OXFORD

Briefings in Functional Genomics, 17(5), 2018, 352–361

doi: 10.1093/bfgp/ely008
Advance Access Publication Date: 16 March 2018
Review paper

Inference of the ancestral vertebrate phenotype through vestiges of the whole-genome duplications

Koh Onimaru and Shigehiro Kuraku

Corresponding author: Shigehiro Kuraku, Laboratory for Phyloinformatics, RIKEN Center for Biosystems Dynamics Research (BDR), 2-2-3 Minatojima-minamimachi, Chuo-ku, Kobe, Hyogo 650-0047 Japan. Tel.: +81-78-306-3048; E-mail: shigehiro.kuraku@riken.jp

Abstract

Inferring the phenotype of the last common ancestor of living vertebrates is a challenging problem because of several unresolvable factors. They include the lack of reliable out-groups of living vertebrates, poor information about less fossilizable organs and specialized traits of phylogenetically important species, such as lampreys and hagfishes (e.g. secondary loss of vertebrae in adult hagfishes). These factors undermine the reliability of ancestral reconstruction by traditional character mapping approaches based on maximum parsimony. In this article, we formulate an approach to hypothesizing ancestral vertebrate phenotypes using information from the phylogenetic and functional properties of genes duplicated by genome expansions in early vertebrate evolution. We named the conjecture as 'chronological reconstruction of ohnolog functions (CHROF)'. This CHROF conjecture raises the possibility that the last common ancestor of living vertebrates may have had more complex traits than currently thought.

Key words: whole-genome duplication; subfunctionalization; cyclostomes; vertebrate ancestor; paired fins; ohnolog

Challenges in reconstructing the phenotypes of the vertebrate ancestor

Having a valid picture of the ancestral state of all vertebrates is essential to understand the evolutionary process that gave rise to the diverse phenotypes of extant vertebrates. In terms of morphological evolution, the fossil record, comparative anatomy and embryology have been major sources to reconstruct the ancestral phenotype of vertebrates [1]. However, the difficulty of such attempts has been increasingly recognized because of the derived characters of phylogenetically important species, hagfishes and lampreys [2]. Hagfishes and lampreys are living jawless fishes, and serve as the out-groups of jawed vertebrates. Thus, the inference of the last common ancestor of living vertebrates critically depends on whether these jawless fishes share traits that jawed vertebrates have. Apparently, hagfishes and lampreys have also experienced secondary modifications in their isolated lineages over the same amount of evolutionary time—about half a billion years—as jawed vertebrates have spent. As a result, the highly derived and mostly degenerative characters of hagfishes and lampreys have confused the view on early vertebrate evolution [2]. In recent studies on hagfishes and lampreys, comparisons of gene expression patterns and regulation during embryogenesis have eased such difficulty, revealing unexpected complexity of the last common ancestor of all living vertebrates, such as the brain organization, the neural crest cells and the migratory muscles [3–6]. Nevertheless, by relying on such divergent species, attempts to infer the ancestral state of living vertebrates still suffer from a risk of false negatives. In other words, the high frequency of secondary trait changes fundamentally limits reliable ancestral reconstruction of phenotypes, as long as one relies on reconstruction under the principle of maximum parsimony. In

Koh Onimaru is a postdoctoral researcher in RIKEN Center for Life Science Technologies [RIKEN Center for Biosystems Dynamics Research (BDR) from April 2018] in Kobe Japan, and has a PhD in biological science (Tokyo Institute of Technology).
Shigehiro Kuraku is leading Phyloinformatics Unit in RIKEN Center for Life Science Technologies [RIKEN Center for Biosystems Dynamics Research (BDR) from April 2018] in Kobe, Japan. He conducts various molecular evolutionary studies while supervising a cutting-edge DNA sequencing and genome informatics core facility.

© The Author(s) 2018. Published by Oxford University Press.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/4.0/), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact journals.permissions@oup.com

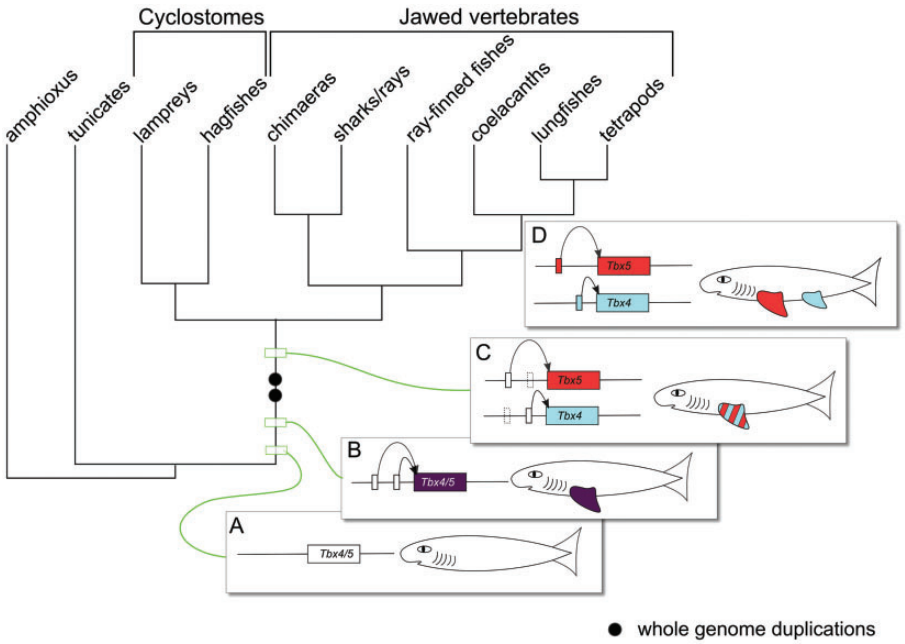

Figure 1. Mapping hypothetical phenotypes to vertebrate phylogeny (**A–D**) schematic representations of a hypothetical vertebrate ancestor before the acquisition of paired fins (**A**), and ones just before and after the WGDs (**B** and **C**), and extant jawed vertebrates (**D**). Black circle, a whole-genome duplication. Note that the first pair of fins appeared with *Tbx4/5* expression (dark purple) before the WGDs. After the WGDs, the paired fins temporally expressed both *Tbx4* and *Tbx5* (red and blue stripes).

addition, the difficulty of ancestral reconstruction is doubled by the lack of a reliable out-group of living vertebrates because of the severe phenotypic gap between them and their immediate out-group, namely, tunicates and cephalochordates.

In this article, we introduce a previously unformulated approach to inferring the ancestral phenotype of vertebrates, which does not heavily rely on either character mapping vulnerable to secondary loss or out-group comparison. As a premise, we first review modern knowledge of vertebrate phylogeny and genome evolution and later illustrate how those lines of knowledge assist the inference of the chronological order of events in early vertebrate evolution.

Three surviving vertebrate lineages: How are they related to each other?

