Cluster expansion of *apolipoprotein D* (*ApoD*) genes in teleost fishes

Langyu Gu[1,2*], Canwei Xia[3]

[1]Key Laboratory of Freshwater Fish Reproduction and Development, Ministry of Education, Laboratory of Aquatic Science of Chongqing, School of Life Sciences, 400715, Southwest University, Chongqing, China. langyugu@gmail.com

[2]Zoological Institute, University of Basel, Vesalgasse 1, 4051, Basel, Switzerland.

[3]Ministry of Education Key Laboratory for Biodiversity and Ecological Engineering, College of Life Sciences, Beijing Normal University, Beijing, China. xiacanwei@126.com

Corresponding author:

*Langyu Gu

langyugu@gmail.com

Abstract

Background: Gene and genome duplication play important roles in the evolution of gene function. Compared to individual duplicated genes, gene clusters attract particular attentions considering their frequent associations with innovation and adaptation. Here, we report for the first time the expansion of the ligand (e.g., pheromone and hormone)-transporter genes, *apolipoprotein D* (*ApoD*) genes in a cluster, specific to teleost fishes.

Results: The single *ApoD* gene in the ancestor expands in two clusters with a dynamic evolutionary pattern in teleost fishes. Based on comparative genomic and transcriptomic analyses, protein 3D structure comparison, evolutionary rate detection and breakpoint detection, orthologous genes show conserved expression patterns. Lineage-specific duplicated genes that are under positive selection evolved specific and even new expression profiles. Different duplicates show high tissue-specific expression patterns (e.g., skin, eye, anal fin pigmentation patterns, gonads, gills, spleen and lower pharyngeal jaw). Cluster analyses based on protein 3D structure comparisons, especially the four loops at the opening side, show segregation patterns with different duplicates. Duplicated *ApoD* genes are predicted to be associated with forkhead transcription factors and MAPK genes, and they are located next to the breakpoints of genome rearrangements.

Conclusions: Here, we report the expansion of *ApoD* genes specific to teleost fishes in a cluster manner for the first time. Neofunctionalization and subfunctionalization were observed at both protein and expression levels after duplication. Evidence from different aspects, i.e. abnormal expression induced disease in human, fish-specific expansion, predicted associations with forkhead transcription factors and MAPK genes, highly specific expression patterns in tissues related to sexual selection and adaptation, duplicated genes that are under positive selection, and their locations next to breakpoints of genome rearrangement, suggests the potential advantageous roles of

ApoD genes in teleost fishes. Cluster expansion of *ApoD* genes specific to teleost fishes thus provides an ideal evo-devo model for studying gene duplication, cluster maintenance and new gene function emergence.

Key words

apolipoprotein D (ApoD), duplication, gene cluster, positive selection, breakpoints, teleost fishes

Background

Gene and genome duplication play important roles in evolution by providing new genetic materials [1]. The gene copies emerging from duplication events (including whole genome duplications (WGD)) can undergo different evolutionary fates, and a number of models have been proposed as to what can happen after duplication [2]. In many instances, one of the duplicates becomes silenced via the accumulation of deleterious mutations (i.e. pseudogenization or nonfunctionalization [1]). Alternatively, the original pre-duplication function can be subdivided between duplicates (i.e. subfunctionalization) [3], or one of the duplicates can gain a new function (i.e. neofunctionalization) [4]. Since the probability of accumulating beneficial substitutions is relatively low, examples for neofunctionalization are sparse. There are, nevertheless, examples for neofunctionalization. For example, the duplication of *dachshund* in spiders and allies has been associated with the evolution of a novel leg segment [5]; the expansion of repetitive regions in a duplicated trypsinogen-like gene led to functional antifreeze glycoproteins in Antarctic notothenioid fish [6]; and the duplication of *opsin* genes is implicated in trichromatic vision in primates [7]. Another selective advantage of gene duplication can be attributed to the increased numbers of gene copies, e.g. in the form of gene dosage effects [8,9]. Multiple mechanisms can act together to shape different phases of gene evolution after duplication [10].

