

Research article

Evolution of Nodal and Nodal-related genes and the putative composition of the heterodimer that triggers the Nodal pathway in vertebrates

Juan C. Opazo¹, Shigehiro Kuraku², Kattina Zavala¹, Jessica Toloza-Villalobos¹ & Federico G. Hoffmann^{3,4}

¹Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile. ²Laboratory for Phyloinformatics, RIKEN Center for Biosystems Dynamics Research (BDR), Kobe, Japan. ³Department of Biochemistry, Molecular Biology, Entomology, and Plant Pathology, Mississippi State University.

⁴Institute for Genomics, Biocomputing and Biotechnology, Mississippi State University

Running title: Evolution of nodal genes in vertebrates

Corresponding author:

Juan C. Opazo
Instituto de Ciencias Ambientales y Evolutivas
Facultad de Ciencias
Universidad Austral de Chile
Valdivia, Chile
jopazo@gmail.com
+56632221674

Abstract

Nodal is a signaling molecule that belongs to the transforming growth factor-beta superfamily that plays key roles during the early stages of development of animals. Nodal forms an heterodimer with a GDF1/3 protein to activate the Nodal pathway. Vertebrates have a paralog of nodal in their genomes labeled Nodal related, but the evolutionary history of these genes is a matter of debate, mainly because of variable numbers of genes in the vertebrate genomes sequenced so far. Thus, the goal of this study was to investigate the evolutionary history of the Nodal and Nodal-related genes with an emphasis in tracking changes in the number of genes among vertebrates. Our results show the presence of two gene lineages (Nodal and Nodal-related) that can be traced back to the ancestor of jawed vertebrates. These lineages have undergone processes of differential retention and lineage-specific expansions. Our results imply that Nodal and Nodal-related duplicated at the latest in the ancestor of gnathostomes, and they still retain a significant level of functional redundancy. By comparing the evolution of the Nodal/Nodal-related with GDF1/3 gene family, it is possible to infer that there are at least four types of heterodimers that can trigger the Nodal pathway among vertebrates.

Keywords: gene family evolution; gene duplication; gene loss; differential retention

1 INTRODUCTION

Understanding the evolution of genes that are involved in early developmental processes represents an important challenge in biology (Gilbert, 2017). In most of the cases, the information that is available is restricted to the presence of a particular developmental pathway, and of the genes that participate, but much less is known regarding the evolutionary history of the genes and pathways (Jones et al., 1995; Levin et al., 1995; Seleiro et al., 1996; Joseph and Melton, 1997; Raya et al., 2004a; Lagadec et al., 2015; Tadjuidje et al., 2016; Li et al., 2017).

The gene *Nodal* encodes for a signaling molecule that belongs to the transforming growth factor-beta superfamily (TGF- β). It is a single-copy gene in most animal groups, but in vertebrates it has a paralog called *Nodal-related*. The *Nodal* pathway is active during the pre-gastrulation and gastrulation stages of development (Shen, 2007; Zinski et al., 2018) and is responsible for multiple fundamental processes such as mesoderm induction and patterning, endoderm function, neural patterning, establishment of left-right asymmetry, dorsal-ventral axis specification, anterior-posterior axis formation and maintenance of undifferentiated embryonic stem cells (Conlon et al., 1994; Jones et al., 1995; Feldman et al., 1998; Schier, 2003; Shen, 2007). Recently, three independent groups have shown that *Nodal* needs to form heterodimers with the GDF1/3 protein, encoded by another member of the TGF- β superfamily, to activate the signaling pathway (Bisgrove et al., 2017; Montague and Schier, 2017; PellicciaJ.L. et al., 2017). To gain insights into the evolutionary history of the *Nodal* pathway, Opazo & Zavala (2018) studied the evolutionary history of the GDF1/3 gene family and found that the vast majority of vertebrates have a single GDF1/3 ortholog in their genome. The two exceptions are amphibians and mammals, in which lineage-specific duplications gave rise the separate GDF1 and 3 genes: GDF1_A and 3_A for amphibians and GDF1_M and 3_M for mammals.

On the other hand, the evolution of its partner, the *Nodal* gene and its paralog, *nodal-related*, remains poorly understood. There is extensive variation in the number of genes in the *Nodal* gene family among vertebrates, ranging from one in mammals and some sauropsids to six in amphibians (e.g., *Xenopus laevis*). In fact, most vertebrate genomes include copies of *Nodal* and *Nodal-related*, but their phyletic distribution is not clear, and nomenclature is problematic, as different names have

been applied to orthologs, and conversely, the same name has been applied to paralogous copies. In addition, the evolutionary history of the Nodal gene and its homologs is a matter of debate mainly because the number of Nodal and Nodal-related genes lineages present in vertebrates is not well defined, and sampling differences complicate comparisons among studies (Fan and Dougan, 2007; Kuraku and Kuratani, 2011).

Accordingly, the goal of this study was to reconstruct the evolutionary history of the Nodal and Nodal-related genes in vertebrates to 1) define the number of Nodal and Nodal-related genes in vertebrates, 2) reconstruct evolutionary relationships among the vertebrate Nodal and Nodal-related to unravel the duplicative history of the genes, with special attention to amphibians and teleost fish, and 3) integrate the history of the Nodal and GDF1/3 gene families to gain insights into the evolution of the heterodimer that is responsible for triggering the Nodal pathway. Our results indicate that the duplication giving rise to Nodal and Nodal-related traces back at least to the last common ancestor of jawed vertebrates. These two genes have undergone differential retention and lineage-specific expansions, where the lack of Nodal orthologs in birds, crocodiles, turtles and snakes, as well as, the expansion of Nodal-related genes in anurans represent the most salient results. Finally, by comparing the presence and absence of the Nodal/Nodal-related and GDF1/3 gene family members in different lineages, we infer that at least four different types of heterodimers can activate the Nodal pathway among vertebrates.

2 MATERIALS AND METHODS

2.1 Sequence data

We implemented bioinformatic searches to retrieve Nodal and Nodal-related genes in mammals, birds, reptiles, amphibians, coelacanth, teleost fish, holostean fish, cartilaginous fish, cyclostomes, urochordates and cephalochordates (Supplementary Table S1). We identified pieces containing Nodal and Nodal-related genes in the Ensembl database (Zerbino et al., 2018) using BLASTN or National Center for Biotechnology Information (NCBI) database (refseq_genomes, htgs, and wgs) (Geer et al., 2010) using TBLASTN (Altschul et al., 1990) with default settings. Conserved synteny was also used to define the genomic region containing Nodal and Nodal-related genes. Once identified, genomic pieces were extracted including the 5' and 3'

flanking genes. After extraction, we curated the existing annotation by comparing known exon sequences to genomic pieces using the program Blast2seq with default parameters (Tatusova and Madden, 1999). Putatively functional genes were characterized by an open intact reading frame with the canonical exon/intron structure typical of Nodal/Nodal-related gene. Sequences derived from shorter records based on genomic DNA or cDNA were also included to attain a broad taxonomic coverage.