Vertebrate phylogeny has been debated and revised repeatedly. Here, we focus on the position of the jawless fishes (hagfishes and lampreys, classified into cyclostomes) because their phylogenetic placement is tightly related to the timing of the genome expansion. An early split of a certain evolutionary lineage sometimes evokes researchers' expectation that descendants of that lineage have maintained phenotypic characters at the time of the early split. Apparently, extant jawless fishes (hagfishes and lampreys, classified into cyclostomes) have been studied for this expectation (reviewed in [7]). Traditionally, phylogeny reconstructions based on phenotypic characters did not produce unequivocal results because of highly variable characters between hagfishes and lampreys (outlined in [8]). As a result, the relationships between the three extant vertebrate lineages, namely, jawed vertebrates, hagfishes and lampreys had been an unresolved issue until recently. In these two decades, however, a wealth of molecular data has accumulated, which supported monophyly of cyclostomes (Figure 1; reviewed in [9]: also see [10] for a different result based partly on phenotypic data). Nowadays, even paleontological articles adopt the cyclostome monophyly (e.g. [11, 12]), and this revised phylogeny serves a more reliable framework for ancestral reconstruction of phenotypic characters.

When in early vertebrate evolution did the alleged genome expansion occur?

Gene duplication is known as a universal mechanism that generates an additional gene copy and is divided into two major types, small-scale duplication (SSD) and whole-genome duplication (WGD). Additional gene copies generated by the latter type of duplications are called ohnologs, after the name of the first advocate of WGDs in the chordate lineage, Susumu Ohno [13]. While duplications of the former types are occurring intermittently during evolution, the latter occurred only a limited number of times, including the early vertebrate period, as far as extant evolutionary lineages are concerned [14]. Through gene duplication, the original function of a gene is inherited by descendant genes. Depending on the mode of the inheritance of gene functions, this process is divided into neofunctionalization that confers a new function to a novel gene and subfunctionalization in which the original function is divided by the duplicates, sometimes with coordination between them [15].

It is a widely accepted theory that two rounds of WGD (2R-WGD) occurred in chordate evolution ([16, 17]; reviewed in [18]; also see in [19, 20]). Sporadic studies on individual gene families involving hagfish or/and lamprey genes sought to determine whether the genome expansion took place before or after the split of the cyclostome lineage from the future jawed vertebrate lineage (e.g. [21, 22]). However, these efforts did not result in a uniform conclusion. To tackle this difficulty, one of the authors performed an analysis using 55 gene families controlled systematically with probabilistic phylogeny inference and its statistical evaluation [9]. This analysis led to the first proposal of the evolutionary scenario that the cyclostome lineage branched off from the future jawed vertebrate lineage after

the two-round WGD (PV4 hypothesis; Figure 1). This hypothesis has been corroborated by subsequent studies, provided that they were based on exhaustive search and identification of homologs and elaborate phylogeny inference (e.g. [23–25]).

If the genome expansion was achieved through 'whole-genome' duplications, why did inferences of its timing yield contradictory results between different gene families? Especially, in phylogenetic trees, why do cyclostome genes stay unsettled in terms of phylogenetic affinity to any jawed vertebrate paralog and frequently form a cluster among them, while the numbers of paralogs are often similar between cyclostomes and jawed vertebrates (e.g. [26])? To gain clues about these questions, a group including one of the authors previously focused on the homeobox-containing *Emx* genes [27], in which a lamprey lineage-specific gene duplication was previously proposed [28]. The reexamination demonstrated a possibility that the exclusive clustering of multiple lamprey genes often observed in molecular phylogeny could be caused by secondary changes in their sequences introduced independently into multiple ancient paralogs [27]. This effect, characteristic of lamprey genes, may be operating on a genome scale (termed 'lamprey dialect'; [29, 30]), and it is worth examining whether hagfish genes exhibit a similar trend. It has more recently been proposed that a duplicated genome experiences massive local recombinations, namely, rediploidization, causing 'lineage-specific ohnolog resolution (LORe)' [31]. This phenomenon can also explain the abovementioned intragenomic discrepancy in gene phylogeny. Moreover, gene phylogeny can also be confused by reciprocal gene loss between the cyclostome and jawed vertebrate lineages (hidden paralogy; reviewed in [32]). These explanations claim that the PV4 hypothesis holds as the default, even though the phylogenetic pattern deduced from this hypothesis is not necessarily supported by molecular phylogeny of some genes (reviewed in [33]).

In addition to the genome expansion in the lineage leading to jawed vertebrates, the lamprey lineage has also been proposed to have experienced WGD [34], following more fragmental evidence presented more than a decade ago [35, 36]. However, the most recently proposed evidence that have revived this hypothesis [34] was based on paralog distance distribution, which can largely be confused by the secondary effect mentioned above (lamprey dialect and/or LORe) and needs to be assessed further by means of exploring genome-wide conservation of duplicated synteny. Below, we discuss ancestral vertebrate phenotypes assuming that the 2R-WGD occurred before the radiation of all extant vertebrates including the monophyletic cyclostomes.

Inference of a pre-WGD state: paired appendages as an example

While the WGDs have been hotly debated regarding the timing, the number, the scale and the actual influences to vertebrate evolution, a consequence of the WGDs also seems to tell us about what the last common ancestor of vertebrates looked like. The basic information that we use to infer the ancestral phenotype is tissue-specific gene regulation that originated before the WGDs. Provided that an ohnolog pair shares overlapping tissue-specific gene regulation, the ancestral gene should have acquired the gene regulation at the time point of the WGD at the latest.

This estimation can be applied only to ohnologs, but not other duplicated genes, for the following reasons. Most importantly, we can unambiguously pinpoint the timing of gene duplication in WGD but not in SSD because WGDs are rare events in chordate evolution. In addition, as far as genes in a tandem cluster are concerned, because they are sometimes coordinately regulated by shared enhancers [37, 38], similar regulation of tandem duplicates can be acquired easily even after gene duplication. On the contrary, the similarity in transcriptional regulation between ohnologs is most likely explained by vestiges of the WGDs. These special properties of ohnologs offer a unique condition for ancestral reconstruction. Below, one model case of an ohnolog-based conjecture is presented.

Tbx5 and *Tbx4* are a pair of ohnologs [39] (WGD-derived duplicates are usually determined by means of identifying duplicated gene arrays spanning large genomic regions—'conserved synteny' [40]), which are differentially expressed in forelimb and hindlimb buds, respectively [41]. This expression pattern is controlled by several forelimb- and hindlimb-specific enhancers located in different regions of *Tbx5* and *Tbx4* [42–44]. The spatial segregation of these ohnologs drove researchers to infer the association to the origin of limbs/fins [45–47]. A hypothetical scenario of the subfunctionlization event between *Tbx4* and *Tbx5* is that when they were still a single ancestral gene, *Tbx4/5*, the gene acquired several enhancers specific to the single pair of ancestral fins, which later originated forelimbs and hindlimbs. Subsequently, after the WGDs, stepwise subfunctionalizations of the limb/fin enhancers resulted in the segregation of *Tbx4* and *Tbx5* expression between pectoral and pelvic fins (forelimbs and hindlimbs), though it remains uncertain whether the emergence of the two pairs of fins was the cause or consequence of the subfunctionalization (see [45] for more details).