Gene functional changes after duplication can occur at the protein level [6,11,12]. For example, the physiological division of labour between the oxygen-carrier function of haemoglobin and the oxygen-storage function of myoglobin in vertebrates [13] (subfunctionalization); and the acquired enhanced digestive efficiencies of duplicated gene encoding of pancreatic ribonuclease in leaf monkeys [14] (neofunctionalization). However, the chance to accumulate beneficial alleles is low, and thus, functional changes after duplication at the protein level are sparse. Instead, changes in the expression level are more tolerable and efficient, since they do not require modification of coding sequences, and they can immediately offer phenotypic consequences. For example, complementary degenerative mutation in the regulatory regions of duplicated genes is a common mechanism for subfunctionalization [2,15]. Many examples have provided evidence that duplicated genes acquiring new expression domains are linked to neofunctionalization (e.g., *dac2*, a novel leg segment in Arachnid [5]; *elnb*, bulbus arteriosus in teleost fishes [16]; and *fhl2b*, egg-spots in cichlid fishes [17]).

In some cases, gene clusters resulting from gene duplication have attracted considerable attention, such as Hox gene clusters [18], globin gene clusters [19], paraHox gene clusters [20], MHC gene clusters [21] and opsin gene clusters [22]. Duplicated genes in clusters are usually related to innovation and adaptation [11,22,23], suggesting advantageous roles during evolution. The expansion of gene clusters can be traced back to WGD and tandem duplication [23,24]. Compared to two rounds of WGD occurring before the split between cartilaginous and bony fish [25], the ancestor of teleost fishes experienced another round of WGD (teleost genome duplication, TGD) after its divergence from non-teleost actinopterygians, including Bichir, Sturgeon, bowfin and spotted gar [26,27]. This extra TGD provides an additional opportunity for gene family expansion in fishes [22,28–30].

Genome rearrangements have been suggested to occur frequently in teleost fishes [31]. If genome rearrangements can capture locally adapted genes or antagonist sex determining genes by reducing recombination, the rearranged genome can promote divergence and reproductive isolation [32,33], and thus contribute to speciation and adaptation. Examples can be found in butterflies [34], mosquitoes [35] and fish [36]. This occurs especially when advantageous genes are located next to the breakpoints of a genome rearrangement due to its low recombination rates [37,38]. Considering that the expansion of gene clusters is usually adaptive (as mentioned above) and is linked to genome instability [33,39,40], it will be interesting to investigate the roles of gene clusters located next to the breakpoints of genome rearrangements. However, related studies are sparse.

Here, we report, for the first time, the expansion of the gene *apolipoprotein D* (*ApoD*) in teleost fishes. The *ApoD* gene belongs to the lipocalin superfamily of lipid transport proteins [41,42]. In humans, ApoD is suggested to function as a multi-ligand, multifunctional transporter (e.g., hormone and pheromone) [42,43], which is important in homeostasis and in the housekeeping of many organs [43]. It is expressed in multiple tissues, most notably in the brain and testis (see e.g. [42,44,45]), and it has been suggested to be involved in the central and peripheral nervous systems [42]. Interestingly, the *ApoD* gene is duplicated in a cluster manner in two chromosomes in different teleost fishes (http://www.genomicus.biologie.ens.fr/genomicus-91.01/cgi-bin/search.pl). However, no detailed analyses have been reported. Here, we investigated the evolutionary history of *ApoD* genes in fishes for the first time.

