Peer Community In
Evolutionary Biology

RESEARCH ARTICLE

Open Access
Open Data
Open Code
Open Peer-Review

Evolution of the DAN gene family in vertebrates

Juan C. Opazo[1,2,3], Federico G. Hoffmann[4,5], Kattina Zavala[1], Scott V. Edwards[6]

[1]Instituto de Ciencias Ambientales y Evolutivas, Facultad de Ciencias, Universidad Austral de Chile, Valdivia, Chile.

[2]David Rockefeller Center for Latin American Studies, Harvard University, Cambridge, MA 02138, USA.

[3]Millennium Nucleus of Ion Channels-Associated Diseases (MiNICAD).

[4]Department of Biochemistry, Molecular Biology, Entomology, and Plant Pathology, Mississippi State University, Mississippi State, 39762, USA.

[5]Institute for Genomics, Biocomputing, and Biotechnology, Mississippi State University, Mississippi State, 39762, USA.

[6]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA.

Cite as: Opazo JC, Hoffmann FG, Zavala K, Edwards SV (2020) Evolution of the DAN gene family in vertebrates. bioRxiv, 794404, ver. 2 peer-reviewed and recommended by PCI Evolutionary Biology. doi: 10.1101/794404

This article has been peer-reviewed and recommended by
Peer Community in Evolutionary Biology
doi: 10.24072/pci.evolbiol.100104

Posted: 28 June 2020

Recommender: Kateryna Makova

Reviewers: 2 anonymous reviewers

Correspondence: jopazo@gmail.com

ABSTRACT

The DAN gene family (DAN, Differential screening-selected gene Aberrant in Neuroblastoma) is a group of genes that is expressed during development and plays fundamental roles in limb bud formation and digitation, kidney formation and morphogenesis and left-right axis specification. During adulthood the expression of these genes are associated with diseases, including cancer. Although most of the attention to this group of genes has been dedicated to understanding its role in physiology and development, its evolutionary history remains poorly understood. Thus, the goal of this study is to investigate the evolutionary history of the DAN gene family in vertebrates, with the objective of complementing the already abundant physiological information with an evolutionary context. Our results recovered the monophyly of all DAN gene family members and divide them into five main groups. In addition to the well-known DAN genes, our phylogenetic results revealed the presence of two new DAN gene lineages; one is only retained in cephalochordates, whereas the other one (GREM3) was only identified in cartilaginous fish, holostean fish, and coelacanth. According to the phyletic distribution of the genes, the ancestor of gnathostomes possessed a repertoire of eight DAN genes, and during the radiation of the group GREM1, GREM2, SOST, SOSTDC1, and NBL1 were retained in all major groups, whereas, GREM3, CER1, and DAND5 were differentially lost.

Keywords: gene family evolution, Cerberus, differential retention, evolutionary medicine, evolutionary slowdown, gremlin.

Introduction

Genetic variation harbored in non-model species represents a powerful resource to gain insights in our understanding of the genetic bases of biological diversity (Opazo et al. 2005; Storz et al. 2007; Faulkes et al., 2015). The comparative approach has gained popularity since the discovery of non-model species that are resistant to diseases that affect human health, and also because some non-model species develop pathologies in a similar way as ourselves (Protas et al. 2006; Gorbunova et al., 2012; Castro-Fuentes and Socas-Pérez, 2013; Manov et al., 2013; Henning et al., 2014; Braidy et al., 2015).

The DAN gene family (DAN stands for Differential screening-selected gene Aberrant in Neuroblastoma) is a group of genes characterized by the presence of a DAN domain (http://pfam.xfam.org/family/PF03045), low inter-paralog conservation and expression during early development. DAN genes play fundamental roles in limb bud formation and digitation, kidney formation, and morphogenesis as well as left-right axis specification (Nolan & Thompson, 2014). In adult organisms, the expression of DAN genes is associated with pathologies such as cancer and nephropathies (Koli et al. 2006; Walsh et al. 2010; Gu et al. 2012; Droguett et al. 2014; Liang et al. 2015). As currently recognized, the DAN gene family comprises seven paralogs: NBL1 (neuroblastoma 1); CER1 (cerberus 1); DAND5 (DAN domain BMP antagonist family member 5); SOST (sclerostin); SOSTDC1 (sclerostin domain containing 1); GREM1 (gremlin 1); and GREM2 (gremlin2) (Avsian-Kretchmer et al. 2004; Walsh et al. 2010; Nolan & Thompson, 2014). Historically, DAN genes have been associated with the inhibition of the bone morphogenetic protein (BMP) signaling pathway, although more recent studies have shown that they also act as antagonists to nodal, wingless-type MMTV integration site (Wnt) and vascular endothelial growth factor (VEGF) signaling cascades (Piccolo et al. 1999; Ellies et al. 2006; Chiodelli et al. 2011). Most of the work on DAN genes has focused on understanding their role in development (Nolan & Thompson, 2014, and references therein), and less effort has focused on their evolutionary history (Avsian-Kretchmer et al. 2004; Walsh et al; 2010; Nolan & Thompson, 2014; Le Petillon et al. 2013; Opazo et al. 2017). Understanding the evolutionary history of gene families permits insights into the ways in which genes originate and diversify, as well as rates of evolution, the relative roles of different mutational processes such as point mutations and duplications and, most importantly, relating all of these phenomena to the physiological and life-history idiosyncrasies of different taxonomic groups.

The goal of this study is to investigate the evolutionary history of the DAN gene family in vertebrates, with the aim of complementing the already abundant physiological information with an evolutionary perspective. Therefore, we examined the diversity of DAN genes in representative species of all major groups of

vertebrates, inferred homologous relationships by examining gene phylogenies and synteny conservation, reconstructed ancestral gene repertoires, examined patterns of differential gene retention and quantified rates of molecular evolution. Our results recovered the monophyly of all DAN gene family members and grouped them into five clades. In addition to the well-known DAN genes, our phylogenetic results revealed the presence of two additional DAN gene lineages; one is only retained in cephalochordates (e.g. amphioxus), whereas the second one (GREM3) has only been retained by cartilaginous fish, holostean fish, and coelacanth. The phylogenetic distribution of genes suggests that the ancestor of gnathostomes possessed a repertoire of eight DAN genes, and that during the radiation of vertebrates, GREM1, GREM2, SOST, SOSTDC1, and NBL1 were retained in all major groups, whereas, GREM3, CER1, and DAND5 were differentially lost, creating complex patterns of retention and diversification.

Methods

DNA data and phylogenetic analyses

*To minimize potential errors in the annotation of the sequences, w*e manually annotated DAN genes in representative species of all major groups of chordates. To do so, we first identified genomic fragments containing DAN genes in Ensembl v86 (Zerbino et al., 2018) or National Center for Biotechnology Information (NCBI) databases (refseq_genomes, htgs, and wgs; Geer et al., 2010) based on conserved synteny s. Once identified, genomic pieces were extracted including the 5' and 3' flanking genes. After extraction, we manually curated the existing annotation or we annotate the novo by comparing known exon sequences (query sequence) from a species that share a common ancestor most recently in time to the species of which the genomic piece (subject sequence) is being analyzed using the program Blast2seq v2.5 with default parameters (Tatusova & Madden, 1999). Sequences derived from shorter records based on genomic DNA or complementary DNA were also included to attain a broad taxonomic coverage. We included representative species from mammals, birds, reptiles, amphibians, lobe-finned fish, holostean fish, teleost fish, cartilaginous fish, cyclostomes, urochordates and cephalochordates (Supplementary Table S1). Amino acid sequences were aligned using the FFT-NS-i strategy from MAFFT v.6 (Katoh & Standley, 2013). Our alignment contained 424 sequences and a length of 797 characters. Phylogenetic relationships were estimated using maximum likelihood and Bayesian approaches. Our goal was to estimate the history of a gene family, not the phylogeny of the organismal lineages themselves, hence we used traditional gene-tree based phylogenetic methods (Edwards, 2009; Yang & Rannala, 2012). We used IQ-Tree (Kalyaanamoorthy et al., 2017) to select the best-

fitting model of amino acid substitution (JTT + I + G4) and to obtain the maximum likelihood tree (Trifinopoulos et al. 2016) and assessed support for the nodes with 1,000 bootstrap pseudoreplicates using the ultrafast routine. Bayesian searches were conducted in MrBayes v.3.1.2 (Ronquist and Huelsenbeck, 2003), setting two independent runs of six simultaneous chains for $5x10^6$ generations, sampling every 2,500 generations, and using default priors. Once the analyses were done, we verified that the estimated sample size (ESS) exceeded the recommended value of 200 using Tracer ver 1.7.1 (Rambaut et al. 2018). The run was deemed converged once the likelihood scores reached an asymptotic value and the average standard deviation of 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 4,000 trees.

Assessments of Conserved Synteny

We examined genes found upstream and downstream of each member of the DAN gene family on species representative of all the major groups of vertebrates. Synteny analyses were performed in humans, chicken, spotted gar and elephant shark. Initial ortholog predictions were derived from the EnsemblCompara database (Vilella et al. 2009) and were visualized using the program Genomicus v84.01 (Muffato et al. 2010). In the case of the elephant shark the genomic segments containing DAN genes were annotated, and predicted genes were then compared with the non-redundant protein database using Basic Local Alignment Search Tool (BLAST) (Altschul et al. 1990).