While we simply revisited a hypothesis conceived decades ago, an intriguing possibility can be raised by fitting the hypothesis with both the timing of the WGDs and the fossil record. Namely, if the 2R-WGDs occurred before the split between cyclostome and jawed vertebrate lineages, the single pair of fins with *Tbx4/5* expression likely existed already in the last common ancestor of all extant vertebrates (Figure 1). On the other hand, in the fossil record, early vertebrates before the cyclostome-jawed vertebrate divergence, such as *Metaspriggina* and *Haikouichthys*, did not have any paired fins [48–51]. Paired fins can be recognized from several early jawless vertebrates, such as euphaneropids, anaspids and thelodonts [1, 52]. However, owing to the considerable diversity of their morphologies, several researchers suggested that the paired fins of euphaneropids, anaspids and thelodonts arose independently from those of jawed vertebrates, and that only ostrastracans (a group of jawless vertebrates) shared paired fin homologs with jawed vertebrate [45, 53–55]. In addition, because euphaneropids and anaspids were once regarded as being ancestral to lampreys, lampreys were thought to have lost anaspid-type paired fins [52]. However, their phylogenetic relations, particularly of euphaneropids, are still oscillating between the stem lineages of cyclostomes and jawed vertebrates [12, 52, 54, 56]. In this confusion, recent evolutionary developmental studies often assume that the lack of paired fins in the extant cyclostomes reflects the ancestral condition of vertebrates [44, 57–59]. However, if the first paired fins with *Tbx4/5* expression existed in the last common ancestor of all extant vertebrates, it is possible that the paired fins of all jawless and jawed vertebrates are homologous to each other. Furthermore, under this hypothetical scenario, the lack of paired fins in lampreys and hagfishes is interpreted as a result of secondary loss in their unique lineages. This example highlights that transcriptional regulation of ohnologs contains clues for ancestral phenotypic reconstruction.

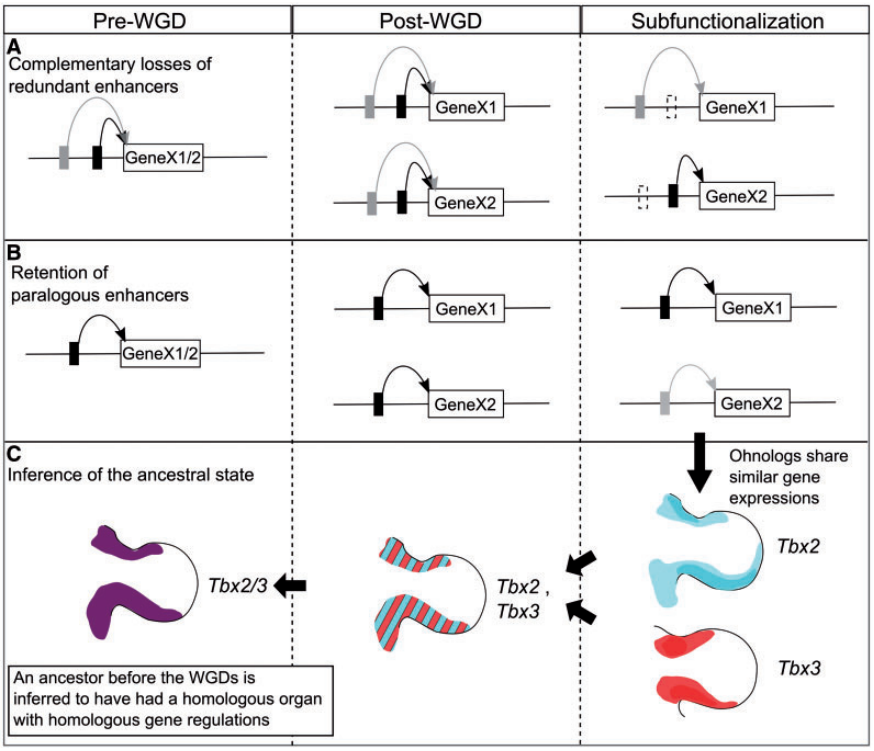

Figure 2. The basic concept of the CHROF conjecture. (**A** and **B**) Subfunctionalization of enhancers after WGDs. Rectangles, arbitrary gene Xs. Small dark and gray boxes, enhancers that are redundant, but have slightly different activities. (**A**) Redundant enhancers that existed before WGDs are lost in a complementary style, resulting in ohnologs flanked by nonhomologous enhancers with similar activities. (**B**) An enhancer is duplicated by WGDs, and retained with slightly different functions. (**C**) Inference of the ancestral state by gene expression. The arrows indicate logical flows. Here, ohnologs, *Tbx2* and *Tbx3* [39] are shown as an example. The right-most figures show the expression patterns of *Tbx2* and *Tbx3* in living tetrapod limb buds. *Tbx2* (red) and *Tbx3* (blue) are expressed in the lateral parts of limb buds, but each expression covers a slightly different region. The middle figure shows hypothetical expression patterns just after the WGDs. The spatial expression patterns are exactly same. The left figure shows a hypothetical ancestral state just before the WGDs. The ancestor had a limb/fin bud with the expression of a single gene, *Tbx2/3* (dark purple).

Chronological reconstruction of ohnolog functions

We refer to the aforementioned conjecture based on the previous conception about *Tbx4/5* as chronological reconstruction of ohnolog functions (CHROF). This conjecture can be formulated to the following two-step logic. The first step is to make an assumption that gene regulatory subfunctionalizations after WGDs are the cause of a similarity between the expression patterns of ohnologs. This assumption is basically derived from the DDC (duplication–degeneration–complementation) model [15]. Consider a single ancestral gene flanked by certain regulatory sequences such as redundant multiple enhancers or a single enhancer. Subsequently, after the WGDs, the gene regulatory sequences experienced subfunctionalization either by complementary losses of redundant enhancers (Figure 2A) or by keeping paralogous enhancers (Figure 2B). We include the former case because the latter seems a rare event [60], and there is increasing evidence that redundant enhancers are ubiquitous [61, 62]. These regulatory subfunctionalizations cause ohnologs to have similar or complementary expression patterns. In the second step, an ancestral state is inferred by reconstructing the chronological events of subfunctionalized genes. Namely, if a group of ohnologs has complementary or similar gene expression patterns in a specific organ of one or several extant species, the organ with such spatial gene regulation can be traced back to the era before the WGD event that gave rise to those ohnologs (Figure 2C; note that this logic by itself cannot predict morphological changes, e.g., limbs were fins, because of using partial information of the gene circuits of morphogenesis). While this idea itself is not new [45], the recent increase in confidence of the timing of the 2R-WGDs and phylogenetic information provides a clearer picture of the ancestral phenotype of all vertebrates.