2.2 Phylogenetic analyses

Amino acid sequences were aligned using the L-INS-i strategy from MAFFT v.7 (Kato and Standley, 2013). We used the proposed model tool of IQ-Tree v1.6 (Trifinopoulos et al., 2016) to select the best-fitting model of amino acid substitution (JTT + F + R6). We employed the maximum-likelihood method to obtain the best tree using the program IQ-Tree v1.6 (Trifinopoulos et al., 2016); support for the nodes was assessed with 1,000 bootstrap pseudoreplicates using the ultrafast routine. Bayesian searches were conducted in MrBayes v.3.1.2 (Ronquist and Huelsenbeck, 2003). Two independent runs of six simultaneous chains for 1×10^7 generations were set, and every 1,000 generations were sampled using default priors. The run was considered to have reached convergence once the likelihood scores formed an asymptote and the average standard deviation of the split frequencies remained < 0.01 . We discarded all trees that were sampled before convergence, and we evaluated support for the nodes and parameter estimates from a majority rule consensus of the last 5,000 trees. Human and opossum growth differentiation factor 10 (GDF10) and bone morphogenetic protein 2 (BMP2) sequences were used as outgroups. Divergence times were obtained from the TimeTree server (Kumar et al., 2017).

2.3 Assessment of Conserved Synteny

We examined genes found upstream and downstream of the Nodal and Nodal-related genes of the species below mentioned to encompass diverse vertebrate lineages. For comparative purposes, we used the estimates of orthology and paralogy derived from the EnsemblCompara database (Herrero et al., 2016); these estimates are obtained from an automated pipeline that considers both synteny and phylogeny to generate orthology mappings. These predictions were visualized using

the program Genomicus v94.01 (Louis et al., 2015). Our assessments were performed in humans (*Homo sapiens*), chicken (*Gallus gallus*), American alligator (*Alligator mississippiensis*), Chinese softshell turtle (*Pelodiscus sinensis*), anole lizard (*Anolis carolinensis*), Western clawed frog (*Xenopus tropicalis*), axolotl (*Ambystoma mexicanum*), coelacanth (*Latimeria chalumnae*), spotted gar (*Lepisosteus oculatus*), zebrafish (*Danio rerio*), and *Callorhinchus milii*, the elephant fish, which is sometimes labeled as elephant shark, and inshore hagfish (*Eptatretus burgeri*). In the case of the elephant fish, American alligator and axolotl, flanking genes were examined using the Entrez gene database from the NCBI (Maglott et al., 2005).

3 RESULTS

3.1 Identification of two Nodal paralogs in vertebrates

We recovered the monophyly of the Nodal and Nodal-related sequences from chordates with strong support in our Bayesian and ML analyses (Fig. 1). These analyses place the single ortholog of lancelets (cephalochordates) and sea squirts (urochordates) as sister to the clade that includes the Nodal and Nodal-related sequences of vertebrates (Fig. 1). The Nodal genes of jawed vertebrates were placed in two strongly supported clades, corresponding to the Nodal and Nodal-related genes respectively (Fig. 1). The Nodal clade includes cartilaginous fish, holostean and teleost fish, coelacanth, amphibians, lizards and mammals, and the Nodal-related clade includes cartilaginous fish, holost and teleost fish, coelacanth, amphibians, and sauropsids (Fig. 1). The Nodal-like genes of cyclostomes are placed in a strongly supported monophyletic group sister to the Nodal gene of jawed vertebrates, although support for this node is weak (Fig. 1). The Nodal gene is absent from the genomes of snakes, turtles, crocodilians, and birds, whereas the Nodal-related gene is absent from the genomes of all mammals examined. In all cases, deviations from the expected organismal phylogeny were weakly supported.

The results of our synteny analyses are consistent with the phylogenetic results: the Nodal and Nodal-related genes are each in a relatively well-conserved chromosomal location (Fig. 2). In almost all vertebrates, Nodal is flanked by an ortholog of EIF4EBP2 upstream and of PALD1 downstream, and Nodal-related is flanked by an EIF4EBP1 ortholog upstream and orthologs of ASH2L and STAR downstream (Fig.

2). In both cases, there is higher synteny conservation among bony vertebrates, from human to spotted gar. Interestingly, there is high synteny conservation for the regions corresponding to the Nodal and Nodal-related genes even in the lineages where these Nodal genes have been lost. Synteny comparisons between cyclostomes and the Nodal or Nodal-related genes of gnathostomes were uninformative.

3.2 Evolution of Nodal and Nodal-related repertoires in fish and amphibians

The cases of fish and amphibians deserve more careful attention, as these groups present additional variation in copy number of Nodal and Nodal-related genes relative to other vertebrates. In both lineages, Nodal has remained as a single-copy ortholog, despite inconsistencies in the nomenclature. For example, the Nodal genes of fish fall in a clade (Figs. 1 and 3) and include the single ortholog gene *cyclops*, also known as *ndr2*, for nodal-related-2, of zebrafish. The three copies of the Nodal-related gene from spotted gar are in an exclusive clade (Fig. 1 and 4), and the teleost Nodal-related genes are grouped into two clades, one of which includes the *southpaw* gene of zebrafish, also known as *ndr3*, and is sister to the clade that includes the three spotted gar paralogs and the other that includes the *squint* gene of zebrafish, also known as *ndr1* (Figs. 1 and 4). Thus, most teleosts have a single copy of Nodal, and duplicate copies of Nodal-related, whereas the spotted gar has a single copy of Nodal, and triplicate copies of Nodal-related.

The amphibian repertoire is the most variable among the groups studied. Axolotl, order Caudata, and frogs, order Anura, include both Nodal and Nodal-related genes in their genomes. There is a single ortholog of Nodal in all amphibians, known as *Xnr4*, for *Xenopus-nodal-related-4*. By contrast, there is a single copy of Nodal-related in axolotl, but three to eleven copies in anurans, including the *Xnr1*, 2, 3, 5 and 6 genes. The three frogs in our study include a single ortholog of *Xnr1*, which is sister to all other Nodal-related genes from frogs (Fig. 4), and a single ortholog of *Xnr2*. *Xnr6* is also present as a single copy in the two clawed frogs but is absent from the current assembly of the High Himalaya frog. There are single copies of *Xnr3* and 5 in the High Himalaya frog, but these two genes have expanded independently to 3 to 5 copies in the Western and African clawed frogs respectively (Figs. 1 and 4).

4 DISCUSSION

We studied the evolution of Nodal and Nodal-related genes with the main goal of integrating their evolutionary history with that of the GDF1/3 gene family to gain insights into the evolution of the heterodimer that is responsible for triggering the Nodal pathway early in development. Further, understanding the evolutionary history and phyletic distribution of these genes is also important as Nodal is expressed in multiple types of cancers (Strizzi et al., 2012) and the nodal pathway has emerged as a potential target for cancer therapy (Kalyan et al., 2017). Thus, our inferences of orthology should facilitate the selection of adequate model organisms for this research, and permit an evolutionarily sensible interpretation of experimental outcomes.