Results and discussion

Gene phylogenies and synteny analyses define homology

Our phylogenetic analyses recovered the monophyly of each members of the genes in the DAN gene family with strong support with the exception of DAND5 (Fig. 1). In addition, our trees suggest an arrangement in which these genes are divided into five main groups (Fig. 1): 1) a clade containing the SOSTDC1 and SOST genes; 2) a clade that contains the CER1 and DAND5 gene lineages; 3) a clade corresponding to 2 cerberus-like sequences of cephalochordates; 4) a clade corresponding to the NBL1 gene; and 5) a clade containing the GREM gene lineages (Fig. 1).

In the first clade, the monophyly of SOSTDC1 and SOST is well supported, as is the sister group relationship between them (shown in shades of red; Fig. 1), suggesting that they share a common ancestor

more recently than with any other member of the family (Supplementary Fig. S1A). The sister group relationship between these two gene lineages has also been recovered in other studies (Avsian-Kretchmer et al. 2004; Walsh et al. 2010; Nolan & Thompson, 2014). Synteny analyses provide further support for the identity of these two gene lineages because genes found up- and downstream are conserved in most surveyed species (Fig. 2). The tree topology in which the SOSTDC1/SOST clade is recovered sister to all other members of the gene family has also been recovered in other studies (Avsian-Kretchmer et al. 2004; Walsh et al. 2010; Nolan & Thompson, 2014).

In the second group, shown in shades of orange (Fig.1), the monophyly of CER1 genes is well supported, whereas the monophyly of DAND5 genes is not (Fig. 1). The lack of support for the DAND5 clade could be due to the presence of a coelacanth DAND5 sequence for which the phylogenetic position is not well resolved (Supplementary Fig. S1B). However, genes found upstream (GADD45GIP1 and RAD23A) and downstream (NFIX, LYL1) of the coelacanth DAND5 sequence confirm that the gene is in the expected position for a DAND5 ortholog, providing independent support for the topology obtained in our phylogenetic analyses (Fig. 1). Thus, the data suggests that the coelacanth DAND5 gene is a true ortholog of the DAND5 gene of vertebrates and that the lack of resolution is likely due to high sequence divergence in the coelacanth gene. In addition to the coelacanth sequence, synteny analyses provide further support for our phylogeny, because genes found up- and downstream of CER1 and DAND5 are conserved across species (Fig. 2). Although we recovered a sister group relationship between CER1 and DAND5 (Fig. 1), the relevant node was not supported (Fig. 1). In the literature, there is no clear pattern regarding the sister group relationship between these paralogs (Avsian-Kretchmer et al. 2004; Walsh et al. 2010; Nolan & Thompson, 2014). Although the study of Walsh et al. (2010) supports the sister group relationship between CER1 and DAND5, the study ofAvsian-Kretchmer et al. (2004) recovered a topology in which CER1 is the sister group of a clade containing DAND5, NBL1 and GREM gene lineages, whereas, Nolan & Thompson (2014) recovered DAND5 sister to a clade containing CER1 and GREM gene lineages. In support of our topology, an amino acid alignment that includes all human DAN family members shows that CER1 and DAND5 are more similar to each other than to any other member of the gene family; moreover, the Ensembl and gene cards platforms suggest that the only paralog of CER1 is DAND5 and vice versa. Finally, according to our analyses, the clade that includes CER1 and DAND5 was recovered sister to the clade that includes the NBL1 and GREM gene lineages and a third clade of DAN genes that includes two cerberus-like sequences from two lancelet species (Fig. 1).

Our phylogenetic analyses recovered a third clade of DAN genes that includes two cerberus-like sequences from two lancelet species (*Branchiostoma floridae* and *B. belcheri*; purple lineage; Fig. 1). These sequences were recovered sister to the GREM/NBL1 clade (Fig. 1). This phylogenetic arrangement would be compatible with two possible evolutionary scenarios. The first suggests that the DAN gene repertoire of lancelets represents the gene complement of the ancestor of Olfactores, from latin olfactus, the group that includes vertebrates and urochordates, and during the radiation of the vertebrates the ancestral gene lineage gave rise to the extant GREM and NBL1 genes. The second scenario implies that the lancelet gene lineage could be an ancient member of the DAN gene family that has only been retained in cephalochordates. In support of the second scenario, we found cephalochordate sequences in the NBL1 and GREM clades (Supplementary Fig. S1C), suggesting that the genomes of chordates possess an already-differentiated member of the GREM and NBL1 clades.

The fourth clade corresponds to the NBL1 gene which was recovered with strong support (pink clade; Fig. 1)(Supplementary Fig. S1C). The conservation of the genes found up- and downstream provides further support for the identity of the NBL1 gene lineage (Fig. 2). The phylogenetic position of NBL1 is still a matter of debate because different phylogenetic hypotheses have been proposed in earlier studies. Nolan et al. (2014) recovered NBL1 as sister to the clade containing GREM1, GREM2, CER, and DAND5 gene lineages; whereas in Walsh et al. (2010) NBL1 sequences were recovered in a trichotomy with the GREM1/GREM2 and CER1/DAND5 clades. However, in support of our study Avsian-Kretchmer et al. (2004) recovered NBL1 as sister to the GREM gene lineages.

The fifth clade corresponds to a well-supported group containing GREM gene lineages (shown in shades of green; Fig. 1). Within this clade, the monophyly of the GREM1 and GREM2 genes is well supported (Fig. 1). As in other cases, synteny analyses support the identity of both gene lineages (Fig. 3). We identified a clade that contains a single copy gene in lancelets and tunicates, which in turn was recovered as sister to the GREM2 clade (Supplementary Fig. S1D). The presence of a single copy gene in cephalochordates and urochordates is expected given that a previous study reported that the vertebrate GREM genes diversified as a product of the two rounds of whole-genome duplications early in vertebrate evolution (Singh et al. 2015; Sacerdot et al. 2018; Simakov et al. 2020), and cephalochordates and urochordates diverged from vertebrates before these whole-genome duplications. Unexpectedly, our phylogenetic analyses identified the presence of a third GREM gene lineage (GREM3) among gnathostome vertebrates that has not been described before (Fig. 1). This new gene lineage was recovered with strong support, and was identified in three distantly related species: elephant shark (*Callorhinchus milii*), spotted gar (*Lepisosteus oculatus*) and

coelacanth (*Latimeria chalumnae*). Synteny analyses provide further support for the identity of this new lineage as genes found up- (RASGRP4, FAM98C and SPRED3) and downstream (RYR1, MAP4K1, and EIF3K) are conserved (Fig. 3). Moreover, we found that human orthologs of the genes syntenic to GREM3 mapped to chromosome 19, suggesting that this chromosome as the putative genomic location of the GREM3 gene in humans (Fig. 3). A similar differential retention of an ancestral gene in a small group of distantly related species has also been observed in other gene families (Wichmann et al. 2016).

It is necessary to highlight that although we report disagreement in tree topologies between published studies and our results, phylogenetic trees are not always directly comparable as they have differences in the taxonomic and/or membership sampling, two factors known to affect phylogenetic inference.

Definition of ancestral gene repertoires

The interpretation of the phylogenetic distribution of genes in combination with the evolutionary relationships among them (i.e. gene tree), in the light of the organismal phylogeny (i.e. species tree), represents an appropriate way to understand the evolution of the DAN gene family. According to our results, GREM1, GREM2, SOST, SOSTDC1, and NBL1 are present in the major groups of gnathostomes (Fig. 4; Supplementary Fig. S1E), and within each paralog clade the species arrangement does not significantly deviate from the organismal phylogeny. This indicates that all these genes were present in the common ancestor of the group, and agrees with the expectations of the simplest model of multigene family diversification, the divergent evolution model (Nei et al. 1997). In the case of GREM3, CER1, and DAND5, although they are not found in all major groups of gnathostomes, their phyletic distribution in combination with the corresponding tree topology (Fig. 4) indicate that these genes were also present in the common ancestor of gnathostomes. Thus, our results suggest that the last common ancestor of gnathostomes, dated between 615 to 476 mya (Hedges et al. 2015), possessed a repertoire of eight DAN genes: GREM1, GREM2, SOST, SOSTDC1, NBL1, GREM3, CER1, and DAND5. We note that the ancestral condition of eight genes is only present today in coelacanths (Fig. 4).. This reconstruction holds under any of the alternative rootings previously proposed in the literature.

Inferences about the repertoire of the common ancestors of vertebrates, olfactores and chordates are somewhat speculative given the limited availability of genomes and the quality of current assemblies. However, the phylogenetic position of key taxonomic groups, such as cyclostomes, urochordates, and cephalochordates, provides significant information to advance hypotheses regarding the complement of

genes in these different ancestors. We identified GREM2 and SOSTDC1 sequences in cyclostomes (Supplementary Figs. S1A and S1D), indicating that at least these two genes were likely present in the common ancestor of vertebrates. However, given that cyclostome sequences are nested within the GREM2 and SOSTDC1 clades and not sister to a clade with more than one DAN gene, it is possible that the absence of the other paralogs in cyclostome genomes is due to gene loss.

Inferences about the common ancestor of olfactores are restricted given the limited number of urochordate genomes and the quality of their assemblies. To further complicate this situation, urochordates have experienced an extraordinary number of gene losses (Dehal et al. 2002; Cañestro et al. 2003; Cañestro et al. 2013; Albalat & Cañestro, 2016), so it is difficult to tell whether the absence of a gene is due to the quality of the data or a real loss. Taking all these uncertainties into account, we were able to identify one urochordate sequence. This sequence was recovered as sister to three cephalochordate GREM genes, which in turn was recovered as sister to GREM2 (Supplementary Fig. S1D). Given that GREM genes diversified as a product of the two rounds of whole genome duplications in the last common ancestor of vertebrates (Singh et al. 2015; Sacerdot et al., 2018; Simakov et al. 2020), a single annotated gene copy in urochordates is expected. Thus, the most probable scenario is that the common ancestor of olfactores possessed one GREM gene. It is not possible to draw inferences for the other members of the gene family at this time. However, given that the ancestor of olfactores existed before the two vertebrate-specific whole genome duplications, we expect its repertoire to have fewer genes compared to the ancestors of vertebrates or gnathostomes.