Applications of the reconstruction approach

We apply the CHROF conjecture to *Pax2*, *Pax5* and *Pax8* genes, which were shown to have triplicated in the 2R-WGD and studied in diverse vertebrates [63–65]. In mouse embryos, all of these genes are expressed in the midbrain–hindbrain boundary, and *Pax2* and *Pax8* share the expression in the pronephros [66–68]. By following the CHROF conjecture, these expression patterns suggest that the last common ancestor of all extant vertebrates was equipped with homologs of the midbrain–hindbrain boundary and the pronephros with *Pax2/5/8* expression. Indeed, lampreys also have an equivalent *Pax2* expression in these sites [69]. In addition, the expression patterns of *Pax2*, *Pax5* and *Pax8* genes seem to be comparable with that of amphioxus *Pax2/5/8* (a nonduplicated homolog of *Pax2*, *Pax5* and *Pax8*) [70]. Interestingly, *Pax2* and *Pax8* neighbor paralogous enhancers that are thought to have duplicated during the WGDs [71]. Both of the paralogous enhancers of *Pax2* and *Pax8* can drive expression in the pronephros and conditionally in the midbrain–hindbrain boundary of *Xenopus* embryos [71], demonstrating an example of Figure 2B. Another interesting aspect of these genes is that

Table 1. Applications of the CHROF conjecture

Organ/tissue	Ohnolog	Evidence for ohnology	Description	Confirmed by independent evidence?	References
MHB, Pronephros	*Pax2/Pax5/Pax8*	See the main text	See the main text	Yes	See the main text
Rhombomere	*Hoxa3/Hoxb3/Hoxd3*, *Hoxa4/Hoxb4/Hoxd4*	[88]	Shared expression boundaries between *Hox* ohnologs suggest that rhombomere segmentation existed before the WGDs	Yes	[3, 89, 90]
Migratory muscle precursor (MMP)	*Lbx1/Lbx2*	[91]	In mouse embryos, only *Lbx1* is expressed in the MMPs, whereas *Lbx2* is expressed in the MMPs of zebrafish embryos	Yes	[6, 92, 93]
Neural crest cells	*Sox8/Sox9/Sox10*[a]	[94]	*Sox8*, *Sox9* and *Sox10* are ohnologs, and expressed in the neural crest cells	Yes	[4, 95, 96]
Paired appendages	*Tbx4/Tbx5*	[39]	See the main text	No	See the main text
Lung	*Wnt2/Wnt2b*, *Bmp2/Bmp4*[a] *Irx1/Irx3*, *Irx2/Irx5*, *Tbx4/Tbx5*	[97] [94, 98] [99] [99] [39]	See the main text	No	See the main text
External genitalia	HoxA/D, *Tbx4/Tbx5*	[88] [39]	External genitalia-specific enhancers exist in HoxA and D clusters. However, co-option of gene regulation from paired appendages is also likely	No	[100–102]

[a]The evidence for ohnology was obtained by retrieving data from a genome-wide study [94].

mouse *Pax8* and *Xenopus Pax2* are expressed in the thyroid gland of these respective species, but orthologs of these two species do not share the expression in this organ [67, 72, 73]. According to the CHROF conjecture, an enhancer drove the expression of a single pre-WGD gene that gave rise to *Pax2* and *Pax8* in the thyroid gland, and after the duplication, its descendants were differentially lost in the respective lineages. In fact, the thyroid gland of vertebrates also has a deeper origin, as amphioxus and tunicates have its homolog called 'endostyle', which expresses the pre-WGD gene *Pax2/5/8* [64, 65]. Therefore, even complementary loss of regulation between species does not undermine the conjecture. Taken together, this example confirms the validity of our CHROF conjecture.

Next, we discuss a more challenging example that need to be confirmed—the origin of the lung. The lung is not frequently discussed in the context of paleontology because soft tissues are poorly fossilized. From the anatomical point of view, the homology between the lung and the swim bladder has long been debated since Richard Owen's era [74], and recently confirmed by molecular data [75, 76]. However, there has been uncertainty about the origin of the lung/swim bladder. As modern cartilaginous fishes do not have its explicit homolog, the lung/swim bladder is often suggested to be a novelty in the bony vertebrate lineage [76, 77]. On the other hand, two exceptional fossils suggested that a placoderm (stem jawed vertebrates) possibly had a lung-like structure [77, 78]. Can the CHROF conjecture provide an insight into the origin of the lung and swim bladder? During mouse lung development, a pair of ohnologs, *Wnt2* and *Wnt2b*, are expressed in the mesoderm of the presumptive lung field, and redundantly induce lung morphogenesis [79]. In addition, other (putative) ohnolgs, such as *Bmp2/Bmp4* and *Tbx4/Tbx5*, are also reportedly co-expressed in the lung mesoderm [41, 80, 81], while *Irx1* and *Irx3* are expressed in its endoderm [82]. On the other hand, some key genes for lung development show 'orphan' expression as follows. One is *Nkx2.1*, an early marker of the lung endoderm [83]. This gene has an ohnolog, *Nkx2.4*, but *Nkx2.4* is not expressed in the lung endoderm [84]. Another one is *Shh*, which is expressed throughout the gut endoderm (therefore, not a lung-specific marker) [85]. One of the *Shh* ohnologs, *Ihh*, shares a similar expression pattern with *Shh* except that it is not expressed in the foregut, including the lung primordium [86]. These facts suggest either that the expression of *Nkx2.1* and *Shh* in the lung was acquired after the WGDs or that expression of their ohnologs in the lung was secondarily lost. Interestingly, the loss-of-functions of these genes do not cause a full lung defect, but only the lack of branchings [83, 87], suggesting that they are not required for inducing the lung bulge. Therefore, although the branched lung is specific to mammalians, a lung-like bulge may have existed since before the WGDs. Identification of the lung-specific enhancers of the above ohnolog pairs would further clarify the possibility. In addition, because lung-like bulges in fishes are proposed to be formed by elongation of most posterior pharyngeal pouches [77], it would be helpful to examine the abovementioned gene expression in the pharyngeal regions of sharks and lampreys.

Including the above cases, we have listed some tentative applications of the CHROF conjecture in Table 1. The examples with independent evidence demonstrate the validity of the CHROF conjecture. The other cases remain to be confirmed but suggest possibilities that the last common ancestor of all extant vertebrates had more complex traits than currently thought. The conjecture introduced here would further help gain insights into the origins of other tissues/organs under debate, such as calcified bones [103, 104]. However, the CHROF conjecture embraces risks for false discoveries, which we discuss below.

Skepticism

The conjecture is still naive, and has at least three potential pitfalls. First, the conjecture assumes that in a group of ohnologs, subfunctionalization of gene regulation is more likely to occur than convergent acquisitions of complementary or redundant enhancers. This expectation may not hold, for example, if tissue-specific enhancers easily duplicate and propagate over genome sequences through transpositions as recently reported [105–107]. One way to discern the evolutionary origin of gene regulation would be orthologous sequence comparisons between species. Conservation of regulatory sequences across divergent vertebrate lineages is a sign that the gene regulatory sequences were not acquired in recent convergent events. Filtering for enhancers that have paralogs duplicated by the WGDs would also lower the false discoveries caused by convergent evolution.

Second, co-option or redeployment is also another source of false discoveries, as a new organ can emerge by redeploying an old morphogenetic gene circuit [108, 109]. Such modification of gene regulation can affect multiple downstream genes, which may include ohnolog pairs. The concrete evidence for co-option of preexisting gene regulatory circuits is still limited, but one well-studied example is recently reported: a *Drosophila melanogaster*-specific organ, the posterior lobe (external genitalia) [110]. The study shows that many enhancers that control genes expressed in the posterior lobe are also active in the developing posterior spiracle, which is a common organ in *Drosophila* species, concluding that the developmental gene network of the posterior lobe was co-opted from that of the posterior spiracle [110]. Therefore, considering enhancers that control transcriptions only in a specific organ could reject the possibility that the enhancer activities are co-opted from other old organs.