4.1 Evolution of the Nodal and Nodal-related genes in jawed vertebrates

The phylogenetic and synteny analyses resolve orthology relationships and the duplicative history of the Nodal and Nodal-related genes of vertebrates, placing them in reciprocally monophyletic groups (Fig. 1). Previous studies had identified a discordant number of gene lineages, probably because of scarce sequence information outside mammals and actinopterygian fish (Fan and Dougan, 2007; Kuraku and Kuratani, 2011). The key difference among these studies lied in the position of the clade that includes the teleost *ndr2/cyclops* genes. With the more extensive taxonomic sampling of our study, we are able to resolve the evolutionary history of the Nodal and Nodal-related genes of jawed vertebrates to two well-supported gene lineages, Nodal and Nodal-related (Fig. 1).

4.2 Inference of ancestral repertoires

Because both Nodal and Nodal-related are present in cartilaginous fish, we infer that these genes were present at least in the ancestor of jawed vertebrates, between 615 and 473 mya. Synteny is relatively well conserved across jawed vertebrates (Fig. 2), and the synteny blocks of Nodal, 8.1, and Nodal-related, 10.3, in human map to chromosomal segments that derive from the same ancestral chordate linkage group (CGL6 from Putnam et al. 2008). Thus, we infer that Nodal and Nodal-related derive from the whole genome duplications early in vertebrate evolution. In agreement with this, most non-vertebrate species screened maintain a single ortholog of the gene

(Grande et al., 2014). Interestingly, we were able to find Nodal-like genes in the current genome assemblies of both hagfish and lampreys, the two extant lineages of cyclostomes, whereas a previous study included no Nodal-like gene in a particular lamprey species, the sea lamprey (*Petromyzon marinus*) (Lagadec et al., 2015). In our results, phylogenetic analyses would suggest the orthology of the identified cyclostome Nodal-like gene to the Nodal gene of jawed vertebrates, rather than Nodal-related, although the support is not high (Figs. 1 and 3).

Taken together, our results indicate that the duplication of Nodal and Nodal-related genes can be traced back at least to the last common ancestor of jawed vertebrates, and possibly to the last common ancestor of extant vertebrates. Unfortunately, as in other gene families such as globins, Myb and KCNA, resolving orthology between cyclostome and gnathostomes remains a hard problem because of GC-, codon-, and amino acid composition biases (Qiu et al., 2011; Smith et al., 2013; Schwarze et al., 2014; Campanini et al., 2015; Opazo et al., 2015). This has complicated inferences of the phenotype of the last common ancestor of vertebrates on the basis of gene orthology (reviewed in (Onimaru and Kuraku, 2018)).

4.3 Differential retention of Nodal and Nodal-related genes

Tracing gains and losses along the vertebrate phylogeny shows that the differential retention of ancestral genes has played an important role in shaping Nodal repertoires (Fig 5). Nodal is present in all major lineages of vertebrates other than birds, crocodiles, turtles, and snakes (Fig. 3), suggesting that it was lost twice independently, that is, in the common ancestor of archelosaurians (birds, crocodiles, and turtles), between 280 and 254 millions of years ago, and in the snake lineage. On the other hand, Nodal-related is present in all main groups of vertebrates other than mammals and cyclostomes (Fig. 4).

The case of teleost fish deserves attention given that this group experienced an additional whole genome duplication in their early evolution (Amores et al., 1998; Petit et al., 2004; Kuraku and Meyer, 2009). Accordingly, we expect a more diverse repertoire in comparison to other vertebrates. This was the case, as teleosts have two Nodal-related genes: ndr1/squint and ndr3/southpaw, and one Nodal ortholog, ndr2/cyclops. Fan & Dougan (2007) suggested that the diversity of Nodal genes in

teleosts arose as a combination of the teleost-specific genome duplication and small-scale duplication events. According to these authors, the teleost-specific genome duplication would have given rise to two ancestral genes. Later, one of them would have undergone a duplication event giving rise to the ndr1/squint and ndr3/southpaw genes (Fan and Dougan, 2007). By contrast, our results indicate that the ancestor of teleosts already had two genes, which as a product of the teleost-specific genome duplication would have given rise to a repertoire of four. After this genome duplication, both Nodal-related ohnologs were retained (ndr1/squint and ndr3/southpaw)(Figs. 1 and 4), whereas only one of the Nodal ohnologs was retained (ndr2/cyclops)(Figs. 1 and 3).

We observed additional lineage-specific expansions of the Nodal-related gene in the spotted gar and amphibians of the order Anura, the group that includes frogs and toads (Fig. 4). The spotted gar has three Nodal-related genes (Figs. 1 and 4) which are very similar among them with pairwise distances ranging from 2.43% to 3.78%, and group with the ndr3/squint gene of teleosts. The most important gene expansion occurred in anurans. The duplications giving rise to the Xnr1, 2, 3, 5 and 6 genes are shared among all frogs examined. The Xnr6 has been apparently lost in the High Himalaya frog, and the Xnr3 and Xnr5 genes have expanded further in the two clawed frog lineages independently. Thus, our analyses indicate that the expansion of Nodal-related genes occurred in two waves, the first was after the divergence of the orders Anura and Caudata but before the radiation of anuran species, whereas the second wave happened independently in the western and African clawed frog species (Fig. 4).

4.4 Gene duplication and gene loss as sources of evolutionary innovations

The differential retention of genes could be seen as a stochastic process, in which the resulting differences in gene complement do not translate into functional consequences. This could be due to some functional redundancy among the paralogs that could work as a backup (i.e., functionally overlapping paralogues) in the case one is lost or inactivated (Cañestro et al., 2009; Félix and Barkoulas, 2015; Albalat and Cañestro, 2016), or a by-product of divergent nature of the genomic location in which the duplicates are found (Hara et al., 2018b, 2018a). Alternatively, it is also possible that the presence of multiples copies opens opportunities for the

emergence of biological novelties (Ohno et al., 1968; Ohno, 1970; Hughes et al., 1994; Force et al., 1999; Zhang, 2003). For example, the differential retention of functional copies of the INSL4 gene in catarrhine primates has been associated to unique reproductive traits of the group (Arroyo et al., 2012a, 2012b; Malone et al., 2017).

The loss of Nodal in birds is informative in this regard because several studies have reported the expression of the Nodal gene during early development of the chicken (Raya et al. 2004a,b, and references therein). Thus, we infer these studies were actually reporting the expression of the Nodal-related gene, which is annotated as Nodal (<https://www.ncbi.nlm.nih.gov/gene/395205>, NCBI). This is an example of how 'hidden paralogy', the differential retention of alternate paralogs, has complicated reconstructions of ancestral states in vertebrate evolution (reviewed in [Kuraku 2010](#)) and highlights the need for applying phylogenetically explicit approaches to make orthology inferences (Gabaldón, 2008; Altenhoff et al., 2016). From a functional standpoint, the Nodal-related gene of birds seems to be fulfilling the role of Nodal (Raya et al. 2004a,b, and references therein), suggesting these paralogs have retained some degree of redundancy.