Finally, we found eight cephalochordate sequences that help us define the repertoire of genes in the ancestor of chordates. One sequence was recovered sister to a lamprey SOSTDC1 gene, a clade that in turn was placed sister to all other SOSTDC1 sequences (Supplementary Fig. S1A). A group of two sequences was recovered sister to the clade that includes NBL1 and GREM genes, and we think this clade represents an ancient member of the gene family that was only retained in this group. Two additional sequences were recovered as sister to the NBL1 lineage; whereas the last three were placed sister to a urochordate GREM sequence, and this latter clade, in turn, was recovered sister to GREM2 (Supplementary Fig. S1C). Our results agree with Le Petillon et al. (2013), who also found copies of NBL1 and GREM orthologs in cephalochordates. However, in contrast to Le Petillon et al. (2013), we did not find a cephalochordate DAND5 gene lineage but instead identified a cephalochordate-specific gene lineage (Fig. 1, purple clade). Bringing these results together, we propose that the chordate ancestor had at least four DAN genes (SOSTDC1, NBL1, GREM, DAND5 and the cephalochordate-specific gene lineage). Le Petillon et al. (2013) also identified sequences in ambulacraria, the group that includes echinoderms and hemichordates, suggesting that the deuterostome

ancestor, which existed between 797 and 684 mya (Hedges et al. 2015), had at least three DAN sequences, NBL1, GREM, and DAND5.

Differential retention of DAN genes during the evolutionary history of gnathostomes

Our results show that copies of GREM1, GREM2, SOST, SOSTDC1, and NBL1 have been maintained in the genome of all major groups of gnathostomes during the last 615 mya (Hedges et al. 2015; Fig. 4), which could suggest they play critical roles during development. In line with this suggestion, the inactivation or deletion of some of these genes produce embryonically lethal effects (Khokha et al. 2003; Gazzero et al. 2006); in other cases, the loss of a gene is associated with pathological conditions (Li et al. 2008), although there are other examples where knockouts develop minor morphological anomalies with no significant consequences (Davis et al. 2015; Voguel et al. 2015). Although these genes are present in all major groups of gnathostomes, some of them could still be absent in a particular group within these more inclusive taxonomic categories. For example, it has been suggested that the GREM2 gene was lost in the common ancestor of ruminants, hippopotamuses, and cetaceans between 56.3 and 63.5 million years ago as a product of a chromosomal rearrangement (Opazo et al. 2017).

Our results also revealed that GREM3, CER1 and DAND5 were differentially retained during the evolutionary history of gnathostomes. The differential retention of genes could be a stochastic process, in which the resulting differences in gene complement do not translate into functional consequences. This phenomenon could be facilitated by a degree of redundancy that could function as a backup with functionally overlapping paralogues, in case that one of the genes is lost or inactivated (Gitelman, 2007; Cañestro et al., 2009; Félix & Barkoulas, 2015; Albalat & Cañestro, 2016). However, it is also possible that possessing multiple copies could help direct the trajectory of physiological evolution by providing opportunities for the emergence of biological novelty (Ohno et al., 1968; Ohno, 1970; Force et al., 1999; Hughes, 1994; Zhang, 2003). For example, the differential retention and duplication of the γ-globin genes might have played an important role in the evolution of life history in anthropoid primates by facilitating an extended fetal development (Goodman et al. 1987; Opazo et al. 2008). Similarly, the differential retention of functional copies of the INSL4 gene may have been associated with unique reproductive characteristics in catarrhine primates (Arroyo et al. 2012a,b).

From another perspective, the absence of genes in natural knockouts represents an opportunity to understand their physiological role (Albertson et a. 2009). This suggestion is based on the orthology-function conjecture: the expectation that orthologous genes are most likely to have equivalent functions in different

organisms (Altenhoff et al. 2012). This approach has been useful for understanding human diseases, especially for those exhibiting simple Mendelian inherence, such as cystic fibrosis, albinism and other diseases (Albertson et al. 2009). For example, it has been shown that in the blind cavefish (*Astyanax mexicanus*), loss of pigmentation is produced by mutations inactivating the protein encoded by the OCA2 gene, which is also the most frequently mutated gene in cases of human albinism (Protas et al. 2006). Thus, both retention and gene loss could be seen as evolutionary events that help us to understand the genetic bases of biological diversity.

Our results point to the GREM3 gene as the most extreme case of differential retention among DAN genes. It has been retained in cartilaginous fish, holostean fish and coelacanths (Fig. 4), suggesting that it was present in the common ancestor of gnathostomes, between 615 and 473 mya (Hedges et al. 2015), and was subsequently lost independently in the common ancestors of tetrapods and teleost fish (Fig. 5). A similar case of differential retention of a newly discovered gene lineage was shown in a tumor suppressor gene family (Wichmann et al. 2016). In this case, in addition to the elephant shark, spotted gar and coelacanth, the newly discovered gene lineage was also retained by teleost fish (Wichmann et al. 2016). The absence of GREM3 in a significant fraction of gnathostome species suggests that GREM3 could be dispensable. However, given that GREM3 represents a new gene lineage of unknown biological function, it is premature to evaluate the physiological and evolutionary impacts of retaining or losing this gene. Experimental evidence for the other members of the clade (GREM1 and GREM2) provides conflicting results (Khokha et al. 2003; Davis et al. 2015; Voguel et al. 2015), making it difficult to anticipate the degree of dispensability of the related GREM3 gene.

CER1 was also present in the common ancestor of gnathostomes and was differentially retained during the radiation of the group (Fig. 4), being lost in the common ancestor of ray-finned fish (Actinopterygii) between 425 and 375 mya (Betancur-R et al. 2013)(Fig. 4). Synteny analyses provide further support for our scenario because the genomic context of CER1 is conserved in fish (Fig. 5). CER1 is a protein possessing nine conserved cysteines and a cysteine knot region expressed during early gastrulation in the primitive endoderm and subsequently in somites and anterior presomitic mesoderm (Shawlot et al. 1998; Pearce et al. 1999; Piccolo et al. 1999; Belo et al. 2000). CER1 has developmental roles associated primarily with anterior-posterior, left-right, and dorso-ventral asymmetries (Belo et al. 2009), accomplishing them by binding to BMPs, Nodal and Wnt ligands, which in turn block the ligand-receptor interaction and activation (Bouwmeester et al. 1996; Piccolo et al. 1999; Belo et al. 2000; Silva et al. 2003; Avsian 2004). The physiological significance of this gene seems to vary with the taxonomic group: for example, CER1 produces

stronger phenotypic effects in amphibians than in mammals (Simpson et al. 1999; Piccolo et al. 1999; Belo et al. 2000). In fish, CER1 developmental functions may be performed by its paralog DAND5, which is functionally similar (Hashimoto et al. 2004; Marques et al. 2004; Belo et al. 2009; Lopes et al. 2010; Schweickert et al. 2010; Hamada et al. 2012; Araujo et al. 2014). Thus, from an evolutionary perspective, the possession of both paralogs, CER1 and DAND5, in gnathostomes other than fish would represent a case of functional redundancy. Under this scenario, the loss of one of these genes could be compensated by the retention of the other.

In the case of DAND5, we found this gene in all main lineages of gnathostomes other than sauropsids and cartilaginous fish (Fig. 4). In the case of the elephant shark, we believe that the absence of the gene could be due to a problem in the current genome assembly; accordingly, our proposed scenario mainly assumes the absence of this gene in sauropsids. Our results indicate that DAND5 was lost in the common ancestor of sauropsids, the group including birds and non-avian reptiles, between 312 and 280 mya (Hedges et al. 2015; Fig. 5), even though the corresponding genomic region is well conserved (Fig. 6). Our proposed loss of DAND5 in birds has also been noticed by other authors (Le Petillon et al. 2013). DAND5 has been described as a gene playing fundamental roles driving asymmetries in the early stages of development by inhibiting Nodal activity (Marques et al. 2004). Although in amphibians, fish, and mammals the expression pattern seems to be different, the molecular mechanism, restricting nodal signaling in the left side of the embryo is conserved (Belo et al. 2009). In addition to the roles in early development, during adulthood DAND5 also helps promote cancer metastasis, as well as reactivation of metastatic cells in lungs by reversing the ability of BMP to inhibit cancer stem cell function (Gu et al. 2012). Experiments in other organs, such as bone and brain, show that cancer-related functions of DAND5 are specific to the lung (Gu et al. 2012). Consequently, it has been shown that patients expressing high levels of DAND5 have overall reduced survival rates (Gu et al. 2012). The lack of this gene in sauropsids is difficult to interpret; we speculate that, here, the loss of DAND5 is compensated by CER1, removing any physiological consequences and consistent with the idea that gene families possess some degree of redundancy (Gitelman, 2007; Cañestro et al., 2009; Félix & Barkoulas, 2015; Albalat & Cañestro, 2016). Thus, CER1 and DAND5 apparently perform similar physiological roles during development (Hashimoto et al. 2004; Belo et al. 2009).