Results

1. *In silico* screen and phylogenetic reconstruction of *ApoD* genes

To investigate the expansion of *ApoD* genes, we performed phylogenetic reconstruction with high quality assembly genome data (Figures 1A and 1B). There is one *ApoD* gene in coelacanth (*Latimeria chalumnae*), and two copies (*A* and *B*) in spotted gar (*Lepisosteus oculatus*), which are located in one cluster. Different numbers of *ApoD* genes are located in two clusters in different teleost fishes, i.e., with two copies in cavefish (*Astyanax mexicanus*) (*A1* and *A2*) and in pufferfish (*Tetraodon nigroviridis*) (*B2a* and *A2*); three copies (*A1*, *A2*, *B2*) in zebrafish (*Danio rerio*) and in cod (*Gadus morhua*) (*A1*, *A2* and *B2b*); four copies in platyfish (*Xiphophorus maculatus*) (*A1*, *A2*, *B2a* and *B2b*); five copies in Amazon molly (*Poecilia formosa*) (*A1*, *A2*, *B2a*, *B2ba1*, *B2ba2*) and in fugu (*Takifugu rubripes*) (*A1*, *A2*, *B1*, *B2a*, *B2b*); six copies in medaka (*Oryzias latipes*) (*A1*, *A2m1*, *A2m2*, *A2m3*, *B2a*, *B2b*); and eight copies in stickleback (*Gasterosteus aculeatus*) (*A1*, *A2s1*, *A2s2*, *B1*, *B2a*, *B2bs1*, *B2bs2*, *B2bs3*) and in tilapia (*Oreochromis niloticus*) (*A1*, *A2t1*, *A2t2*, *A2t3*, *A2t4*, *B1*, *B2a*, *B2b*). Noticeably, although the copy *B2b* of cod is clustered within the *B2a* clade based on a maximum likelihood (ML) tree, the bootstrap value is very low which could be due to recent duplication (Figure 1B). Considering its gene direction and syntenic with *B2b* genes in other species (Figure 1A), we named it copy *B2b*. Sequence alignment for tree construction can be found in Additional file 1.

To further retrieve the evolutionary history of *ApoD* genes in fishes, we performed an *in silico* screen across the whole phylogeny of teleost fishes using draft genomes (Figure 1C). No *ApoD* gene was detected in *Stylephorus chordatus*. Unlike copy *A1* which shows no lineage-specific duplication, copy *A2* exhibits variable lineage-specific copies in different fishes, with the highest lineage-specific duplication present in the cichlid fish tilapia (four copies). Copy *B1* is absent in the clade of Gadiformes but is kept in Acanthopterygii. Species in *Danio rerio*, *Osmerus eperlanus* and *Parasudis fraserbrunneri* possess copy *B2*. The co-existence of *B2a* and *B2b* is common in Percomorphaceae. The largest numbers of lineage-specific duplicated *ApoD* genes were

found in tilapia (copy *A2*), medaka (copy *A2*) and stickleback (copy *B2b*) in Percomorphaceae. The predicted *ApoD* gene sequences can be found in Additional file 2.

To infer the relationship between *ApoD* gene duplication and TGD, syntenic analyses were conducted, revealing that two paralogous regions in teleost fishes correspond to one ohnolog region in spotted gar and in chicken (connected by the same coloured lines in Figure 2A). For example, the region with yellow colour in linkage group (LG) 9 in spotted gar (left above), or in chromosome (chr) 2 in chicken (left down) have paired paralogous regions (connected by yellow lines) in chr17 (*ApoD* cluster I located on) and in chr20 (*ApoD* cluster II located on) in medaka. Similarly, the region with purple colour in LG3 in spotted gar and in chr1 in chicken has paired paralogous regions (connected by purple lines) in chr17 and in chr20 in medaka.

2. *ApoD* clusters next to the breakpoints of genome rearrangements

Syntenic analyses clearly show that more than one chromosome in spotted gar and in chicken are syntenic to the corresponding paralogous regions where *ApoD* clusters are located in fishes. For example, chromosomes (or LGs) containing *ApoD* clusters (chr17 and chr 20 in medaka, LG III and LG XXI in stickleback, LG9 and LG18 in tilapia, chr2 and chr24 in zebrafish, Figures 2A and 2B) are syntenic to chromosomes LG10, LG9, LG19, LG3 and LG14 in spotted gar and chr8, chr2, chr28, chr1 and chr9 in chicken (Figures 2A and 2B). Noticeably, *ApoD* clusters are located next to the breakpoints of genome rearrangements (Figures 2A and 2B). The rearranged segments (approx. 80 kb to 200 kb syntenic with LG14 in spotted gar, Figure 2B) include *ApoD* clusters and their neighbour genes (e.g., *and2*, *samd7*, *sec62*, *nadkb*, *gpr160*, *skila*, *prkci* and *phc3* for Cluster I; *otos*, *myeov2*, *and1*, *tmtopsb*, *tnk2a* and *tfr1b* for Cluster II) (Figure 2B). Analyses of available cichlid genomes further show that inversion of the segments (approx. 600 kb to 800kb)

with *ApoD* clusters as the breakpoints also occurred in the species belonging to the most species-rich lineage, the haplochromine lineage in cichlid fishes, supported by split-read analysis using Delly (Figure 2C).