4.5 Evolution of the Nodal-GDF1/3 heterodimers

There has been evidence for the co-dependence of the Nodal and GDF1/3 gene families during early developmental processes in all main group of vertebrates for some time, but their interactions were not clear. This has recently changed, as three independent groups using different approaches reported that the formation heterodimers between Nodal and GDF1/3 is necessary to activate the Nodal pathway (Bisgrove et al., 2017; Montague and Schier, 2017; PellicciaJ.L. et al., 2017).

From an evolutionary standpoint, the vertebrate Nodal/Nodal-related and GDF1/3 gene families have undergone lineage-specific expansions and retentions in different vertebrate lineages (Opazo and Zavala, 2018). As a result, the signaling heterodimer, which triggers the Nodal developmental pathway, is composed by proteins that derive from different genes in different vertebrate groups (Fig. 5). Mapping the presence of the corresponding genes indicates that cyclostomes,

cartilaginous fish, teleosts, holostean fish, caudates, and lizards possess a single ortholog of Nodal and a single ortholog of GDF1/3 gene (Fig. 5 and 6). So, these groups would be able to form heterodimers following the conformation recently reported (Bisgrove et al., 2017; Montague and Schier, 2017; PellicciaJ.L. et al., 2017). In birds, turtles, crocodiles, and snakes, however, there is no Nodal gene, as far as the species analyzed so far are concerned. Evidence from chicken suggests that Nodal-related is involved in forming the heterodimers with GDF1/3 (Raya et al., 2004a, 2004b), and this is probably the case for all the groups listed above, in a case that illustrates the evolutionary significance of functional redundancy.

The case of anurans and mammals is also interesting; as they both possess a single ortholog of Nodal and independently derived paralogs of GDF1/3, the GDF1_A and GDF3_A genes in amphibians derive from an amphibian-specific duplication whereas the GDF1_M and GDF3_M genes of mammals derive from a mammalian-specific duplication (Fig. 5) (Opazo and Zavala, 2018). Thus, the Nodal genes of mammals and anurans are 1:1 orthologs, but their partners for the heterodimer, GDF1 or GDF3, are not (Opazo and Zavala, 2018). Thus, mammals activate the pathway with a heterodimer containing Nodal and GDF1_M (Fig. 5), whereas anurans trigger the Nodal developmental pathway with Nodal and GDF1_A (Vg1)(Fig. 5).

In summary, there are at least four different heterodimer conformations that activate the Nodal pathway in vertebrates (Fig. 5): 1) Nodal + GDF1/3, in cyclostomes, cartilaginous fish, teleosts, holostean fish, caudates and lizards; 2) Nodal-related + GDF1/3, in birds, crocodiles, turtles, and snakes; 3) Nodal and GDF1_A in anurans; and 4) Nodal + GDF1_M in mammals. The diversity of heterodimers that can trigger the same developmental pathway highlights the redundancy present in the genomes of vertebrates and the versatility of the evolutionary process. In this case, the duplication that gave rise to Nodal and Nodal-related predates the divergence between cartilaginous fish and the rest of extant jawed vertebrates (615-473 mya), and one of the duplicates, Nodal, was lost in the common ancestor of archelosaurs (312-254 mya). Despite their relatively old divergence, both nodal and nodal-related seem to have retained the ability to activate the pathway in the absence of nodal, suggesting that functional redundancy can last long periods of time even if not directly needed. Furthermore, it is important to note that the role of Nodal-related in

species that also have Nodal is poorly understood, which suggests that these species could form a much more diverse array of heterodimers. Thus, comparing the Nodal pathway in representative species belonging to the four groups of heterodimer conformations will shed light regarding the robustness of the nodal pathway.

ACKNOWLEDGMENTS

This work was supported by the Fondo Nacional de Desarrollo Científico y Tecnológico (FONDECYT 1160627) grant to JCO. FGH's work was partially funded by the National Science Foundation through grant EPSCoR RII Track-2 FEC 1736026.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

REFERENCES

- Albalat, R., and Cañestro, C. 2016. Evolution by gene loss. *Nat. Rev. Genet.* 17: 379–391.
- Altenhoff, A.M., Boeckmann, B., Capella-Gutierrez, S., Dalquen, D.A., DeLuca, T., Forslund, K., Huerta-Cepas, J., Linard, B., Pereira, C., Przych, L.P., et al. 2016. Standardized benchmarking in the quest for orthologs. *Nat. Methods* 13: 425–430.
- Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215: 403–410.
- Amores, A., Force, A., Yan, Y., Joly, L., Amemiya, C., Fritz, A., Ho, R.K., Langeland, J., Prince, V., Wang, Y., et al. 1998. Zebrafish hox Clusters and Vertebrate Genome Evolution. *Science (80-.)*. 282: 1711–1715.
- Arroyo, J.I., Hoffmann, F.G., Good, S., and Opazo, J.C. 2012a. INSL4 pseudogenes help define the relaxin family repertoire in the common ancestor of placental mammals. *J. Mol. Evol.* 75:.
- Arroyo, J.I., Hoffmann, F.G., and Opazo, J.C. 2012b. Gene turnover and differential retention in the relaxin/insulin-like gene family in primates. *Mol. Phylogenet. Evol.* 63:.
- Bisgrove, B.W., Su, Y.-C., and Yost, H.J. 2017. Maternal Gdf3 is an obligatory cofactor in Nodal signaling for embryonic axis formation in zebrafish. *Elife* 6: e28534.
- Campanini, E.B., Vandewege, M.W., Pillai, N.E., Tay, B., Jones, J.L., Venkatesh, B., and Hoffmann, F.G. 2015. Early Evolution of Vertebrate Mybs : An Integrative. *Genome Biol. Evol.* 7: 3009–3021.
- Cañestro, C., Catchen, J.M., Rodriguez-Mari, A., Yokoi, H., and Postlethwait, J.H. 2009. Consequences of Lineage-Specific Gene Loss on Functional Evolution of Surviving Paralogs : ALDH1A and Retinoic Acid Signaling in Vertebrate Genomes. *PLoS Genet.* 5: 1–19.
- Conlon, F.L., Lyons, K.M., Takaesu, N., Barth, K.S., and Kispert, A. 1994. A primary requirement for nodal in the formation and maintenance of the primitive streak in the mouse. *Development* 120: 1919–1928.
- Fan, X., and Dougan, S.T. 2007. The evolutionary origin of nodal-related genes in teleosts. *Dev. Genes Evol.* 217: 807–813.
- Feldman, B., Gates, M.A., Egan, E.S., Dougan, S.D., Rennebeck, G., Sirotkin, Howard İçSchier, A.F., and Talbot, W.S. 1998. Zebrafish organizer development and germ-layer formation require nodal-related signals. *Nature* 395: 181–185.
- Félix, M., and Barkoulas, M. 2015. Pervasive robustness in biological systems. *Nat. Rev. Genet.* 16: 483–496.
- Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y., and Postlethwait, J. 1999. Preservation of Duplicate Genes by Complementary , Degenerative Mutations. *Genetics* 151: 1531–1545.
- Gabaldón, T. 2008. Large-scale assignment of orthology: back to phylogenetics. *Genome Biol.* 9: 235.
- Geer, L.Y., Marchler-bauer, A., Geer, R.C., Han, L., He, J., He, S., Liu, C., Shi, W., and Bryant, S.H. 2010. The NCBI BioSystems database. *Nucleic Acids Res.* 38: 492–496.
- Gilbert, S.F. 2017. Developmental biology , the stem cell of biological disciplines.