Rates of molecular evolution: the mammalian slowdown

The phylogenetic trees indicate that the rates of molecular evolution, as measured by branch lengths, vary within each paralog clade (Fig. 1). The most notorious case is the observed slowdown in branches leading to mammals relative to sauropsids (Supplementary Fig. S1A to S1D). This is interesting because sauropsids such as birds, crocodilians, lizards, snakes and turtles tend to have genome-wide rates of evolution much lower than in mammals (Green et al. 2014). Using the relative rate test in MEGA 7 (Kumar et al. 2016) we confirmed that rates in amino acid change are significantly slower in mammals relative to sauropsids (Supplementary Table S2). In mammals, the divergence values ranged from 0.0112 (SOSTDC1) to 0.5143 (DAND5) substitutions per site, whereas in sauropsids they ranged from 0.0326 (GREM2) to 0.9248 (DAND5). In all cases, except for GREM2, the saurospid divergences were higher than in mammals, and the ratio of rates in sauropsids vs. mammals varied from 1.79 (DAND5) to 14.66 (GREM1). We expect genes expressed in early stages of life to be subject to more stringent purifying selection than those expressed in later development, and therefore to have lower evolutionary rates (Goodman, 1961, 1963; Roux & Robinson-Rechavi, 2008). Thus, the slower rate of evolution in mammals could reflect an increased level of functional or structural constraints on these genes. A slowdown in a specific set of genes should be distinguished from a lineage-specific slowdown. In the first case, the reduced evolutionary rate is probably due to differences in the strength of purifying selection between the different lineages, whereas lineage-specific slowdowns are more plausibly be explained by changes in the substitution rate, which could be reduced by lowering of the mutation rate by improved systems of DNA repair mechanisms, longer generation times, or other factors (Goodman, 1985).

Conclusions

We performed an evolutionary study of the differential screening-selected gene aberrant in neuroblastoma (DAN) gene family in vertebrates. According to our results, this gene family has evolved by a combination of models including divergent evolution, birth-and-death, and differential retention. We recovered the monophyly of all recognized gene family members, which in turn were arranged into five main clades. Importantly, in this work we described the presence of two new DAN gene lineages; one that is only present in cephalochordates (e.g. amphioxus), and other (GREM3) that was only identified in cartilaginous fish, holostean fish and coelacanths. The ancestor of gnathostome vertebrates possessed a repertoire of eight DAN genes. During the radiation of the group, some of them (GREM1, GREM2, SOST, SOSTDC1 and NBL1) were retained in the genome of all major groups, while others (GREM3, CER1 and DAND5) were differentially

retained during the evolutionary history of the group. The rate of molecular evolution of mammals is in general low, suggesting an increased evolutionary constraint regime. Finally, we expect that with all these new information researchers in the biomedical field could put their results in an evolutionary perspective.

Data accessibility

Data and supplementary material are available online: https://zenodo.org/deposit/3911345

Acknowledgements

This work was supported by the Fondo Nacional de Desarrollo Científico y Tecnológico from Chile (FONDECYT 1160627) and Millennium Nucleus of Ion Channels Associated Diseases (MiNICAD), Iniciativa Científica Milenio, Ministry of Economy, Development and Tourism from Chile to JCO, National Science Foundation (EPS-0903787, DBI-1262901 and DEB-1354147) to F.G.H. S.V.E. acknowledges NSF grant DEB 1355343. Version 2 of this preprint has been peer-reviewed and recommended by Peer Community In Evol Biol (https://doi.org/10.24072/pci.evolbiol.100104)

Conflict of interest disclosure

The authors of this preprint declare that they have no financial conflict of interest with the content of this article. Juan C. Opazo is recommender for PCI Evolutionary Biology and PCI Genomics.

References

Albalat R., Cañestro C. 2016. Evolution by gene loss. *Nature Reviews Genetics* 17:379–391. DOI: 10.1038/nrg.2016.39.

Albertson RC., Cresko W., Detrich HW., Postlethwait JH. 2009. Evolutionary mutant models for human disease. *Trends in Genetics*. DOI: 10.1016/j.tig.2008.11.006.

Altenhoff AM., Studer RA., Robinson-Rechavi M., Dessimoz C. 2012. Resolving the ortholog conjecture: Orthologs tend to be weakly, but significantly, more similar in function than paralogs. *PLoS Computational Biology*. DOI: 10.1371/journal.pcbi.1002514.

Altschul SF., Gish W., Miller W., Myers EW., Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215:403–410. DOI: 10.1016/S0022- 2836(05)80360-2.

Amigo JD., Id JCO., Jorquera R., Wichmann IA., Garcia-bloj BA., Alarcon MA., Owen GI., Corval AH. 2018. The Reprimo Gene Family : A Novel Gene Lineage in Gastric Cancer with Tumor Suppressive Properties. :1–15. DOI: 10.3390/ijms19071862.

Araujo AC., Marques S., Belo JA. 2014. Targeted inactivation of cerberus like-2 leads to left ventricular cardiac hyperplasia and systolic dysfunction in the mouse. *PLoS ONE*. DOI: 10.1371/journal.pone.0102716.

Arroyo JI., Hoffmann FG., Good S., Opazo JC. 2012. INSL4 pseudogenes help define the relaxin family repertoire in the common ancestor of placental mammals. *Journal of Molecular Evolution* 75. DOI: 10.1007/s00239-012-9517-0.

Arroyo JI., Hoffmann FG., Opazo JC. 2012. Gene turnover and differential retention in the relaxin/insulin-like gene family in primates. *Molecular Phylogenetics and Evolution* 63. DOI: 10.1016/j.ympev.2012.02.011.

Arroyo JI., Hoffmann FG., Opazo JC. 2014. Evolution of the relaxin/insulin-like gene family in anthropoid primates. *Genome Biology and Evolution* 6. DOI: 10.1093/gbe/evu023.

Avsian-Kretchmer O., Hsueh AJW. 2004. Comparative Genomic Analysis of the Eight- Membered Ring Cystine Knot-Containing Bone Morphogenetic Protein Antagonists. *Molecular Endocrinology* 18:1–12. DOI: 10.1210/me.2003-0227.

Belo JA., Bachiller D., Agius E., Kemp C., Borges AC., Marques S., Piccolo S., De Robertis EM. 2000. Cerberus-like is a secreted BMP and nodal antagonist not essential for mouse development. *Genesis*. DOI: 10.1002/(SICI)1526- 968X(200004)26:4<265::AID-GENE80>3.0.CO;2-4.

Belo JA., Silva AC., Borges AC., Filipe M., Bento M., Gonçalves L., Vitorino M., Salgueiro AM., Texeira V., Tavares AT., Marques S. 2009. Generating asymmetries in the early vertebrate embryo: The role of the Cerberus-like family. *International Journal of Developmental Biology*. DOI: 10.1387/ijdb.072297jb.

Belyi VA., Ak P., Markert E., Wang H., Hu W., Puzio-Kuter A., Levine AJ. 2010. The origins and evolution of the p53 family of genes. *Cold Spring Harbor perspectives in biology*. DOI: 10.1101/cshperspect.a001198.

Betancur-R. R, Broughton RE, Wiley EO, Carpenter K, López JA, Li C, Holcroft NI, Arcila D, Sanciangco M, Cureton II JC, Zhang F, Buser T, Campbell MA, Ballesteros JA, Roa-Varon A, Willis S, Borden WC, Rowley T, Reneau PC, Hough DJ, Lu G, Grande T, Arratia G, Ortí G. The Tree of Life and a New Classification of Bony Fishes. PLOS Currents Tree of Life. 2013 Apr 18 . Edition 1. doi: 10.1371/currents.tol.53ba26640df0ccaee75bb165c8c26288.

Bieging KT., Mello SS., Attardi LD. 2014. Unravelling mechanisms of p53-mediated tumour suppression.

Nature Reviews Cancer 14:359–370. DOI: 10.1038/nrc3711.

Bouwmeester T., Kim S., Sasai Y., Lu B., De Robertis E. 1996. Cerberus is a head- inducing secreted factor expressed in the anterior endoderm of Spemann's organizer. *Nature* 382:595–601. DOI: 10.1038/382595a0.

Braidy N., Poljak A., Jayasena T., Mansour H. 2015. Accelerating Alzheimer ' s research through ' natural ' animal models. :155–164. DOI: 10.1097/YCO.0000000000000137.

Cañestro C., Albalat R., Irimia M., Garcia-Fern??ndez J. 2013. Impact of gene gains, losses and duplication modes on the origin and diversification of vertebrates. *Seminars in Cell and Developmental Biology* 24:83–94. DOI: 10.1016/j.semcdb.2012.12.008.

Cañestro C., Bassham S., Postlethwait JH. 2003. Seeing chordate evolution through the Ciona genome sequence. *Genome Biology* 4.

Cañestro C., Catchen JM., Rodriguez-Mari A., Yokoi H., Postlethwait JH. 2009. Consequences of Lineage-Specific Gene Loss on Functional Evolution of Surviving Paralogs : ALDH1A and Retinoic Acid Signaling in Vertebrate Genomes. *PLoS Genetics* 5:1–19. DOI: 10.1371/journal.pgen.1000496.

Chiodelli P., Mitola S., Ravelli C., Oreste P., Rusnati M., Presta M. 2011. Heparan sulfate proteoglycans mediate the angiogenic activity of the vascular endothelial growth factor receptor-2 agonist gremlin. *Arteriosclerosis, Thrombosis, and Vascular Biology*. DOI: 10.1161/ATVBAHA.111.235184.