Third, the CHROF conjecture implicitly assumes some level of specificity and modularity of gene regulation. For example, consider an ohnolog pair of housekeeping genes that are ubiquitously expressed. Apparently, such widespread expression does not readily indicate that all parts of the body existed before the WGDs, but simply that their ubiquitous expression existed since before the WGDs. This sort of pitfall applies also to the *Tbx4* and *Tbx5* expression in the limb buds mentioned earlier. *Tbx4* and *Tbx5* are expressed in almost whole part of hindlimb or forelimb buds, including digits [41]. This fact does not suggest that the ancestral fins before the WGDs had digits. To discuss the origin of digits, one needs to focus on digit-specific gene expression or regulation.

Another problem in the above case studies is the arbitrary choice of genes, which results from our incomplete knowledge of gene regulation and gene phylogenies. More thorough catalogs of duplicated syntenies (e.g. [94]) should aid in grasping the whole picture of ohnolog regulation through subfunctionalization. To enhance the confidence of our conjecture, systematic analyses of enhancer functions and evolution would be required in the future.

Extension of the CHROF conjecture

So far, we have focused on inferring the ancestral morphology through spatial gene regulation, but this conjecture can be extended to inferring ancestral functions of individual molecules. As another controversial theme in vertebrate evolution, we here discuss the origins of mammalian neurohypophyseal hormones, vasopressin and oxytocin [111]. It was shown that the peptide hormone vasopressin and its orthologous peptide vasotocin in nonmammalians are recognized as ligands by different vasopressin receptors (VTRs) depending on the biological context [112]. Phylogenetic and genomic studies indicated the origins of at least a subset of those VTRs in the 2R-WGD [113, 114]. In this model case, the vasopressin ligand–receptor relationship is inferred to have existed already before the 2R-WGD, according to the CHROF conjecture. Consistently, the lampreys possess the orthologs of vasopressin and its multiple receptors [24, 115, 116].

In contrast, the receptor of the other neurohypophyseal hormone oxytocin, namely, OTR, is known as the single form, and its taxonomic distribution is confined to jawed vertebrates [24, 114]. In this situation, the CHROF conjecture cannot be used, and thus the origin of the oxytocin binding capacity cannot be traced back as anciently as the 2R-WGDs. Again, the fact that the OTR orthologs are confined to jawed vertebrates is consistent with the absence of the oxytocin ortholog in the lamprey genome [116]. In the species with both the oxytocin and vasopressin orthologs, these two ligand genes are found as a tandem cluster in the genome [116]. As these genes are thought to have duplicated in a SSD, its timing should not coincide with that of the abovementioned duplications between the VTR paralogs. Obviously, one cannot apply the CHROF conjecture to these ligand genes, which are not WGD-derived duplicates.

It should be noted that the CHROF conjecture would be qualified to infer ancestral molecular functions but not its physiological output, such as hormonal roles in social and reproductive behaviors, in the case of the vasopressin hormone and receptors discussed above. The CHROF conjecture can further be exerted for molecules with more diverse functions including enzymatic capacity. Again, the possibility that ohnologs secondarily acquired their shared functional capacities independently after the WGD should always be carefully assessed before drawing any conclusion.

Conclusion

This article has introduced a conjecture to infer the ancestral phenotype of all extant vertebrates by reconstructing subfunctionalized ohnologs, which is based on the DDC model [15]. While the idea has been around for decades [45], we have reviewed and formulated the conjecture in the light of modern knowledge of molecular phylogenetics, genome evolution and developmental regulation. Among the model cases presented, the triplication of the *Pax2/5/8* genes serves as a proof of principle confirmed by independent evidence of pre-WGD gene functions in invertebrate chordates. One controversial hypothesis drawn from the application of the CHROF conjecture would be the origin of paired fins before the WGD, which remains to be carefully examined. One advantage of this conjecture is the independence from out-group comparisons, which could compensate traditional ancestral

reconstruction approaches based on phylogenetic character mapping. Moreover, information of gene expression or regulation for a single species can basically allow an inference based on this conjecture, although it would be more reliable to resort to information from diverse species. Finally, we hope that the CHROF conjecture would help unveil earlier origins of some traits previously regarded as novelties in the jawed vertebrate lineage, and remove unnecessary prejudice that the common ancestor of vertebrates was simple and primitive.

Key Points

- Growing evidence supports the monophyly of cyclostomes and two rounds of WGDs as a genomic synapomorphy of all extant vertebrates.
- Functions shared between paralogs expanded in WGDs help reconstruct phenotypes of the vertebrate ancestor, and this conjecture is termed CHROF.
- The CHROF conjecture hypothesizes the antiquity of some phenotypic characters of vertebrate ancestors, such as paired fins.

Acknowledgements

The authors acknowledge Kazuaki Yamaguchi, Per E. Ahlberg, Dan Larhammar, Leif Andersson and Lars G. Lundin for insightful discussions. The authors' gratitude extends to the researchers who previously presented the seminal ideas of the ancestral reconstruction approach formulated in this article.

Funding

This study is supported by a research grant from MEXT to the RIKEN Center for Life Science Technologies, JSPS KAKENHI Grant Number 17K07426 to S.K. and 17K15132 to K.O.