3. Protein-protein association predictions and protein 3D structure comparisons

To assess the biological functions of different *ApoD* genes, protein-protein associations were predicted. Protein domain architecture analyses revealed that ApoD proteins in different fishes are composed of approx. 20 amino acids as the signal peptide and approx. 144 amino acids as the lipocalin domain (Figure 2D). The common associations of *ApoD* genes are with the immunity-related gene *pla2g15* [47,48], the high-density lipoprotein biogenesis-related gene *lcat* [49], different copies of *zg16* related to pathogenic fungi recognition [50] and *MAPK* genes [51,52]. Different *ApoD* genes in teleost fishes can be subdivided into two classes based on the association predictions. *ApoD* genes from one class (copy *A2*, copy *B2a* and copy *B1*) are associated with forkhead transcription factors. *ApoD* genes from the other class (copy *A1* and copy *B2b*) lost their associations with forkhead transcription factors but are associated with the lipoprotein-related gene *apoa1* [53]. Noticeably, the only *ApoD* gene in coelacanth possesses both associations with forkhead transcription factors and with *apoa1*. *ApoD* in coelacanth is also associated with genes encoding ligands that can activate the Notch signalling pathway (*jag1*, *jag2*) [54], and the gene belongs to the annexin family (*anxaII*) (Figure 2D). Noticeably, more members of forkhead transcription factors and MAPK genes are associated with *ApoD* genes after duplication. Unlike in other fishes, copy *A1* in zebrafish is associated with bone resorption-related duplicated genes (*ostf1a*, *ostf1b*, *ostf1c*) [55], the cell growth and division-related gene *ppp2cb* [56], the neurodevelopment-related gene *rab3gap2* [57] and the guanylate-binding gene *gbp1* [58] (Figure 2D).

Homology protein structure modelling of different *ApoD* genes shows a conserved structure, including a cup-like central part made by eight antiparallel β-sheet strands and two ends connected by loops (a wide opening side formed by loops 1, 3, 5 and 7; and a narrow, closed bottom formed by loops 2, 4 and 6) (Figure 3A). Interestingly, unlike the very conserved cup-like central part, the loops are highly variable (Figure 3A). Cluster analyses based on the whole protein 3D structure comparison can clearly segregate different duplicates, e.g., copy *A1*, copy *A2*, copy *B2a* and copy *B2b*. Especially for lineage-specific duplicated genes which are clustered together. Copy *B1s* are clustered together, nesting within the copy *A2* clade (Figure 3B). The same analyses focused on only the four loops (loops 1, 3, 5 and 7) at the opening side show a similar segregation pattern (Figure 3C). However, cluster analyses focused on only the three loops (loops 2, 4 and 6) at the bottom side did not show a segregation pattern (Figure 3D). Details about the PDB files and the cluster results can be found in Additional file 3.

4. Positive selection detection on *ApoD* genes

Positive selection was detected in many *ApoD* genes before and after duplication, for example, the branch of copy *A2* in zebrafish, the ancestral branch of copies *B2a* and *B2b* in fugu, medaka and platyfish; and, especially, in lineage-specific duplicated genes, such as in stickleback, cichlid fishes, and amazon molly (Figure 4 and Additional file 4). Positive selection sites under Bayes empirical Bayes (BEB) with posterior probability >95% were given in Figure 5 and Additional file 4. Noticeably, most sites under positive selection are located on the highly variable loops or on the connections between loops and the cup-like central part at the opening side (loops 3 and 5) (Figures 3A and 5). Our strategy here is relatively conservative. As we can see, if we only focus on one prior hypothesis on a particular branch without doing multiple test correction, we will

find more branches with dN/dS that are significantly larger than 1 ($p<0.05$) (Additional file 4). However, we chose to use our strategy to make the analyses more rigorous. Even with this strict but comprehensive strategy, our results still show that positive selection occurred in branches before and after *ApoD* gene duplication, especially for lineage-specific duplicated genes. Details about the codon alignment for the PAML analyses can be found in Additional file 4.