Chu P-Y., Zhuo Y-X., Wang F-I., Jeng C-R., Pang VF., Chang P-H., Chin S-C., Liu C-H. 2012. Spontaneous neoplasms in zoo mammals, birds, and reptiles in Taiwan a 10-year survey. *Animal Biology* 62:95–110. DOI: 10.1163/157075611X616941.

Davis H., Irshad S., Bansal M., Rafferty H., Boitsova T., Bardella C., Jaeger E., Lewis A., Freeman-Mills L., Giner F., Rodenas-Cuadrado P., Mallappa S., Clark S., Thomas H., Jeffery R., Poulsom R., Rodriguez-Justo M., Novelli M., Chetty R., Silver A., Sansom O., Greten F., Wang L., East J., Tomlinson I., Leedham S. 2015. Aberrant epithelial GREM1 expression initiates colonic tumorigenesis from cells outside the stem cell niche. *Nature Medicine* 21:62–70. DOI: 10.1038/nm.3750.

Dehal P., Satou Y., Campbell RK., Chapman J., Degnan B., De Tomaso A., Davidson B., Gregorio A Di., Gelpke M., Goodstein DM., Harafuji N., Hastings KEM., Ho I., Hotta K., Huang W., Kawashima T., Lemaire P., Martinez D., Meinertzhagen IA., Necula S., Nonaka M., Putnam N., Rash S., Saiga H., Satake M., Terry A., Yamada L., Wang H-G., Awazu S., Azumi K., Boore J., Doyle S., Francino P., Keys DN., Haga S., Hayashi H., Hino K., Imai KS., Kobayashi K., Kobayashi M., Lee B-I., Makabe KW., Manohar C., Matassi G., Medina M., Mochizuki Y., Sasaki A., Sasakura Y., Shoguchi E., Shin-I T., Spagnuolo A., Stainier D., Suzuki MM., Tassy O., Takatori N.,

Tokuoka M., Yagi K., Yoshizaki F., Wada S., Zhang C., Hyatt PD., Larimer F., Detter C., Doggett N., Glavina T., Hawkins T., Richardson P., Lucas S., Kohara Y., Levine M., Satoh N., Rokhsar DS. 2002. The Draft Genome of Ciona intestinalis: Insights into Chordate and Vertebrate Origins. *Science* 298:2157–2167. DOI: 10.1126/science.1080049.

Edwards S V. 2009. Is a new and general theory of molecular systematics emerging ? :1–19. DOI: 10.1111/j.1558-5646.2008.00549.x.

Effron M., Griner L., Benirschke K. 1977. Nature and Rate of Neoplasia Found in Captive Wild Mammals, Birds, and Reptiles at Necropsy 1. *Journal of the National Cancer Institute* 59:185–198. Ellies DL., Viviano B., Mccarthy J., Rey J-P., Itasaki N., Saunders S., Krumlauf R. 2006. Bone Density Ligand, Sclerostin, Directly Interacts With LRP5 but Not LRP5 G171V to Modulate Wnt Activity. *J Bone Miner Res* 21:1738–1749. DOI: 10.1359/JBMR.060810.

Félix M., Barkoulas M. 2015. Pervasive robustness in biological systems. *Nature Reviews Genetics* 16:483–496. DOI: 10.1038/nrg3949.

Force A., Lynch M., Pickett FB., Amores A., Yan Y., Postlethwait J. 1999. Preservation of Duplicate Genes by Complementary , Degenerative Mutations. *Genetics* 151:1531–1545.

Fox H. 1910. Observations upon neoplasms in wild animals in the philadelphia zoological gardens.l. *Proc. Path. Soc. Bila*.

Gao H., Chakraborty G., Lee-Lim AP., Mo Q., Decker M., Vonica A., Shen R., Brogi E., Brivanlou AH., Giancotti FG. 2012. The BMP inhibitor Coco reactivates breast cancer cells at lung metastatic sites. *Cell*. DOI: 10.1016/j.cell.2012.06.035.

Gazzerro E., Canalis E. 2006. Bone morphogenetic proteins and their antagonists. *Reviews in Endocrine and Metabolic Disorders*. DOI: 10.1007/s11154-006-9000- 6.

Geer LY., Marchler-bauer A., Geer RC., Han L., He J., He S., Liu C., Shi W., Bryant SH. 2010. The NCBI BioSystems database. *Nucleic Acids Research* 38:492–496. DOI: 10.1093/nar/gkp858.

Gitelman I. 1996. Evolution of the Vertebrate Twist Family and Synfunctionalization : A Mechanism for Differential Gene Loss through Merging of Expression Domains. DOI: 10.1093/molbev/msm120.

Goodman M. 1961. The role of immunochemical differences in the phyletic development of human behavior. *Human Biology* 33:131–162.

Goodman M. 1962. Evolution of the immunologic species specificity of human serum proteins. *Human*

Biology 61:104–150.

Goodman M. 1985. Rates of Molecular Evolution: The Hominoid Slowdown. *BioEssays* 3:9–14. DOI: 10.1002/bies.950030104.

Goodman M., Czelusniak J., Koop BF., Tagle DA., Slightom JL. 1987. Globins: a case study in molecular phylogeny. *Cold Spring Harbor symposia on quantitative biology* 52:875–90. DOI: 10.1101/SQB.1987.052.01.096.

Gorbunova V., Hine C., Tian X., Ablaeva J., Gudkov A V., Nevo E. 2012. Cancer resistance in the blind mole rat is mediated by concerted necrotic cell death mechanism. 109:19392–19396. DOI: 10.1073/pnas.1217211109.

Green RE., Braun EL., Armstrong J., Earl D., Nguyen N., Hickey G., Vandewege MW., St John JA., Capella-Gutiérrez S., Castoe TA., Kern C., Fujita MK., Opazo JC., Jurka J., Kojima KK., Caballero J., Hubley RM., Smit AF., Platt RN., Lavoie CA., Ramakodi MP., Finger JW., Suh A., Isberg SR., Miles L., Chong AY., Jaratlerdsiri W., Gongora J., Moran C., Iriarte A., McCormack J., Burgess SC., Edwards SV., Lyons E., Williams C., Breen M., Howard JT., Gresham CR., Peterson DG., Schmitz J., Pollock DD., Haussler D., Triplett EW., Zhang G., Irie N., Jarvis ED., Brochu CA., Schmidt CJ., McCarthy FM., Faircloth BC., Hoffmann FG., Glenn TC., Gabaldón T., Paten B., Ray DA. 2014. Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. *Science* 346. DOI: 10.1126/science.1254449.

Hamada H., Meno C., Watanabe D., Saijoh Y. 2002. Establishment of vertebrate left- right asymmetry. *Nature Reviews Genetics* 3:103–113. DOI: 10.1038/nrg732.

Hashimoto H., Rebagliati M., Ahmad N., Muraoka O., Kurokawa T., Hibi M., Suzuki T. 2004. The Cerberus/Dan-family protein Charon is a negative regulator of Nodal signaling during left-right patterning in zebrafish. *Development* 131:1753. DOI: 10.1242/dev.01070.

Herrero J., Muffato M., Beal K., Fitzgerald S., Gordon L., Pignatelli M., Vilella AJ., Searle SMJ., Amode R., Brent S., Spooner W., Kulesha E., Yates A., Flicek P. 2016. Ensembl comparative genomics resources. *Database* 2016:baw053. DOI: 10.1093/database/baw053.

Hughes AL., Hughes MK., Howell CY., Nei M. 1994. Natural Selection at the Class II Major Histocompatibility Complex Loci of Mammals. *Philosophical Transactions of the Royal Society B* 345:359–367. DOI: 10.1098/rstb.1994.0153.

Katoh K., Standley DM. 2013. MAFFT Multiple Sequence Alignment Software Version 7: Improvements in Performance and Usability. *Molecular Biology and Evolution* 30:772–780. DOI: 10.1093/molbev/mst010.

Khokha MK., Hsu D., Brunet LJ., Dionne MS., Harland RM. 2003. Gremlin is the BMP antagonist required for maintenance of Shh and Fgf signals during limb patterning. *Nature Genetics* 34:303–307. DOI: 10.1038/ng1178.

Koli K., Myllä M., Vuorinen K., Salmenkivi K., Ryynä Nen MJ., Kinnula VL., Keski-Oja J. 2006. Bone Morphogenetic Protein-4 Inhibitor Gremlin Is Overexpressed in Idiopathic Pulmonary Fibrosis. *Cardiovascular, Pulmonary and Renal Pathology*. DOI: 10.2353/ajpath.2006.051263.

Kumar S., Stecher G., Suleski M., Hedges SB. 2017. TimeTree : A Resource for Timelines , Timetrees , and Divergence Times. *Molecular Biology and Evolution* 34:1812–1819. DOI: 10.1093/molbev/msx116.

Kumar S., Stecher G., Tamura K. 2016. MEGA7 : Molecular Evolutionary Genetics Analysis Version 7 . 0 for Bigger Datasets Brief communication. 33:1870–1874. DOI: 10.1093/molbev/msw054.

Kalyaanamoorthy S., Minh BQ., Wong TKF., von Haeseler A., Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetics estimates. Nature Methods. 14:587-589.