References

1. Janvier P. *Early Vertebrates*. New York: Oxford University Press, 1996.
2. Donoghue P. Evolution: divining the nature of the ancestral vertebrate. *Curr Biol* 2017;**27**(7):R277–R9.
3. Parker HJ, Bronner ME, Krumlauf R. A Hox regulatory network of hindbrain segmentation is conserved to the base of vertebrates. *Nature* 2014;**514**(7523):490–3.
4. Ota KG, Kuraku S, Kuratani S. Hagfish embryology with reference to the evolution of the neural crest. *Nature* 2007; **446**(7136):672–5.
5. Sugahara F, Pascual-Anaya J, Oisi Y, *et al*. Evidence from cyclostomes for complex regionalization of the ancestral vertebrate brain. *Nature* 2016;**531**(7592):97–100.
6. Okamoto E, Kusakabe R, Kuraku S, *et al*. Migratory appendicular muscles precursor cells in the common ancestor to all vertebrates. *Nat Ecol Evol* 2017;**1**(11):1731–6.
7. Shimeld SM, Donoghue PCJ. Evolutionary crossroads in developmental biology: cyclostomes (lamprey and hagfish). *Development* 2012;**139**(12):2091–9.
8. Janvier P. Early jawless vertebrates and cyclostome origins. *Zoolog Sci* 2008;**25**(10):1045–56.
9. Kuraku S, Meyer A, Kuratani S. Timing of genome duplications relative to the origin of the vertebrates: did cyclostomes diverge before or after?. *Mol Biol Evol* 2009;**26**(1):47–59.
10. Near TJ. Conflict and resolution between phylogenies inferred from molecular and phenotypic data sets for hagfish, lampreys, and gnathostomes. *J Exp Zool Part B Mol Dev Evol* 2009;**312B**(7):749–61.
11. Goudemand N, Orchard MJ, Urdy S, *et al*. Synchrotron-aided reconstruction of the conodont feeding apparatus and implications for the mouth of the first vertebrates. *Proc Natl Acad Sci USA* 2011;**108**(21):8720–4.
12. Janvier P. Facts and fancies about early fossil chordates and vertebrates. *Nature* 2015;**520**(7548):483–9.
13. Byrne KP, Wolfe KH. The Yeast Gene Order Browser: combining curated homology and syntenic context reveals gene fate in polyploid species. *Genome Res* 2005;**15**(10):1456–61.
14. Van de Peer Y, Maere S, Meyer A. The evolutionary significance of ancient genome duplications. *Nat Rev Genet* 2009; **10**(10):725–32.
15. Force A, Lynch M, Pickett FB, *et al*. Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 1999;**151**(4):1531–45.
16. Larhammar D, Sundström G, Dreborg S, *et al*. Major genomic events and their consequences for vertebrate evolution and endocrinology. *Ann N Y Acad Sci* 2009;**1163**:201–8.
17. Dehal P, Boore JL. Two rounds of whole genome duplication in the ancestral vertebrate. *PLoS Biol* 2005;**3**(10):e314.
18. Vandepoele K, De Vos W, Taylor JS, *et al*. Major events in the genome evolution of vertebrates: paranome age and size differ considerably between ray-finned fishes and land vertebrates. *Proc Natl Acad Sci USA* 2004;**101**(6): 1638–43.
19. Smith JJ, Keinath MC. The sea lamprey meiotic map improves resolution of ancient vertebrate genome duplications. *Genome Res* 2015;**25**(8):1081–90.
20. Abbasi AA. Piecemeal or big bangs: correlating the vertebrate evolution with proposed models of gene expansion events. *Nat Rev Genet* 2010;**11**(2):166.
21. Furlong RF, Holland PWH. Bayesian phylogenetic analysis supports monophyly of ambulacraria and of cyclostomes. *Zoolog Sci* 2002;**19**(5):593–9.
22. Escriva H, Manzon L, Youson J, *et al*. Analysis of lamprey and hagfish genes reveals a complex history of gene duplications during early vertebrate evolution. *Mol Biol Evol* 2002;**19**(9): 1440–50.
23. Campanini EB, Vandewege MW, *et al*. Early evolution of vertebrate Mybs: an integrative perspective combining synteny, phylogenetic, and gene expression analyses. *Genome Biol Evol* 2015;**7**(11):3009–21.
24. Lagman D, Ocampo Daza D, Widmark J, *et al*. The vertebrate ancestral repertoire of visual opsins, transducin alpha subunits and oxytocin/vasopressin receptors was established by duplication of their shared genomic region in the two rounds of early vertebrate genome duplications. *BMC Evol Biol* 2013;**13**(1):238.
25. Gutierrez-Mazariegos J, Kumar E, Studer RA, *et al*. Evolutionary diversification of retinoic acid receptor ligand-binding pocket structure by molecular tinkering. *R Soc Open Sci* 2016;**3**(3):150484.
26. Qiu H, Hildebrand F, Kuraku S, *et al*. Unresolved orthology and peculiar coding sequence properties of lamprey genes: the KCNA gene family as test case. *BMC Genomics* 2011;**12**(1):325.
27. Noro M, Sugahara F, Kuraku S. Reevaluating *Emx* gene phylogeny: homopolymeric amino acid tracts as a potential factor obscuring orthology signals in cyclostome genes. *BMC Evol Biol* 2015;**15**(1):78.

28. Tank EM, Dekker RG, Beauchamp K, *et al*. Patterns and consequences of vertebrate Emx gene duplications. *Evol Dev* 2009;**11**(4):343–53.
29. Manousaki T, Qiu H, Noro M, *et al*. Molecular evolution in the lamprey genomes and its relevance to the timing of whole genome duplications. In: Orlov, A and Beamish, R (ed). *Jawless Fishes of the World*, Vol. 1. Newcastle upon Tyne: Cambridge Scholars Publishing, 2011, 2–16.
30. Smith JJ, Kuraku S, Holt C, *et al*. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nat Genet* 2013;**45**:415–21, 421e1–2.
31. Robertson FM, Gundappa MK, Grammes F, *et al*. Lineage-specific rediploidization is a mechanism to explain time-lags between genome duplication and evolutionary diversification. *Genome Biol* 2017;**18**(1):111.
32. Kuraku S. Palaeophylogenomics of the vertebrate ancestor—impact of hidden paralogy on hagfish and lamprey gene phylogeny. *Integr Comp Biol* 2010;**50**(1):124–9.
33. Kuraku S. Impact of asymmetric gene repertoire between cyclostomes and gnathostomes. *Semin Cell Dev Biol* 2013; **24**(2):119–27.
34. Mehta TK, Ravi V, Yamasaki S, *et al*. Evidence for at least six Hox clusters in the Japanese lamprey (*Lethenteron japonicum*). *Proc Natl Acad Sci USA* 2013;**110**(40):16044–9.
35. Fried C, Prohaska SJ, Stadler PF. Independent Hox-cluster duplications in lampreys. *J Exp Zool B Mol Dev Evol* 2003; **299**(1):18–25.
36. Stadler PF, Fried C, Prohaska SJ, *et al*. Evidence for independent *Hox* gene duplications in the hagfish lineage: a PCR-based gene inventory of *Eptatretus stoutii*. *Mol Phylogenet Evol* 2004;**32**:686–94.
37. Spitz F, Gonzalez F, Duboule D. A global control region defines a chromosomal regulatory landscape containing the HoxD cluster. *Cell* 2003;**113**(3):405–17.
38. Moriyama Y, Kawanishi T, Nakamura R, *et al*. The medaka *zic1/zic4* mutant provides molecular insights into teleost caudal fin evolution. *Curr Biol* 2012;**22**(7):601–7.
39. Ruvinsky I, Silver LM. Newly identified paralogous groups on mouse chromosomes 5 and 11 reveal the age of a T-Box cluster duplication. *Genomics* 1997;**40**(2):262–6.
40. Kuraku S, Meyer A. Detection and phylogenetic assessment of conserved synteny derived from whole genome duplications. *Methods Mol Biol* 2012;**855**:385–95.
41. Gibson-Brown JJ, I. Agulnik S, Silver LM, *et al*. Expression of T-box genes *Tbx2-Tbx5* during chick organogenesis. *Mech Dev* 1998;**74**:165–9.
42. Menke DB, Guenther C, Kingsley DM. Dual hindlimb control elements in the *Tbx4* gene and region-specific control of bone size in vertebrate limbs. *Development* 2008;**135**(15): 2543–53.
43. Minguillon C, Nishimoto S, Wood S, *et al*. Hox genes regulate the onset of *Tbx5* expression in the forelimb. *Development* 2012;**139**(17):3180–8.
44. Adachi N, Robinson M, Goolsbee A, *et al*. Regulatory evolution of *Tbx5* and the origin of paired appendages. *Proc Natl Acad Sci USA* 2016;**113**(36):10115–20.
45. Ruvinsky I, Gibson-Brown JJ. Genetic and developmental bases of serial homology in vertebrate limb evolution. *Development* 2000;**127**(24):5233–44.
46. Horton AC, Mahadevan NR, Minguillon C, *et al*. Conservation of linkage and evolution of developmental function within the *Tbx2/3/4/5* subfamily of T-box genes: implications for the origin of vertebrate limbs. *Dev Genes Evol* 2008;**218**(11-12): 613–28.
47. Minguillon C, Gibson-Brown JJ, Logan MP. *Tbx4/5* gene duplication and the origin of vertebrate paired appendages. *Proc Natl Acad Sci USA* 2009;**106**(51):21726–30.
48. Shu D-G, Morris SC, Han J, *et al*. Head and backbone of the early Cambrian vertebrate *Haikouichthys*. *Nature* 2003; **421**(6922):526–9.
49. Morris SC, Caron J-B. A primitive fish from the Cambrian of North America. *Nature* 2014;**512**(7515):419–22.
50. Zhang XG, Hou XG. Evidence for a single median fin-fold and tail in the Lower Cambrian vertebrate, *Haikouichthys ercaicunensis*. *J Evol Biol* 2004;**17**:1162–6.
51. Shu D-G, Luo H-L, Conway Morris S, *et al*. Lower Cambrian vertebrates from south China. *Nature* 1999; **402**(6757):42–6.
52. Janvier P, Arsenault M. The anatomy of *Euphanerops longaevus* Woodward, 1900, an anaspid-like jawless vertebrate from the Upper Devonian of Miguasha, Quebec, Canada. *Geodiversitas* 2007;**1**:143–216.
53. Sansom RS, Gabbott SE, Purnell MA. Unusual anal fin in a Devonian jawless vertebrate reveals complex origins of paired appendages. *Biol Lett* 2013;**9**:20130002.
54. Sansom RS, Freedman K, Gabbott SE, *et al*. Taphonomy and affinity of an enigmatic Silurian vertebrate, *Jamoytius kerwoodi* White. *Palaeontology* 2010;**53**(6):1393–409.
55. Coates M. The evolution of paired fins. *Theory Biosci* 2003; **122**(2-3):266–87.
56. Blom H. New birkeniid anaspid from the Lower Devonian of Scotland and its phylogenetic implications. *Palaeontology* 2012;**55**(3):641–52.
57. Onimaru K, Shoguchi E, Kuratani S, *et al*. Development and evolution of the lateral plate mesoderm: comparative analysis of amphioxus and lamprey with implications for the acquisition of paired fins. *Dev Biol* 2011;**359**(1):124–36.
58. Tulenko FJ, McCauley DW, Mackenzie EL, *et al*. Body wall development in lamprey and a new perspective on the origin of vertebrate paired fins. *Proc Natl Acad Sci USA* 2013;**110**(29): 11899–904.
59. Freitas R, Zhang G, Cohn MJ. Evidence that mechanisms of fin development evolved in the midline of early vertebrates. *Nature* 2006;**442**(7106):1033–7.
60. Matsunami M, Saitou N. Vertebrate paralogous conserved noncoding sequences may be related to gene expressions in brain. *Genome Biol Evol* 2013;**5**:140–50.
61. Barolo S. Shadow enhancers: frequently asked questions about distributed cis-regulatory information and enhancer redundancy. *Bioessays* 2012;**34**(2):135–41.
62. Cannavò E, Khoueiry P, Garfield DA, *et al*. Shadow enhancers are pervasive features of developmental regulatory networks. *Curr Biol* 2016;**26**:38–51.
63. Goode DK, Elgar G. The *PAX258* gene subfamily: a comparative perspective. *Dev Dyn* 2009;**238**(12):2951–74.
64. Bassham S, Cañestro C, Postlethwait JH. Evolution of developmental roles of Pax2/5/8 paralogs after independent duplication in urochordate and vertebrate lineages. *BMC Biol* 2008;**6**(1):35.
65. Kozmik Z, Holland ND, Kalousova A, *et al*. Characterization of an amphioxus paired box gene, *AmphiPax2/5/8*: developmental expression patterns in optic support cells, nephridium, thyroid-like structures and pharyngeal gill slits, but not in the midbrain-hindbrain boundary region. *Development* 1999;**126**:1295–304.
66. Nornes HO, Dressler GR, Knapik EW, *et al*. Spatially and temporally restricted expression of *Pax2* during murine neurogenesis. *Development* 1990;**109**(4):797–809.