PLoS Biol. 15: 1–8.

- Grande, C., Martín-durán, J.M., Kenny, N.J., Truchado-garcía, M., and Hejnal, A. 2014. Evolution, divergence and loss of the Nodal signalling pathway: new data and a synthesis across the Bilateria. *Int. J. Dev. Biol.* 58: 521–532.
- Hara, Y., Takeuchi, M., Kageyama, Y., Tatsumi, K., Hibi, M., and Kiyonari, H. 2018a. Madagascar ground gecko genome analysis characterizes asymmetric fates of duplicated genes. *BMC Biol.* 16: 1–19.
- Hara, Y., Yamaguchi, K., Onimaru, K., Kadota, M., Koyanagi, M., Keeley, S.D., Tatsumi, K., Tanaka, K., Motone, F., Kageyama, Y., et al. 2018b. Shark genomes provide insights into elasmobranch evolution and the origin of vertebrates. *Nat. Ecol. Evol.* 2: 1761–1771.
- Herrero, J., Muffato, M., Beal, K., Fitzgerald, S., Gordon, L., Pignatelli, M., Vilella, A.J., Searle, S.M.J., Amode, R., Brent, S., et al. 2016. Ensembl comparative genomics resources. *Database* 2016: baw053.
- Hughes, A.L., Hughes, M.K., Howell, C.Y., and Nei, M. 1994. Natural Selection at the Class II Major Histocompatibility Complex Loci of Mammals. *Philos. Trans. R. Soc. B* 345: 359–367.
- Jones, C.M., Kuehn, M.R., Hogan, B.L.M., Smith, J.C., and Wright, C.V.E. 1995. Nodal-related signals induce axial mesoderm and dorsalize mesoderm during gastrulation. *Development* 121: 3651–3662.
- Joseph, E.M., and Melton, D.A. 1997. Xnr4: A Xenopus Nodal-Related Gene Expressed in the Spemann Organizer. *Dev. Biol.* 184: 367–372.
- Kalyan, A., Carneiro, B.A., Chandra, S., Kaplan, J., Chae, Y.K., Matsangou, M., Hendrix, M.J.C., and Giles, F. 2017. Nodal Signaling as a Developmental Therapeutics Target in Oncology. *Mol. Cancer Ther.* 16: 787–793.
- Katoh, K., and Standley, D.M. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Mol. Biol. Evol.* 30: 772–780.
- Kumar, S., Stecher, G., Suleski, M., and Hedges, S.B. 2017. TimeTree: A Resource for Timelines, Timetrees, and Divergence Times. *Mol. Biol. Evol.* 34: 1812–1819.
- Kuraku, S., and Kuratani, S. 2011. Genome-Wide Detection of Gene Extinction in Early Mammalian Evolution. *Genome Biol. Evol.* 3: 1449–1462.
- Kuraku, S., and Meyer, A. 2009. The evolution and maintenance of Hox gene clusters in vertebrates and the teleost-specific genome duplication. *Int. J. Dev. Biol.* 773: 765–773.
- Lagadec, R., Laguerre, L., Menuet, A., Amara, A., Rocancourt, C., Pericard, P., Godard, B., Rodicio, M.C., Rodriguez-Moldes, I., Mayeur, H., et al. 2015. The ancestral role of nodal signalling in breaking. *Nat. Commun.* 6: 1–10.
- Levin, M., Johnson, R.L., Stern, C.D., Kuehn, M., and Tabin, C. 1995. Determining Left-Right Asymmetry in Chick Embryogenesis. *Cell* 82: 803–814.
- Li, G., Liu, X., Xing, C., Zhang, H., Shimeld, S.M., and Wang, Y. 2017. Cerberus – Nodal – Lefty – Pitx signaling cascade controls left – right asymmetry in amphioxus. *Proc. Natl. Acad. Sci.* 114: 3684–3689.
- Louis, A., Nguyen, N.T.T., Muffato, M., and Roest Crolius, H. 2015. Genomicus update 2015: KaryoView and MatrixView provide a genome-wide perspective to multispecies comparative genomics. *Nucleic Acids Res.* 43: D682–D689.
- Maglott, D., Ostell, J., Pruitt, K.D., and Tatusova, T. 2005. Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.* 33: 54–58.
- Malone, L., Opazo, J.C., Ryan, P.L., and Hoffmann, F.G. 2017. Progressive erosion