Li X., Ominsky MS., Niu Q-T., Sun N., Daugherty B., D ’agostin D., Kurahara C., Gao Y., Cao J., Gong J., Asuncion F., Barrero M., Warmington K., Dwyer D., Stolina M., Morony S., Sarosi I., Kostenuik PJ., Lacey DL., Simonet WS., Ke HZ., Paszty C. 2008. Targeted Deletion of the Sclerostin Gene in Mice Results in Increased Bone Formation and Bone Strength. *J Bone Miner Res* 23:860–869. DOI: 10.1359/JBMR.080216.

Liang W., Guan H., He X., Ke W., Xu L., Liu L., Xiao H., Li Y. 2015. Down-regulation of SOSTDC1 promotes thyroid cancer cell proliferation via regulating cyclin A2 and cyclin E2. *Oncotarget* 6.

Lombard LS., Witte EJ. 1959. Frequency and Types of Tumors in Mammals and Birds of the Philadelphia Zoological Garden. *Cancer Research*.

Lopes SS., Lourenço R., Pacheco L., Moreno N., Kreiling J., Saúde L. 2010. Notch signalling regulates left-right asymmetry through ciliary length control. *Development* 137:3625–3632. DOI: 10.1242/dev.054452.

Louis A., Nguyen NTT., Muffato M., Roest Crollius H. 2015. Genomicus update 2015: KaryoView and MatrixView provide a genome-wide perspective to multispecies comparative genomics. *Nucleic Acids Research* 43:D682–D689. DOI: 10.1093/nar/gku1112.

Manov I., Hirsh M., Iancu TC., Malik A., Sotnichenko N., Band M., Avivi A. 2013. Pronounced cancer resistance in a subterranean rodent , the blind mole-rat , Spalax : in vivo and in vitro evidence Pronounced cancer resistance in a subterranean rodent , the blind mole-rat , Spalax : in vivo and in vitro evidence.

Marques S., Borges AC., Silva AC., Freitas S., Cordenonsi M., Belo JA. 2004. The activity of the Nodal

antagonist Cerl-2 in the mouse node is required for correct L/R body axis. *Genes and Development*. DOI: 10.1101/gad.306504.

Møller AP., Erritzøe J., Soler JJ. 2017. Life history, immunity, Peto's paradox and tumours in birds. *Journal of Evolutionary Biology*. DOI: 10.1111/jeb.13060.

Natarajan C., Jendroszek A., Kumar A., Weber RE., Tame JRH., Fago A., Storz JF. 2018. Molecular basis of hemoglobin adaptation in the high-flying bar-headed goose. *PLoS Genetics* 14. DOI: 10.1371/journal.pgen.1007331.

Nei M., Gu X., Sitnikova T. 1997. Evolution by the birth-and-death process in multigene families of the vertebrate immune system. *Proceedings of the National Academy of Sciences of the United States of America* 94:7799–7806.

Nolan K., Thompson TB. 2014. The DAN family: Modulators of TGF-beta signaling and beyond. *Protein Science*. DOI: 10.1002/pro.2485.

Ohno S. 1970. *Evolution by gene duplication*. Springer-Verlag. Ohno S., Wolf U., Atkin NB. 1968. Evolution from fish to mammals by gene duplication. *Hereditas* 59:169–187.

Opazo JC., Hoffmann FG., Storz JF. 2008. Differential loss of embryonic globin genes during the radiation of placental mammals. *Proceedings of the National Academy of Sciences of the United States of America* 105:12950–5. DOI: 10.1073/pnas.0804392105.

Opazo JC., Palma RE., Melo F., Lessa EP. 2005. Adaptive evolution of the insulin gene in caviomorph rodents. *Molecular biology and evolution* 22:1290–8. DOI: 10.1093/molbev/msi117.

Opazo JC., Zavala K., Krall P., Arias RA. 2017. Evolution of gremlin 2 in cetartiodactyl mammals: Gene loss coincides with lack of upper jaw incisors in ruminants. *PeerJ* 2017. DOI: 10.7717/peerj.2901.

Pearce JJH., Penny GE., Rossant J. 1999. A Mouse Cerberus/Dan-Related Gene Family. *Developmental Biology* 209:98–110. DOI: 10.1006/dbio.1999.9240. Le Petillon Y., Oulion S., Escande ML., Escriva H., Bertrand S. 2013. Identification and expression analysis of BMP signaling inhibitors genes of the DAN family in amphioxus. *Gene Expression Patterns*. DOI: 10.1016/j.gep.2013.07.005.

Piccolo S., Agius E., Leyns L., Bhattacharyya S., Grunz H., Bouwmeester T., De Robertis EM. 1999. The head inducer Cerberus is a multifunctional antagonist of Nodal, BMP and Wnt signals Embryological and genetic evidence indicates that the vertebrate head is induced by a different set of signals from those that organize trunk±tail development. *Nature* 397. DOI: 10.1038/17820.

Protas ME., Hersey C., Kochanek D., Zhou Y., Wilkens H., Jeffery WR., Zon LI., Borowsky R., Tabin CJ. 2006. Genetic analysis of cavefish reveals molecular convergence in the evolution of albinism. *Nature Genetics* 38:107–111. DOI: 10.1038/ng1700.

Rambaut A., Drummond AJ., Xie D., Baele G., Suchard MA. 2018. Posterior summarization in Bayesian phylogenetics using tracer 1.7. Sys. Bio. 67: 901-904.

Ratcliffe HL. 1933. Incidence and nature of tumors in captive wild mammals and birds. *American Journal of Cancer*. DOI: 10.1158/ajc.1933.116.

Ronquist F., Huelsenbeck JP. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics (Oxford, England)* 19:1572–4.

Roux J., Robinson-Rechavi M. 2008. Developmental constraints on vertebrate genome evolution. *PLoS Genetics*. DOI: 10.1371/journal.pgen.1000311.

Sacerdot C., Louis A., Bon C., Berthelot C., Crollius HR. 2018. Chromosome evolution at the origin of the ancestral vertebrate evolution. *Genome Biology. 166*

Schweickert A., Vick P., Getwan M., Weber T., Schneider I., Eberhardt M., Beyer T., Pachur A., Blum M. 2010. The Nodal Inhibitor Coco Is a Critical Target of Leftward Flow in Xenopus. *Current Biology*. DOI: 10.1016/j.cub.2010.02.061.

Shawlot W., Deng M., Behringer RR. 1998. Expression of the mouse cerberus- related gene, Cerr1, suggests a role in anterior neural induction and somitogenesis. *Developmental Biology* 95:6198–6203.

Silva AC., Filipe M., Kuerner K-M., Steinbeisser H., Belo JA. 2003. Endogenous Cerberus activity is required for anterior head specification in Xenopus. *Development* 130:4953. DOI: 10.1242/dev.00705.

Simakov O., Marlétaz F., Yue J. *et al.* 2020. Deeply conserved synteny resolves early events in vertebrate evolution. Nat Ecol Evol 4: 820–830. https://doi.org/10.1038/s41559-020-1156-z

Simpson EH., Johnson AK., Hunsicker P., Suffolk R., Jordan SA., Jackson IJ. 1999. The Mouse Cer1 (Cerberus related or homologue) Gene Is Not Required for Anterior Pattern Formation. *Devedlopmental Biology* 213:202–206. DOI: 10.1006/dbio.1999.9372.

Singh PP., Arora J., Isambert H. 2015. Identification of Ohnolog Genes Originating from Whole Genome Duplication in Early Vertebrates, Based on Synteny Comparison across Multiple Genomes. *PLOS Computational Biology* 11:e1004394. DOI: 10.1371/journal.pcbi.1004394.

Tatusova TA., Madden TL. 1999. BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. *FEMS microbiology letters* 174:247–50.

Trifinopoulos J., Nguyen L-T., von Haeseler A., Minh BQ. 2016. W-IQ-TREE: a fast online phylogenetic tool for maximum likelihood analysis. *Nucleic Acids Research* 44:W232–W235. DOI: 10.1093/nar/gkw256.

Vogel P., Liu J., Platt KA., Read RW., Thiel M., Vance RB., Brommage R. 2015. Malformation of Incisor Teeth in Grem2 - / - Mice. 52:224–229. DOI: 10.1177/0300985814528218.