67. Plachov D, Chowdhury K, Walther C, *et al*. *Pax8*, a murine paired box gene expressed in the developing excretory system and thyroid gland. *Development* 1990;**110**(2):643–51.
68. Adams B, Dorfler P, Aguzzi A, *et al*. *Pax-5* encodes the transcription factor BSAP and is expressed in B lymphocytes, the developing CNS, and adult testis. *Genes Dev* 1992;**6**(9): 1589–607.
69. McCauley DW, Bronner-Fraser M. Conservation of Pax gene expression in ectodermal placodes of the lamprey. *Gene* 2002;**287**(1–2):129–39.
70. Holland LZ. Evolution of new characters after whole genome duplications: insights from amphioxus. *Semin Cell Dev Biol* 2013;**24**(2):101–9.
71. Ochi H, Tamai T, Nagano H, *et al*. Evolution of a tissue-specific silencer underlies divergence in the expression of *pax2* and *pax8* paralogues. *Nat Commun* 2012;**3**:848.
72. Heller N, Brändli AW. *Xenopus Pax-2/5/8* orthologues: novel insights into Pax gene evolution and identification of *Pax-8* as the earliest marker for otic and pronephric cell lineages. *Dev Genet* 1999;**24**(3–4):208–19.
73. Wendl T, Lun K, Mione M, *et al*. *Pax2.1* is required for the development of thyroid follicles in zebrafish. *Development* 2002; **129**(15):3751–60.
74. Owen R. Lectures on *The Comparative Anatomy and Physiology of the Vertebrate Animals, Delivered at the Royal College of Surgeons of England, in 1844 and 1846*. London: Printed for Longman, Brown, Green, and Longmans, 1846.
75. Sagai T, Amano T, Maeno A, *et al*. Evolution of *Shh* endoderm enhancers during morphological transition from ventral lungs to dorsal gas bladder. *Nat Commun* 2017;**8**: 14300.
76. Tatsumi N, Kobayashi R, Yano T, *et al*. Molecular developmental mechanism in polypterid fish provides insight into the origin of vertebrate lungs. *Sci Rep* 2016;**6**:30580.
77. Perry SF, Wilson RJA, Straus C, *et al*. Which came first, the lung or the breath? *Comp Biochem Physiol A Mol Integr Physiol* 2001;**129**:37–47.
78. Arsenault M, Desbiens S, Janvier P, *et al*. New data on the soft tissues and external morphology of the antiarch *Bothriolepis canadensis* (Whiteaves, 1880), from the Upper Devonian of Miguasha, Quebec. In: Arratia G, Wilson MVH, Cloutier R (eds). *Recent Advances in the Origin and Early Radiation of Vertebrates*. Munich: Verlag Dr. Friedrich Pfeil, 2004, 439–54.
79. Goss AM, Tian Y, Tsukiyama T, *et al*. Wnt2/2b and β-catenin signaling are necessary and sufficient to specify lung progenitors in the foregut. *Dev Cell* 2009;**17**(2):290–8.
80. Sakiyama J, Yokouchi Y, Kuroiwa A. Coordinated expression of *Hoxb* genes and signaling molecules during development of the chick respiratory tract. *Dev Biol* 2000;**227**(1):12–27.
81. Chapman DL, Garvey N, Hancock S, *et al*. Expression of the T-box family genes, *Tbx1-Tbx5*, during early mouse development. *Dev Dyn* 1996;**206**(4):379–90.
82. Houweling C, Dildrop R, Peters T, *et al*. Gene and cluster-specific expression of the Iroquois family members during mouse development. *Mech Dev* 2001;**107**:169–74.
83. Minoo P, Su G, Drum H, *et al*. Defects in tracheoesophageal and lung morphogenesis in *Nkx2.1*(-/-) mouse embryos. *Dev Biol* 1999;**209**(1):60–71.
84. Small EM, Vokes SA, Garriock RJ, *et al*. Developmental expression of the *Xenopus Nkx2-1* and *Nkx2-4* genes. *Mech Dev* 2000;**96**(2):259–62.
85. Bitgood MJ, McMahon AP. Hedgehog and Bmp genes are coexpressed at many diverse sites of cell- cell interaction in the mouse embryo. *Dev Biol* 1995;**172**(1):126–38.
86. Mao J, Kim B-M, Rajurkar M, *et al*. Hedgehog signaling controls mesenchymal growth in the developing mammalian digestive tract. *Development* 2010;**137**(10):1721–9.
87. Yuan B, Li C, Kimura S, *et al*. Inhibition of distal lung morphogenesis in *Nkx2.1*(-/-) embryos. *Dev Dyn* 2000;**217**(2): 180–90.
88. Larhammar D, Lundin LG, Hallböök F. The human Hox-bearing chromosome regions did arise by block or chromosome (or even genome) duplications. *Genome Res* 2002; **12**(12):1910–20.
89. Murakami Y, Pasqualetti M, Takio Y, *et al*. Segmental development of reticulospinal and branchiomotor neurons in lamprey: insights into the evolution of the vertebrate hindbrain. *Development* 2004;**131**(5):983–95.
90. Takio Y, Pasqualetti M, Kuraku S, *et al*. Evolutionary biology: lamprey *Hox* genes and the evolution of jaws. *Nature* 2004; **429**(6989):262.
91. Wotton KR, Weierud FK, Dietrich S, *et al*. Comparative genomics of Lbx loci reveals conservation of identical Lbx ohnologs in bony vertebrates. *BMC Evol Biol* 2008;**8**:171.
92. Kusakabe R, Kuraku S, Kuratani S. Expression and interaction of muscle-related genes in the lamprey imply the evolutionary scenario for vertebrate skeletal muscle, in association with the acquisition of the neck and fins. *Dev Biol* 2011;**350**(1):217–27.
93. Ochi H, Westerfield M. *Lbx2* regulates formation of myofibrils. *BMC Dev Biol* 2009;**9**:13.
94. Singh PP, Arora J, Isambert H. Identification of ohnolog genes originating from whole genome duplication in early vertebrates, based on synteny comparison across multiple genomes. *PLoS Comput Biol* 2015;**11**(7):e1004394.
95. Green SA, Simoes-Costa M, Bronner ME. Evolution of vertebrates as viewed from the crest. *Nature* 2015;**520**(7548):474–82.
96. Hong CS, Saint-Jeannet JP. Sox proteins and neural crest development. *Semin Cell Dev Biol* 2005;**16**(6):694–703.
97. Garriock RJ, Warkman AS, Meadows SM, *et al*. Census of vertebrate Wnt genes: isolation and developmental expression of *Xenopus Wnt2*, *Wnt3*, *Wnt9a*, *Wnt9b*, *Wnt10a*, and *Wnt16*. *Dev Dyn* 2007;**236**(5):1249–58.
98. Feiner N, Begemann G, Renz AJ, *et al*. The origin of *bmp16*, a novel Bmp2/4 relative, retained in teleost fish genomes. *BMC Evol Biol* 2009;**9**:277.
99. Peters T, Dildrop R, Ausmeier K, *et al*. Organization of mouse *Iroquois* homeobox genes in two clusters suggests a conserved regulation and function in vertebrate development. *Genome Res* 2000;**10**(10):1453–62.
100. Long JA, Mark-Kurik E, Johanson Z, *et al*. Copulation in antiarch placoderms and the origin of gnathostome internal fertilization. *Nature* 2015;**517**(7533):196–9.
101. Lonfat N, Montavon T, Darbellay F, *et al*. Convergent evolution of complex regulatory landscapes and pleiotropy at *Hox* loci. *Science* 2014;**346**(6212):1004–6.
102. Douglas NC, Heng K, Sauer MV, *et al*. Dynamic expression of Tbx2 subfamily genes in development of the mouse reproductive system. *Dev Dyn* 2012;**241**(2):365–75.
103. Venkatesh B, Lee AP, Ravi V, *et al*. Elephant shark genome provides unique insights into gnathostome evolution. *Nature* 2014;**505**(7482):174–9.
104. Ryll B, Sanchez S, Haitina T, *et al*. The genome of *Callorhinchus* and the fossil record: a new perspective on *SCPP* gene evolution in gnathostomes. *Evol Dev* 2014;**16**(3):123–4.
105. Lynch VJ, Leclerc RD, May G, *et al*. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nat Genet* 2011;**43**(11):1154–9.