- of the Relaxin1 gene in bovids. *Gen. Comp. Endocrinol.* 252:.
- Montague, T.G., and Schier, A.F. 2017. Vg1-Nodal heterodimers are the endogenous inducers of mesendoderm. *Elife* 6: e28183.
- Ohno, S. 1970. Evolution by gene duplication (Springer-Verlag).
- Ohno, S., Wolf, U., and Atkin, N.B. 1968. Evolution from fish to mammals by gene duplication. *Hereditas* 59: 169–187.
- Onimaru, K., and Kuraku, S. 2018. Inference of the ancestral vertebrate phenotype through vestiges of the whole-genome duplications. *Brief. Funct. Genomics* 17: 352–361.
- Opazo, J.C., Lee, A.P., Hoffmann, F.G., Toloza-Villalobos, J., Burmester, T., Venkatesh, B., and Storz, J.F. 2015. Ancient duplications and expression divergence in the globin gene superfamily of vertebrates: Insights from the elephant shark genome and transcriptome. *Mol. Biol. Evol.* 32: 1684–1694.
- Opazo, J.C., and Zavala, K. 2018. Phylogenetic evidence for independent origins of GDF1 and GDF3 genes in anurans and mammals. *Sci. Rep.* 1: 1–7.
- PellicciaJ.L., Jindal, G.A., and Burdine, R.D. 2017. Gdf3 is required for robust Nodal signaling during germ layer formation and left-right patterning. *Elife* 6: e28635.
- Petit, J., Stange-thomann, N., Mauceli, E., Bouneau, L., Jaillon, O., Aury, J., Ozouf-costaz, C., Bernot, A., Nicaud, S., Jaffe, D., et al. 2004. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. *Nature* 431: 946–957.
- Qiu, H., Hildebrand, F., Kuraku, S., and Meyer, A. 2011. Unresolved orthology and peculiar coding sequence properties of lamprey genes : the KCNA gene family as test case. *BMC Genomics* 12: 1–11.
- Raya, A., Carlos, J., and Belmonte, I. 2004a. Unveiling the establishment of left – right asymmetry in the chick embryo. *Mech. Dev.* 121: 1043–1054.
- Raya, A., Kawakami, Y., Rodriguez-Esteban, C., Ibañes, M., Rasskin-Gutman, D., Rodriguez-Leon, J., Buscher, D., Feijo, J.A., and Izpisua Belmonte, J.C. 2004b. Notch activity acts as a sensor for extracellular calcium during vertebrate left – right determination. *Nature* 427: 121–128.
- Ronquist, F., and Huelsenbeck, J.P. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.
- Schier, A.F. 2003. Nodal signaling in vertebrate development. *Annu. Rev. Cell Dev. Biol.* 19: 589–621.
- Schwarze, K., Campbell, K.L., Hankeln, T., Storz, J.F., Hoffmann, F.G., and Burmester, T. 2014. The Globin Gene Repertoire of Lampreys : Convergent Evolution of Hemoglobin and Myoglobin in Jawed and Jawless Vertebrates. *Mol. Biol. Evol.* 31: 2708–2721.
- Seleiro, E.A.P., Connolly, D.J., and Cooke, J. 1996. Early developmental expression and experimental axis determination by the chicken Vg1 gene. *Curr. Biol.* 6: 1476–1486.
- Shen, M.M. 2007. Nodal signaling : developmental roles and regulation. *Development* 134: 1023–1034.
- Smith, J.J., Kuraku, S., Holt, C., Sauka-spengler, T., Jiang, N., Campbell, M.S., Yandell, M.D., Manousaki, T., Meyer, A., Bloom, O.E., et al. 2013. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. *Nat. Genet.* 45: 415–421.
- Strizzi, L., Hardy, K.M., Kirschmann, D.A., Ahrlund-richter, L., and Hendrix, M.J.C. 2012. Nodal Expression and Detection in Cancer : Experience and Challenges. *Cancer Res.* 72: 1915–1921.

- Tadjuidje, E., Kofron, M., Mir, A., Wylie, C., Heasman, J., and Cha, S. 2016. Nodal signalling in *Xenopus* : the role of Xnr5 in left / right asymmetry and heart development. *Open Biol.* 6: 150187.
- Tatusova, T.A., and Madden, T.L. 1999. BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS Microbiol. Lett.* 174: 247–250.
- Trifinopoulos, J., Nguyen, L.-T., von Haeseler, A., and Minh, B.Q. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Res.* 44: W232–W235.
- Zerbino, D.R., Achuthan, P., Akanni, W., Amode, M.R., Barrell, D., Bhai, J., Billis, K., Cummins, C., Gall, A., Gil, L., et al. 2018. Ensembl 2018. *Nucleic Acids Res.* 46: 754–761.
- Zhang, J. 2003. Evolution by gene duplication : an update. *Trends Ecol. Evol.* 18: 292–298.
- Zinski, J., Tajer, B., and Mullins, M.C. 2018. TGF- b Family Signaling in Early Vertebrate Development. *Cold Spring Harb. Perspect. Biol.* 10: 1–76.

FIGURE LEGENDS

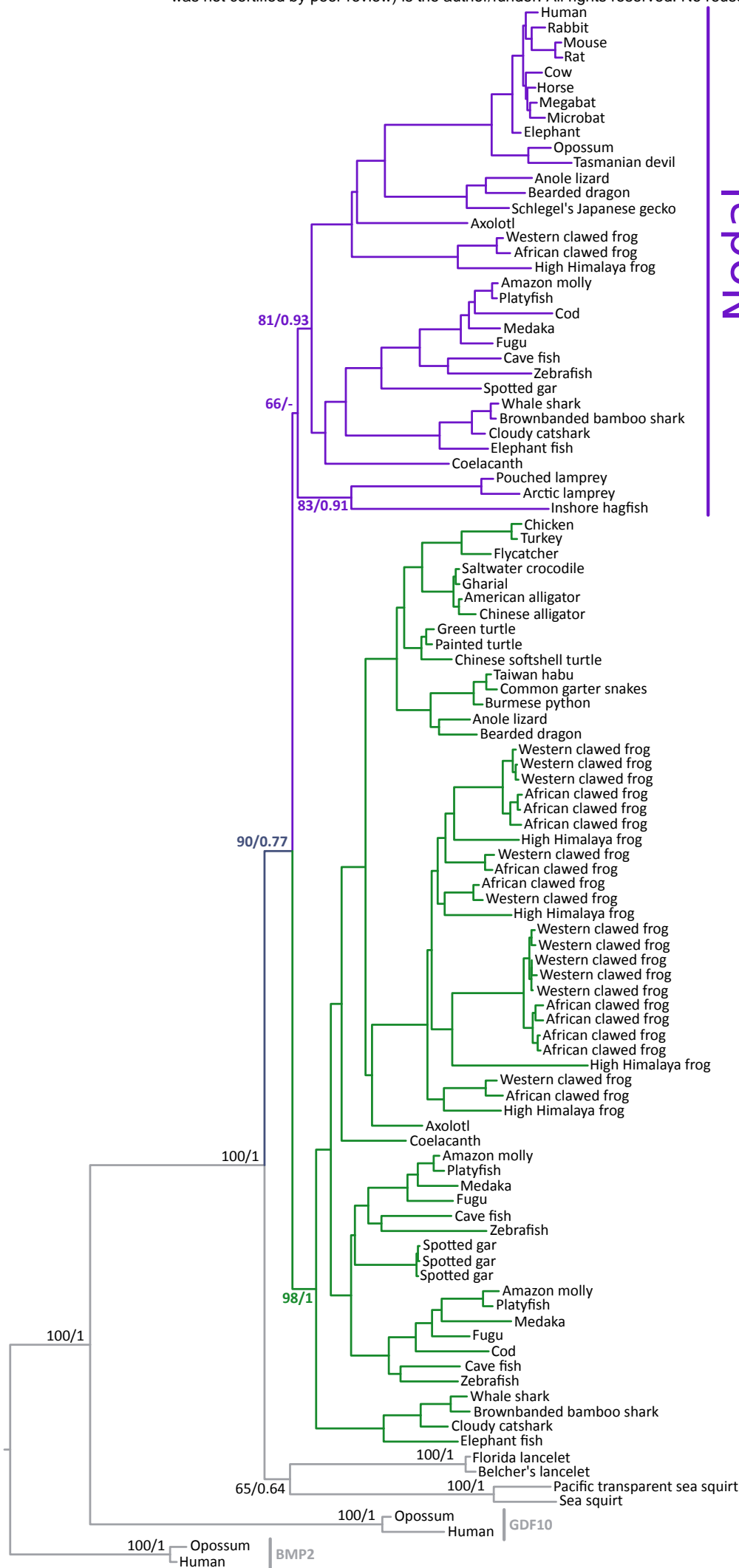
Figure 1. A maximum-likelihood tree depicting evolutionary relationships among the Nodal and Nodal-related genes of chordates. Numbers on the nodes represent maximum likelihood ultrafast bootstrap and Bayesian posterior probability support values. BMP2 sequences from human (*Homo sapiens*) and opossum (*Monodelphis domestica*) were used as outgroups.