5. Gene expression profile detection of *ApoD* genes in different species

Copy *A1* is mainly expressed in the skin and the eye. Noticeably, copy *A1* is also highly expressed in the novel anal fin pigmentation patterns of a cichlid fish, *Astatotilapia burtoni*, which is in consistent with our previous study [59]. Copy *A2* shows redundant expression patterns in the gills, eye, skin and gonads in zebrafish and cavefish, but its orthologous *A2* genes including lineage-specific duplicated ones in the species of Acanthomorphata, which we test here (cod, stickleback, medaka, tilapia and *A. burtoni*), are expressed in the gills and in the related apparatus, the lower pharyngeal jaw. Copy *B2* in zebrafish shows redundant expression profiles (in the skin and gonads) overlapping with the profiles of copies *A1* and *A2*. Copies *B2a* and *B2b* show specific expression patterns in different fishes of Acanthomorphata, which we test here. For example, copy *B2b* is expressed in the ovary and in the liver in cod, while *B2a* and *B2bs* are mainly expressed in the spleen and in the liver of the stickleback. The expression of *B2a* and *B2b* was not detected in adult medaka, tilapia or *A. burtoni*. Interestingly, based on the available transcriptomic data for different developmental stages of gonads in tilapia, we found that *B2a* and *B2b* are highly specifically expressed at the early developmental stage of gonad tissues (five days after hatching, 5dah) (Figure 6). Copy *B1* is highly specifically expressed in the liver in tilapia and in *A. burtoni* but not in stickleback in which *B1* is inverted (Figure 1A). Noticeably, no expression profile was detected for copy *B* in gar, at least in the tissues we test here. Instead, expression profiles including

the skin, eye, gill, liver, testis and brain were detected for copy *A* in gar. Details can be found in Figure 6 and Additional file 5.

Discussion

Dynamic evolution of *ApoD* genes in teleost fishes

Here, we report the cluster expansion of *ApoD* genes specific to teleost fishes for the first time. Phylogenetic reconstruction and syntenic analyses clearly show the expansion of *ApoD* genes in two clusters with lineage-specific tandem duplications after TGD. The interplay between genome duplication and tandem duplication to prompt the expansion of the gene family has already been shown in a few studies [60–62]. It has been suggested that the fixation of duplications is much more common in genome regions where the rates of mutations are elevated due to the presence of already-fixed duplication, which is the so-called “snow-ball” effect [63]. Both genome duplication and tandem duplication can produce raw genetic materials to fuel the diversification of teleost fishes at a later stage, e.g., morphological complexity and ecological niche expansion, which is compatible with the “*time lag model*” [64]. Neofunctionalization and subfunctionalization after duplication are important steps to realize this diversification, and both can be detected in *ApoD* duplicates.

Functional divergence of *ApoD* genes occurred at the protein level. On the one hand, based on protein-protein association predictions, although common associations with MAPK genes were shared among *ApoD* genes, subdivided associations with forkhead transcription factors and *apoa1* were detected in different paralogous *ApoD* genes, indicating subfunctionalization. Noticeably, more members of forkhead transcription factors and MAPK genes are associated with *ApoD* genes

after duplication. In this case, neofunctionalization could also occur but is not limited to the dosage effects. On the other hand, divergence was also detected at the protein structure level. Protein structures are often related to functional divergence [65–68]. New fold can even evolve novel function [69,70]. Even a few residues mutations can induce structural changes [71,72]. Therefore, structure-based inference is important for understanding molecular function. Our protein 3D structure modelling showed a conserved backbone conformation in spite of sequence divergence. Indeed, this is a feature of the lipocalin family [66], to which ApoDs belong [41,42,73]. The cup-like central part is used to transport large varieties of ligands, such as hormones and pheromones [73]. Interestingly, cluster analyses based on the whole protein structure or only on the four loops at the opening side can segregate different duplicates, although not for copy *B1*. Considering that copy *B1* has been lost multiple times, its function might be not necessary and could share functions with other copies; thus, it is not surprising if they are clustered together within the copy *A2* clade. The cluster results of two other individual genes (copy *B2* of zebrafish and *B2b* of medaka) do not affect the general segregation pattern. Noticeably, unlike the four loops at the opending side, cluster analyses based on only the three loops (loops 2, 4 and 6) at the bottom side did not show a clear segregation pattern (Figure 3D). It has been suggested that the loops of lipocalin proteins can affect ligand binding specificities, which are similar to the binding modes of antibodies [74]; and mediate protein-protein interactions [66]. Their segregation with different duplicates might indicate functional divergence occurred at the loops at the opening side. The evidence that most amino acids under positive selection are located on the loops or the connections between loops and strands (Figures 3A and 5, Additonal file 4) further indicates the potential important roles of these loops during evolution. Actually, reshaping different parts of the protein structure is an efficient way to produce functional divergence at the protein level in a short time [73], and this can be one of the ways in which divergence at the protein level occurred in *ApoD* genes.