106. Nishihara H, Kobayashi N, Kimura-Yoshida C, *et al*. Coordinately co-opted multiple transposable elements constitute an enhancer for *wnt5a* expression in the mammalian secondary palate. *PLoS Genet* 2016;**12**(10):e1006380.
107. Xie M, Hong C, Zhang B, *et al*. DNA hypomethylation within specific transposable element families associates with tissue-specific enhancer landscape. *Nat Genet* 2013;**45**(7):836–41.
108. Shubin N, Tabin C, Carroll S. Deep homology and the origins of evolutionary novelty. *Nature* 2009;**457**(7231):818–23.
109. Peter IS, Davidson EH. Evolution of gene regulatory networks controlling body plan development. *Cell* 2011;**144**(6):970–85.
110. Glassford WJ, Johnson WC, Dall NR, *et al*. Co-option of an ancestral Hox-regulated network underlies a recently evolved morphological novelty. *Dev Cell* 2015;**34**(5):520–31.
111. Acher R. Molecular evolution of fish neurohypophysial hormones: neutral and selective evolutionary mechanisms. *Gen Comp Endocrinol* 1996;**102**(2):157–72.
112. Thibonnier M, Berti-Mattera LN, Dulin N, *et al*. Signal transduction pathways of the human V1-vascular, V2-renal, V3-pituitary vasopressin and oxytocin receptors. *Prog Brain Res* 1999;**119**:147–61.
113. Ocampo Daza D, Lewicka M, Larhammar D. The oxytocin/vasopressin receptor family has at least five members in the gnathostome lineage, inclucing two distinct V2 subtypes. *Gen Comp Endocrinol* 2012;**175**(1):135–43.
114. Yamaguchi Y, Kaiya H, Konno N, *et al*. The fifth neurohypophysial hormone receptor is structurally related to the V2-type receptor but functionally similar to V1-type receptors. *Gen Comp Endocrinol* 2012;**178**:519–28.
115. Mayasich SA, Clarke BL. The emergence of the vasopressin and oxytocin hormone receptor gene family lineage: clues from the characterization of vasotocin receptors in the sea lamprey (*Petromyzon marinus*). *Gen Comp Endocrinol* 2016;**226**:88–101.
116. Gwee P, Tay B, Brenner S, *et al*. Characterization of the neurohypophysial hormone gene loci in elephant shark and the Japanese lamprey: origin of the vertebrate neurohypophysial hormone genes. *BMC Evol Biol* 2009;**9**(1):47.