Figure 2. Gene order in the chromosomal region containing the Nodal and Nodal-related genes of vertebrates. Asterisks denote that the orientation of the genomic piece is from 3' to 5', gray lines represent intervening genes that do not contribute to conserved synteny.

Figure 3. Maximum likelihood tree depicting evolutionary relationships among Nodal genes in vertebrates. Numbers on the nodes correspond to maximum likelihood ultrafast bootstrap support and Bayesian posterior probability support values. This tree topology does not represent a novel phylogenetic analysis; it is the Nodal clade that was recovered from Fig. 1.

Figure 4. Maximum likelihood tree depicting evolutionary relationships among Nodal-related genes in vertebrates. Numbers on the nodes correspond to maximum likelihood ultrafast bootstrap support and Bayesian posterior probability support values. This tree topology does not represent a novel phylogenetic analysis; it is the Nodal clade that was recovered from Fig. 1.

Fig. 5. Phyletic distribution of Nodal, Nodal-related, GDF1/3, GDF1 and GDF3 genes among vertebrates, including the putative composition of the heterodimer. The different colors refer to the putative four types of heterodimers that can trigger the Nodal pathway in vertebrates. The phyletic distribution of the GDF1/3, GDF1 and GDF3 genes was obtained from Opazo & Zavala (2018).