Functional divergence of *ApoD* genes also occurred at the expression level. Different *ApoD* genes in zebrafish and cavefish exhibited redundant expression profiles but became more specific as the numbers of tandem duplicates increased, indicating subfunctionalization. Noticeably, new expression profiles were detected in duplicated paralogs, e.g., copy *A1* in novel anal fin pigmentation patterns and copy *A2s* in the lower pharyngeal jaw in *A. burtoni*, which is a representative cichlid fish of the most species-richness haplochromine lineage. These two traits are key innovations that are associated with adaptive radiation in cichlid fishes [75]. With expression changes, duplication can be an important source for the emergence of novelty [76], especially if they are adaptive. The relaxed selection pressure induced by the changing expression profiles can even further prompt the accumulation of mutations at the protein level.

Potential advantageous roles of *ApoD* genes in teleost fishes

The *ApoD* gene has been thoroughly studied in humans and mice, and its abnormal expression was reported to be related to human diseases, such as Parkinson's disease and Alzheimer's disease [77–79]. Interestingly, as our study revealed here, *ApoD* dynamically expanded in fishes in two clusters in different chromosomes. Furthermore, these expanded duplicates are transcriptionally active. For example, they are highly expressed, and are restricted to different tissues related to sexual selection (skin [80], eye [81,82], gonads, anal fin pigmentation pattern [59]) and adaptation (gills [83–86], spleen [87,88], lower pharyngeal jaw [75,89]). These tissues are also derived from the neural crest, which is a key innovation in vertebrates [75,90–92]. Besides, many duplicated *ApoD* genes are under positive selection, especially for lineage-specific duplicated genes. Combined with their associations with MAPK genes and forkhead transcription factors, and their functions as pheromone and hormone transporters, we report evidences that suggests the potential importance of *ApoD* genes have in fishes.

Noticeably, two clusters exhibited distinct expression patterns. *ApoD* genes in cluster I in most fishes are expressed in tissues related to sexual selection (skin, eye, gonad and anal fin pigmentation), but in cluster II, *ApoD* genes are mainly expressed in tissues related to adaptation (gills and lower pharyngeal jaw). Interestingly, both clusters were maintained with their neighbouring genes and are located next to the breakpoints of genome rearangements during evolution. An inversion with *ApoD* clusters next to the breakpoints occurred again in cichlid fishes belonging to the haplochromine lineage, which is the most species-rich lineage [93]. If genome rearrangements can capture locally adaptive genes, or genes related to sexual antagonism, it would accelerate divergence by reducing recombination rates, thus prompting speciation and adaptation, even the emergence of neo-sex chromosome [94]. Given that many *ApoD* genes are under positive selection, their location at the breakpoints of genome rearrangements gives them more opportunities to be involved in speciation and adaptation, which deserves further investigation.

Conclusions

Here, we report the cluster expansion of *ApoD* genes in a cluster manner specific to fishes for the first time. Different evidences based on computational evolutionary analyses strongly suggest the potentially advantageous roles of *ApoD* genes in fishes. An in-depth functional characterization of *ApoD* genes could help consolidate a model for the study of subfunctionalization and neofunctionalization. Moreover, finding the regulatory mechanisms behind the cluster expansion of *ApoD* genes and the reason why the *ApoD* gene expanded specific to fishes remain open questions. As more fish genomes with high assembly quality are available, especially closely related species such as cichlid fishes, the roles of *ApoD* genes in fish speciation and adaptation could be further investigated. Above all, the *ApoD* gene clusters reported here, provide an ideal

evo-devo model for studying gene duplication, cluster maintenance, and the gene regulatory mechanism and their roles in speciation and adaptation.