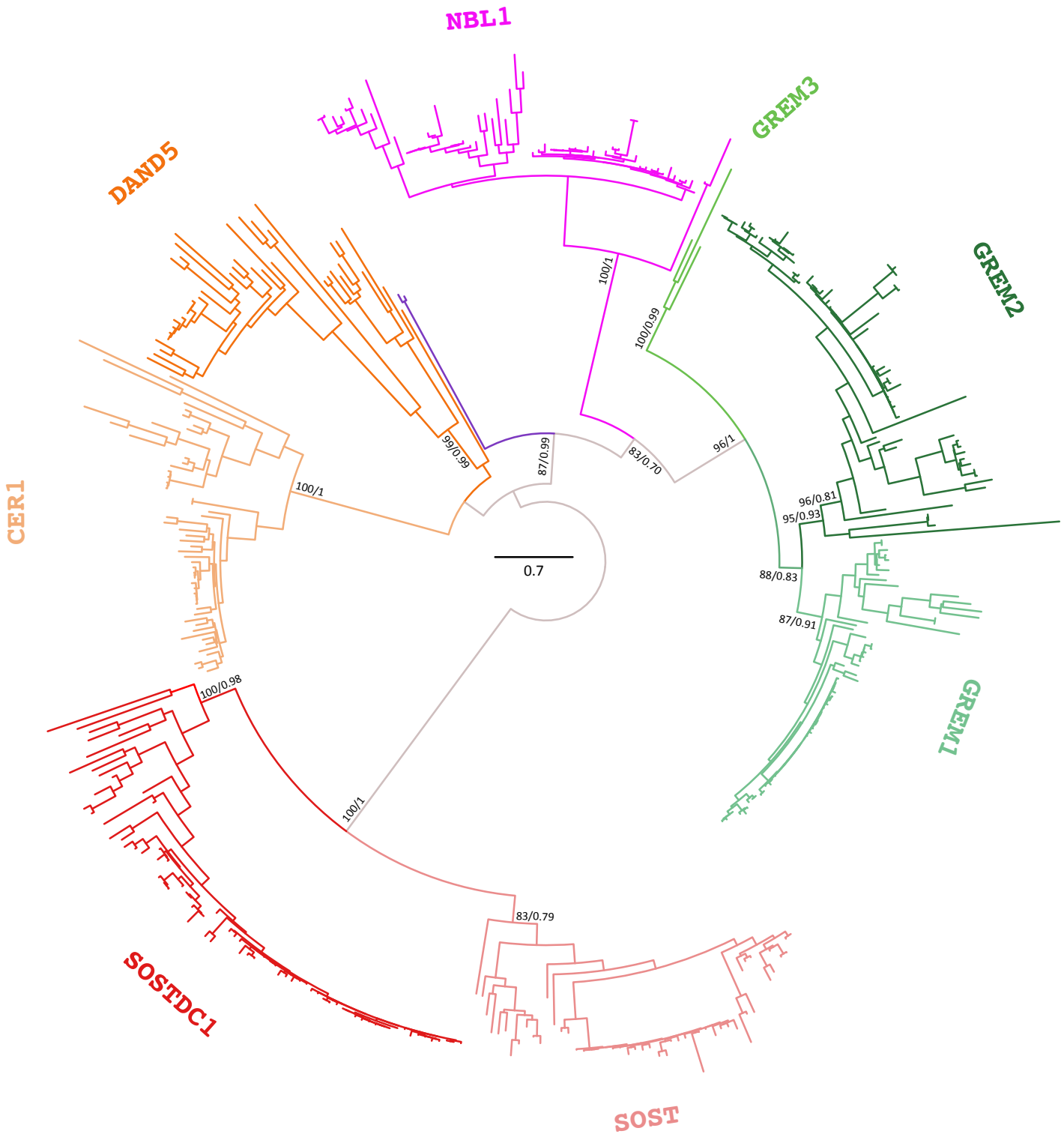
Walsh DW., Godson C., Brazil DP., Martin F. 2010. Extracellular BMP-antagonist regulation in development and disease: Tied up in knots. *Trends in Cell Biology*. DOI: 10.1016/j.tcb.2010.01.008.

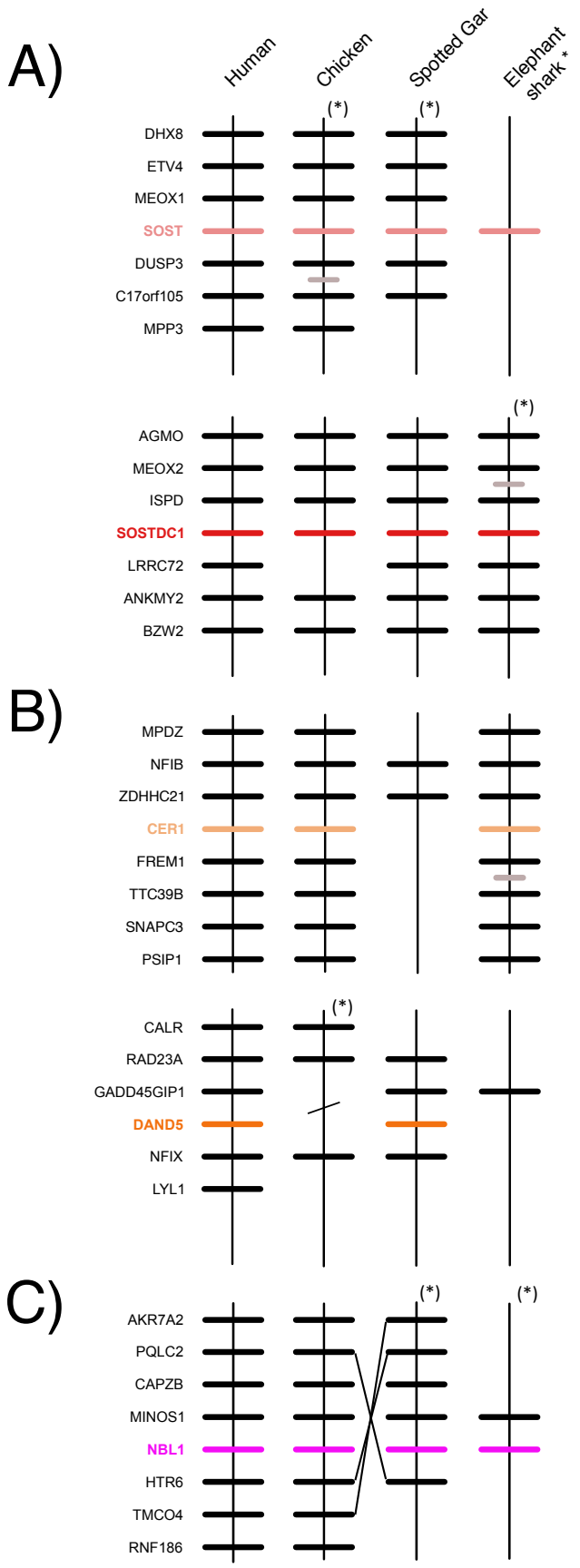
Wichmann IA., Zavala K., Hoffmann FG., Vandewege MW., Corvalán AH., Amigo JD., Owen GI., Opazo JC. 2016. Evolutionary history of the reprimo tumor suppressor gene family in vertebrates with a description of a new reprimo gene lineage. *Gene* 591. DOI: 10.1016/j.gene.2016.07.036.

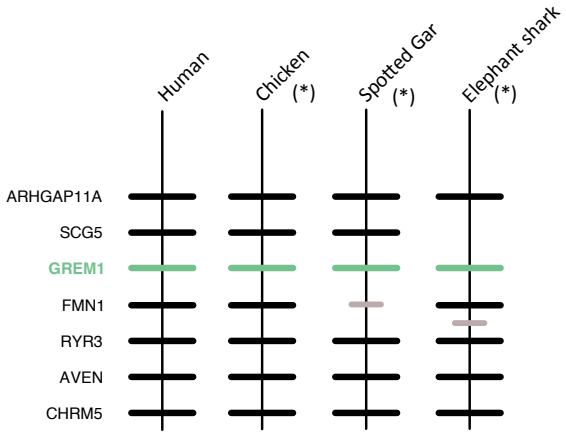
Yang Z., Rannala B. 2012. Molecular phylogenetics: principles and practice. 13. DOI: 10.1038/nrg3186.

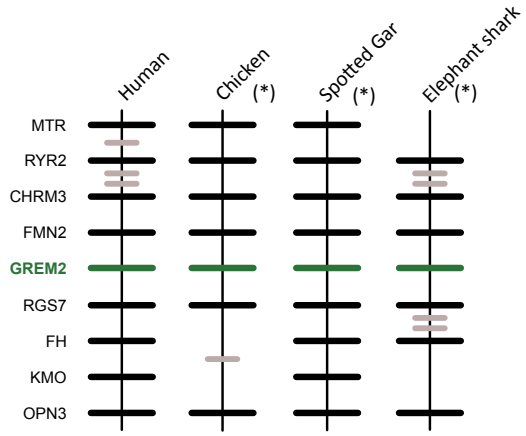
Zerbino DR., Achuthan P., Akanni W., Amode MR., Barrell D., Bhai J., Billis K., Cummins C., Gall A., Gil L., Gordon L., Haggerty L., Haskell E., Hourlier T., Izuogu OG., Janacek SH., Juettemann T., To K., Laird MR., Lavidas I., Liu Z., Loveland JE., Maurel T., Mclaren W., Moore B., Mudge J., Murphy N., Newman V., Nuhn M., Ogeh D., Ong CK., Parker A., Patricio M., Riat HS., Schuilenburg H., Sheppard D., Sparrow H., Taylor K., Thormann A., Vullo A., Walts B., Zadissa A., Frankish A., Hunt SE., Kostadima M., Langridge N., Martin FJ., Muffato M., Perry E., Ruffier M., Staines DM., Trevanion SJ., Aken BL., Cunningham F., Yates A., Flicek P. 2018. Ensembl 2018. *Nucleic Acids Research* 46:754–761. DOI: 10.1093/nar/gkx1098.

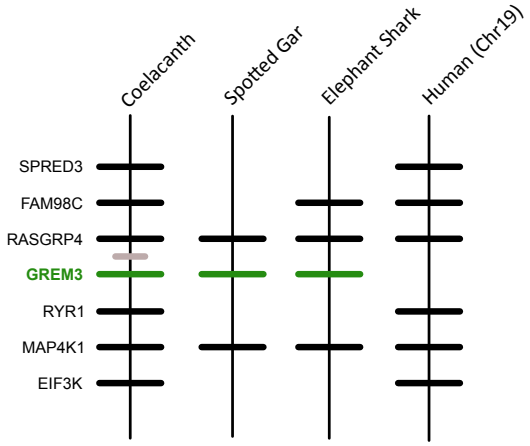
Zhang J. 2003. Evolution by gene duplication : an update. *Trends in Ecology and Evolution* 18:292–298. DOI: 10.1016/S0169-5347(03)00033-8.

