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 unresolved 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
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 ap-
proach to inferring the ancestral phenotype of vertebrates,
which does not heavily rely on either character mapping vul-
nerable to secondary loss or out-group comparison. As a prem-
ise, 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 phylo-
genetic 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 pro-
duce 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, how-
ever, a wealth of molecular data has accumulated, which sup-
ported 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 pheno-
typic characters.

When in early vertebrate evolution did the
alleged genome expansion occur?
Gene duplication is known as a universal mechanism that gen-
erates an additional gene copy and is divided into two major
types, small-scale duplication (SSD) and whole-genome dupli-
cation (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 inter-
mittently 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 des-
cendant 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 subfunctionali-
zation in which the original function is divided by the dupli-
cates, 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 de-
termine 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 con-
trolled systematically with probabilistic phylogeny inference
and its statistical evaluation [9]. This analysis led to the first
proposal of the evolutionary scenario that the cyclostome lin-
eage 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 Enx 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. Providing 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 subfunctionalization 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 euphancropods, anaspids and thelodonts [1, 52]. However, owing to the considerable diversity of their morphologies, several researchers suggested that the paired fins of euphancropods, anaspids and thelodonts arose independently from those of jawed vertebrates, and that only ostracodans (a group of jawless vertebrates) shared paired fin homologs with jawed vertebrate [45, 53–55]. In addition, because euphancropods 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 euphancropods, 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.
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]. Interestingly, Pax2 and Pax8 genes seem to be comparable with that of amphiesmos 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

![Figure 2](https://example.com/figure2.png)
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 also likely have its homolog called ‘endostyle’, which expresses the pre-WGD gene. From the anatomical point of view, the thyroid gland of tetrapods is specific to mammalians, a lung-like bulge may have existed after the WGDs or that expression of their ohnologs in the lung was acquired in the lung 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.

Table 1. Applications of the CHROF conjecture

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

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


113. Ocampo Daza D, Lewicka M, Larhammar D. The oxytocin/vasopressin receptor family has at least five members in the gnathostome lineage, including two distinct V2 subtypes. *Gen Comp Endocrinol* 2012;175(1):135–43.