Methods

In silico screen and phylogenetic reconstruction to infer gene duplication

To retrieve *ApoD* gene duplication in teleost fishes, we first extracted orthologs and paralogs in fishes with available genome data from Ensembl (Release 84) [95] and the NCBI database (https://www.ncbi.nlm.nih.gov/genome/). To confirm gene copy numbers, all orthologs and paralogs were used as queries in a tblastx search against the corresponding genomes. For all unannotated positive hits, a region spanning approx. 2 kb was extracted, and open reading frames (ORF) were predicted using Augustus (http://augustus.gobics.de/) [96]. The predicted coding sequences were then BLAST to the existing transcriptome database to retrieve the corresponding cDNA sequences. The cDNAs were then re-mapped to the corresponding predicted genes to re-check the predicted exon-intron boundary. Coding sequence of genes from human and spotted gar, as well as their neighbour genes were used to perform tblastx against the genomes of lamprey (*Lampetra fluviatilis*) and amphioxus (*Branchiostoma belcheri*). To infer gene duplication, a ML analysis was performed with *ApoD* genes retrieved from available assembled genomes in RAxML v8.2.10 [97] with the GTR+G model and '-f a' option (which generates the optimal tree and conducts 10,000 rapid bootstrap searches).

To further retrieve *ApoD*s in other fishes across the whole phylogeny with draft genomes [21], all sequences retrieved above were used as queries in a tblastx search with the threshold *e* value 0.001. The hit scaffolds were retrieved, and genes within scaffolds were predicted using

Augustus. The predicted *ApoD*s were then translated and re-aligned with known *ApoD* ORF to further confirm the exon-intron boundary. All predicted orthologs and paralogs were used as a query again in a tblastx search against the corresponding genome data until no more *ApoD* genes were predicted.

Syntenic analyses and inversion detection

To further confirm the relationship between *ApoD* gene duplication and TGD, genes regions adjacent to the duplicates and to the outgroup species that did not experience TGD (including spotted gar and chicken) were retrieved. To this end, a window of 5 Mb around the *ApoD* clusters in teleost fishes as well as the corresponding chromosomes in spotted gar and chicken were retrieved from the Ensemble database and NCBI database. Syntenic analyses were performed with SyMap [98]. To further detect the structural variation around *ApoD* genes in cichlid fishes, we retrieved the available cichlid genomes raw data from [99]. The program Delly [100] based on paired-end split-reads analyses was used to detect the inversion and its corresponding breakpoints of the segments that contain the *ApoD* clusters in cichlid fishes, with tilapia as the reference.

Protein-protein interaction prediction, protein 3D structure modelling and comparison

To predict the biological functions of *ApoD* genes, the protein domains of *ApoD* genes and protein-protein interactions were predicted with the Simple Modular Architecture Research Tool (SMART) [101,102] and the stringdb database http://string-db.org/ [103]. To see whether there are divergences at the protein level for different ApoDs, protein 3D structures were simulated with Swiss-model [104–106] using the human ApoD crystal protein structure (PDB ID 2hzr) as the template. The results were further visualized, evaluated and analysed with Swiss-PdbViewer [107].

To compare the protein structures, we first extracted and converted the corresponding information in the PDB file using Swiss-PdbViewer. Protein 3D structure comparisons were then conducted using Vorolign [108], which can compare closely related protein structures even when structurally flexible regions exist [108]. The protein 3D structure comparisons and cluster analyes were conducted with the ProCKSI-server (http://www.procksi.net/). Clustering results were further visualized using Figtree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree). Copy *B2ba2* of of amazon molly was not included in the cluster analyses due to its relatively short sequence. The PDB files were used as the input files. For the global protein structure comparisons, PDB files were extracted directly from the simulation results of Swiss-model [104–106]. To only get PDB files of the loop region, we revised the PDB files manually to get rid of the cup-like central part. To consider the highly variable connections between the loops and the cup-like central part (Figure 3A), we also included the structures of two more residues next to the loops. The resulting PDB files were further checked using Swiss-PdbViewer.