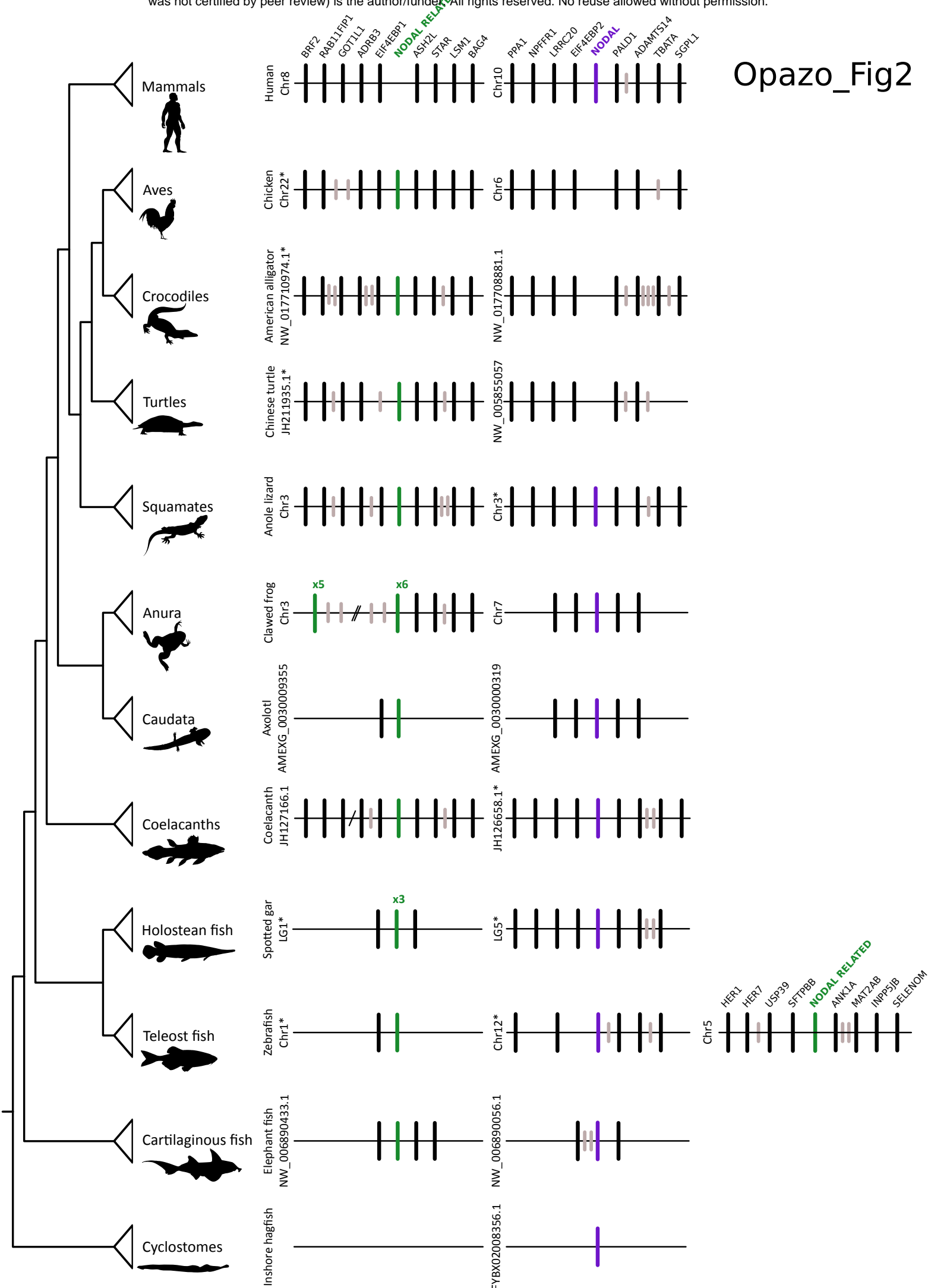
Nodal

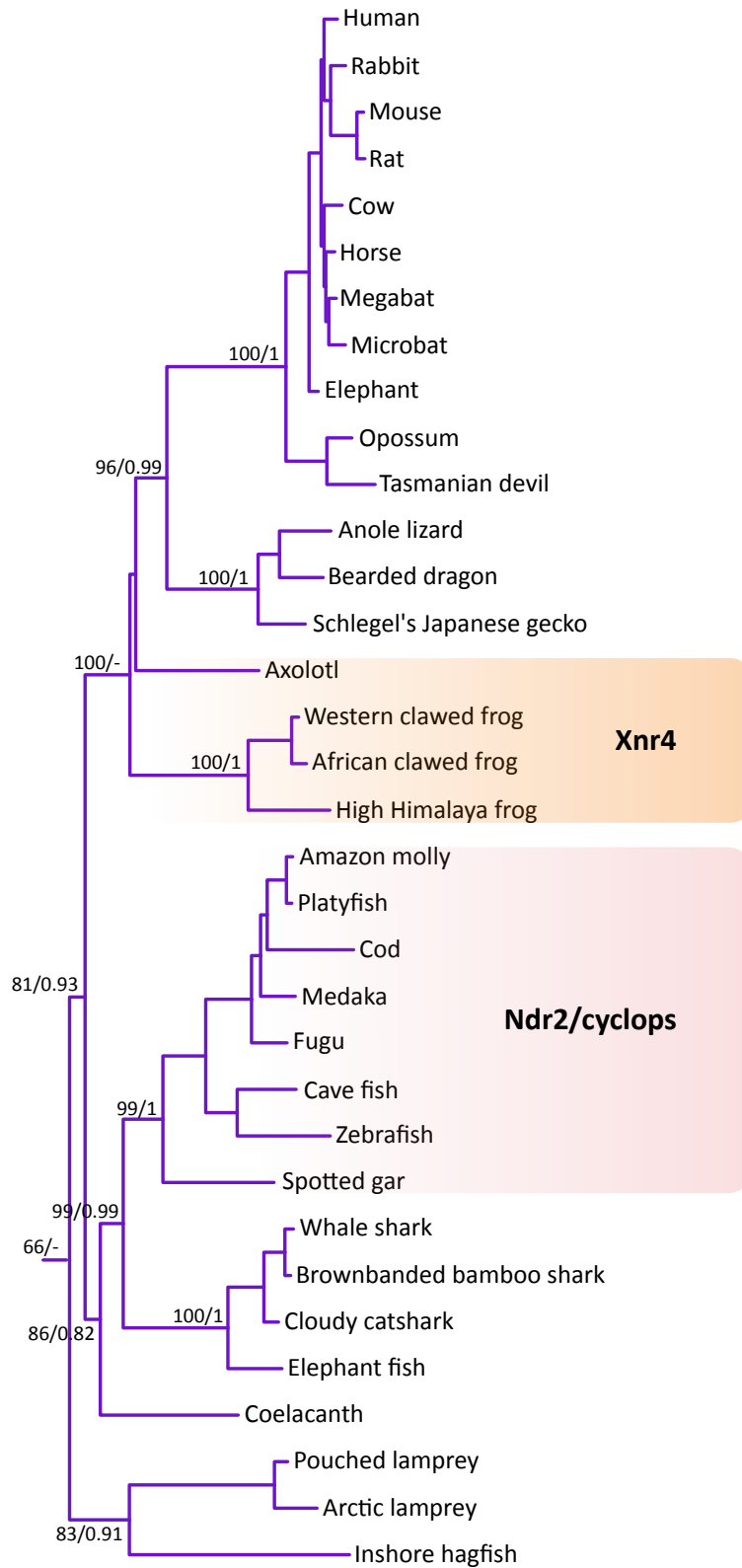
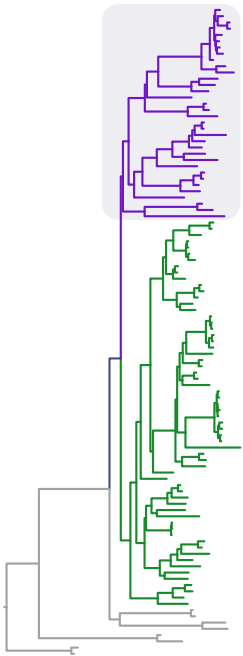
Nodal related

BMP2

GDF10

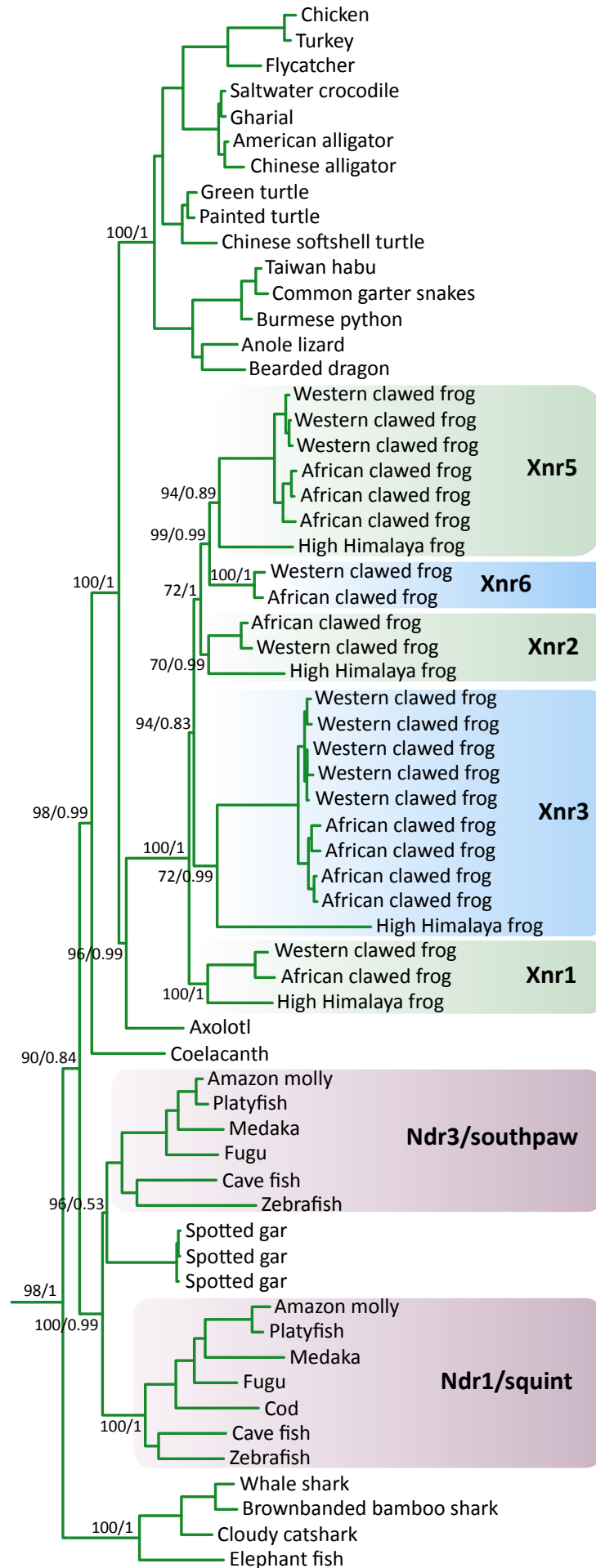
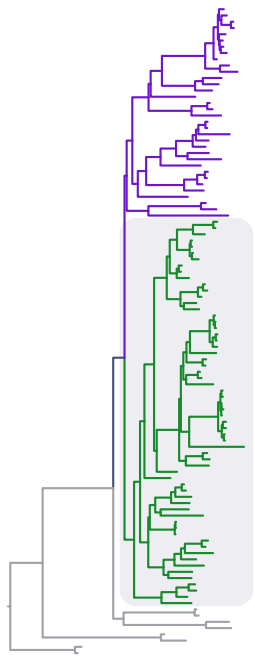
Opazo_Fig2





Nodal

0.7



Nodal related