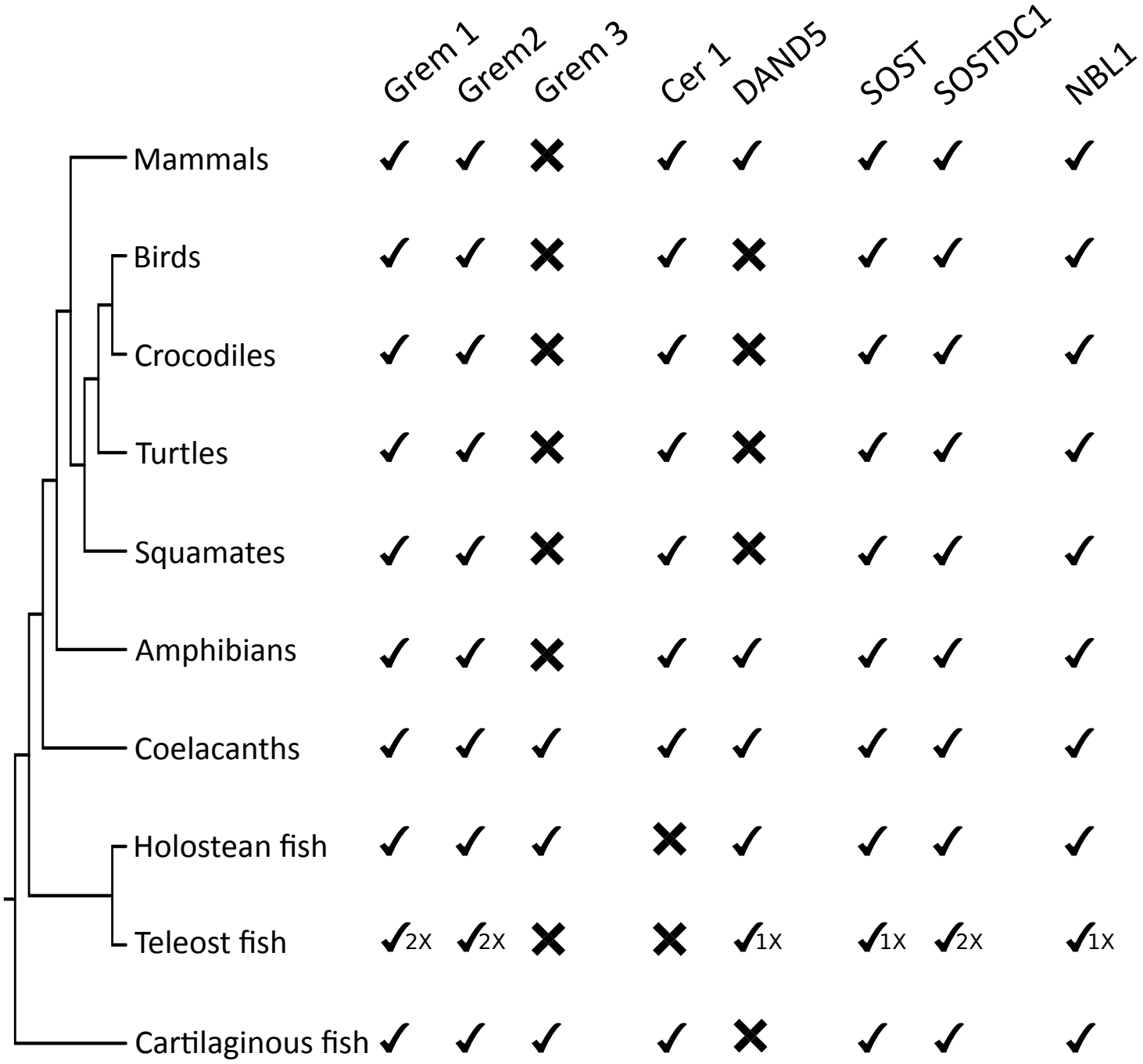

NBL1
GREM3
GREM2
GREM1
SOST
SOSTDC1
CER1
DAND5
100/1
100/0.99
96/1
83/0.70
87/0.99
99/0.99
100/1
96/0.81
95/0.93
88/0.83
87/0.91
0.7
100/0.98
100/1
83/0.79

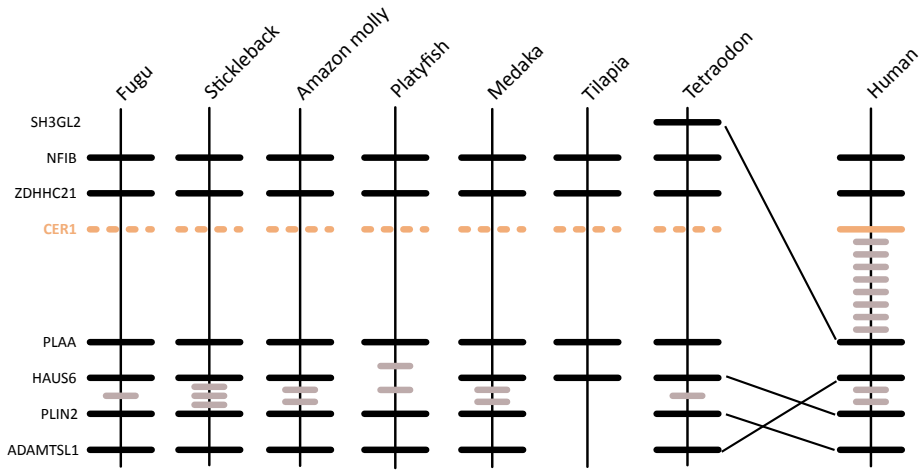

A)
Human
Chicken
Spotted Gar
Elephant shark *
(*)
(*)
DHX8
ETV4
MEOX1
SOST
DUSP3
C17orf105
MPP3
(*)
AGMO
MEOX2
ISPD
SOSTDC1
LRRC72
ANKMY2
BZW2
B)
MPDZ
NFIB
ZDHHC21
CER1
FREM1
TTC39B
SNAPC3
PSIP1
(*)
CALR
RAD23A
GADD45GIP1
DAND5
NFIX
LYL1
C)
(*)
(*)
AKR7A2
PQLC2
CAPZB
MINOS1
NBL1
HTR6
TMCO4
RNF186

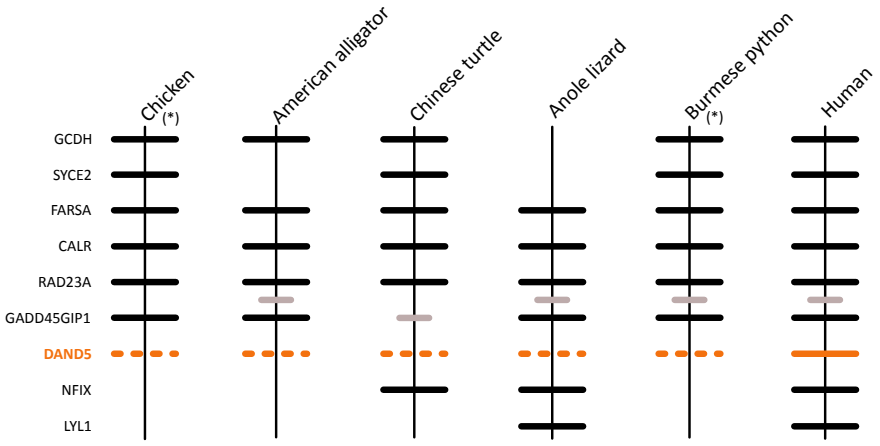

Human
Chicken (*)
Spotted Gar (*)
Elephant shark (*)
ARHGAP11A
SCG5
GREM1
FMN1
RYR3
AVEN
CHRM5


Human
Chicken (*)
Spotted Gar (*)
Elephant shark (*)
MTR
RYR2
CHRM3
FMN2
GREM2
RGS7
FH
KMO
OPN3


Coelacanth
Spotted Gar
Elephant Shark
Human (Chr19)
SPRED3
FAM98C
RASGRP4
GREM3
RYR1
MAP4K1
EIF3K


Grem 1
Grem2
Grem 3
Cer 1
DAND5
SOST
SOSTDC1
NBL1
Mammals ✓ ✓ ✗ ✓ ✓ ✓ ✓ ✓
Birds ✓ ✓ ✗ ✓ ✗ ✓ ✓ ✓
Crocodiles ✓ ✓ ✗ ✓ ✗ ✓ ✓ ✓
Turtles ✓ ✓ ✗ ✓ ✗ ✓ ✓ ✓
Squamates ✓ ✓ ✗ ✓ ✗ ✓ ✓ ✓
Amphibians ✓ ✓ ✗ ✓ ✓ ✓ ✓ ✓
Coelacanths ✓ ✓ ✓ ✓ ✓ ✓ ✓ ✓
Holostean fish ✓ ✓ ✓ ✗ ✓ ✓ ✓ ✓
Teleost fish ✓2X ✓2X ✗ ✗ ✓1X ✓1X ✓2X ✓1X
Cartilaginous fish ✓ ✓ ✓ ✓ ✗ ✓ ✓ ✓


Fugu
Stickleback
Amazon molly
Platyfish
Medaka
Tilapia
Tetraodon
Human
SH3GL2
NFIB
ZDHHC21
CER1
PLAA
HAUS6
PLIN2
ADAMTSL1


Chicken
(*)
American alligator
Chinese turtle
Anole lizard
Burmese python
(*)
Human
GCDH
SYCE2
FARSA
CALR
RAD23A
GADD45GIP1
DAND5
NFIX
LYL1