertebrate lineage. *Mol. Biol. Evol.* msaf072. <https://doi.org/10.1093/molbev/msaf072> (2025).

96. Lin, C.-Y. et al. Chromosome-level genome assemblies of 2 hemichordates provide new insights into deuterostome origin and chromosome evolution. *PLoS Biol.* **22**, e3002661 (2024).

97. Benito-Gutiérrez, È, Weber, H., Bryant, D. V. & Arendt, D. Methods for generating year-round access to *Amphioxus* in the laboratory. *PLoS ONE* **8**, e71599 (2013).

98. Hamburger, V. & Hamilton, H. L. A series of normal stages in the development of the chick embryo. *Dev. Dyn.* **195**, 231–272 (1992).

99. Gillis, J. A. et al. Big insight from the little skate: *Leucoraja erinacea* as a developmental model system. *Curr. Top. Dev. Biol.* **147**, 595–630 (2022).

100. Choi, H. M. T. et al. Third-generation *in situ* hybridization chain reaction: multiplexed, quantitative, sensitive, versatile, robust. *Development* **145**, dev165753 (2018).

101. Andrews, T. G., Gattoni, G., Busby, L., Schwimmer, M. A. & Benito-Gutiérrez, È. Hybridization chain reaction for quantitative and multiplex imaging of gene expression in amphioxus embryos and adult tissues. in *In Situ Hybridization Protocols* (eds Nielsen, B. S. & Jones, J.) 179–194. <https://doi.org/10.1007/978-1-0716-0623-0>. (Springer Nature, 2020).

102. Rees, J. M. et al. A pre-vertebrate endodermal origin of calcitonin-producing neuroendocrine cells. *Development* **151**, dev202821 (2024).

103. Gumnit, E. et al. Evolution of Cajal-Retzius cells in vertebrates from an ancient class of *Tp73*+ neurons. *bioRxiv* <https://doi.org/10.1101/2023.01.04.522730> (2025).

104. Woych, J. et al. Cell-type profiling in salamanders identifies innovations in vertebrate forebrain evolution. *Science* **377**, eabp9186 (2022).

105. Satija, R., Farrell, J. A., Gennert, D., Schier, A. F. & Regev, A. Spatial reconstruction of single-cell gene expression data. *Nat. Biotechnol.* **33**, 495–502 (2015).

106. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

107. Chengzan, L., Yanfei, H., Jianhui, L. & Lili, Z. ScienceDB: A Public Multidisciplinary Research Data Repository for eScience. In *Proc. IEEE 13th International Conference on e-Science (e-Science)* 248–255. <https://doi.org/10.1109/eScience.2017.83> (IEEE, Auckland, 2017).

108. Petryszak, R. et al. Expression atlas update—an integrated database of gene and protein expression in humans, animals and plants. *Nucleic Acids Res.* **44**, D746–D752 (2016).

109. Putnam, N. H. et al. The amphioxus genome and the evolution of the chordate karyotype. *Nature* **453**, 1064–1072 (2008).

110. Lotharukpong, J. S., Laumer, C. E. & Benito-Gutiérrez, È. Phylogenetic discordance and genic innovation at the emergence of modern cephalochordates. *bioRxiv* <https://doi.org/10.1101/2025.10.14.682400> (2025).

111. Thorvaldsdottir, H., Robinson, J. T. & Mesirov, J. P. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief. Bioinform.* **14**, 178–192 (2013).

112. Fornes, O. et al. JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **gkz1001**. <https://doi.org/10.1093/nar/gkz1001> (2019).

113. Gattoni, G., Shew, C. & Benito Gutierrez, E. Evolutionary dynamics of FoxQ2 transcription factors across metazoans reveals three ancient paralogs. *Zenodo* <https://doi.org/10.5281/ZENODO.17143655> (2025).

Acknowledgements

We thank Michael Schwimmer for helping with the analysis of amphioxus adult RNAseq, Jordi Paps for advice and feedback on the phylogenetic analysis, Panagiotis Oikonomou and Nandan L. Nerurkar for providing chicken embryos and Mansi Srivastava for help with the *Hofstenia miamia* dataset. We acknowledge funding from the Whitten Programme to G.G., the Wellcome Trust Mathematical Genomics Medicine Programme at the

University of Cambridge (PFZH/158 RG92770) to D.K., and Life Sciences Research Foundation postdoctoral fellowship sponsored by Walder Foundation to J.R.Y. Work in the E.B.G. lab was supported by the CRUK (C9545/A29580). Work in the J.K.Y. lab was supported by Academia Sinica (AS-GC-111-L01) and National Science and Technology Council, Taiwan (113-2621-B-001-004-MY3). Work in the C.L. lab was supported by the National Institutes of Health (R01GM116538), the National Science Foundation (1764421) and the Simons Foundation (SFARI 597491-RWC). Work in the J.A.G. lab was supported by the Royal Society University Research Fellowship (UF130182 and URF\RV191007) and a Royal Society Research Grant (RG140377).

Author contributions

G.G. and E.B.G. designed the project; E.B.G. directed the project; G.G. performed phylogenetic and microsynteny analyses, collected amphioxus and chick samples and carried out *in situ* HCR and imaging on amphioxus, skate, zebrafish and chick samples; C.Y.L. and J.K.Y. performed macrosynteny analyses; G.G. and C.S. carried out transcriptomic analyses; J.R.Y. and C.L. collected and performed *in situ* HCR and imaging on lamprey samples; J.A.G. collected and sectioned zebrafish and skate samples; G.G. and D.K. carried out the analysis of regulatory sequences; G.G., C.Y.L., and C.S. generated figures for data visualization; G.G. wrote the first draft of the manuscript; all authors reviewed the manuscript.

Competing interests

E.B.G. and C.S. have been employees of Genentech since September 2022 and April 2023, respectively. All other authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s42003-025-09368-y>.

Correspondence and requests for materials should be addressed to Giacomo Gattoni or Elia Benito-Gutiérrez.

Peer review information *Communications Biology* thanks Gregor Bucher, Didier Casane, and the other, anonymous, reviewer for their contribution to the peer review of this work. Primary Handling Editors: Manuel Breuer and George Inglis.

Reprints and permissions information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2025