Positive selection detection

To examine whether *ApoD* duplicates underwent adaptive sequence evolution, a branch-site model was used to test positive selection affecting a few sites along the target lineages (called foreground branches) in codeml within PAML [109,110]. The rates of non-synonymous to synonymous substitutions (ω or dN/dS) with a priori partitions for foreground branches (PAML manual) were calculated. Considering the very dynamic evolutionary pattern of *ApoD* genes, we tested positive selection in every branch in each species. In this case, we designated each branch as the foreground branch to run the branch-site model multiple times. Noticeably, there are two different ways to designate the foreground branch in the branch-site model. One way is to designate only the branch as the foreground branch. The other way is to designate the whole clade (all the

branches within the clade including the ancestral branch) as the foreground branch. We included both perspectives in our data analyses (Additonal file 4). If the whole clade is under positive selection, we will label all the branches within this clade including the ancestral branch with ω>1 in the results (Figure 4).

All model comparisons in PAML were performed with fixed branch lengths (fix_blength = 2) derived under the M0 model in PAML. Alignment gaps and ambiguity characters were removed (Cleandata = 1). A likelihood ratio test was used to test for statistical significance. In addition, Bonferroni correction was conducted for multiple test correction [111]. The BEB was used to identify sites that are under positive selection. Sites under positive selection under BEB with posterior probability >0.95 are given in Figure 5 and Additional file 4.

Gene expression profile analyses

To see the expression profiles of *ApoD* duplicates in different fishes, raw transcriptomic data from zebrafish, cavefish, cod, medaka, tilapia and *A. burtoni* were retrieved from NCBI https://www.ncbi.nlm.nih.gov/ (Additional file 5). Raw reads were mapped to the corresponding cDNAs (http://www.ensembl.org/) to calculate the RPKM (reads per kilobase per million mapped reads) value. Quantitative polymerase chain reaction (qPCR) was used to detect the expression profiles of different *ApoD* duplicates in the tissues that are not included in the existing transcriptomes (Additional file 5). Prior to tissue dissection, specimens were euthanized with MS 222 (Sigma-Aldrich, USA) following an approved procedure (permit nr. 2317 issued by the cantonal veterinary office, Switzerland; Guidelines for the Care and Use of Laboratory Animals prescribed by the Regulation of Animal Experimentation of Chongqing, China). RNA isolation was performed according to the TRIzol protocol (Invitrogen, USA). DNase treatment was performed

with the DNA-free™ Kit (Ambion, Life Technologies, USA). The RNA quantity and quality were determined with a NanoDrop1000 spectrophotometer (Thermo Scientific, USA). cDNA was produced using the High Capacity RNA-to-cDNA Kit (Applied Biosystems, USA). Housekeeping gene elongation factor 1 alpha *(elfa1)* [112], ubiquitin (*ubc*) [113] and ribosomal protein L7 (*rpl7*) [17] [114] were used as endogenous control. qPCR was performed on a StepOnePlusTM Real-Time PCR system (Applied Biosystems, Life Technologies) using the SYBR Green Master Mix (Roche, Switzerland) with an annealing temperature of 58°C and following the manufacturer's protocols. Primers are available in Additional file 6.

List of abbreviations

ApoD: apolipoprotein D. WGD: whole genome duplication. TGD: teleost genome duplication. ML: maximum likelihood. LG: linkage group. chr: chromosome. BEB: Bayes empirical Bayes. ORF: open reading frame. SMART: Simple Modular Architecture Research Tool. RPKM: Reads Per Kilobase per Million mapped reads. qPCR: quantitative polymerase chain reaction. *elfa*1: elongation factor 1 alpha. *ubc:* ubiquitin. *rpl7:* ribosomal protein L7.

Declarations

Ethics approval and consent to participate

Animal experiments reported in this study have been approved by the cantonal veterinary office in Basel under permit number 2317, Switzerland; Guidelines for Care and Use of Laboratory Animals prescribed by the Regulation of Animal Experimentation of Chongqing, China.

Consent for publication

Not applicable

Availability of data and material

The datasets supporting the results of this article are available in the Dryad repository, doi:10.5061/dryad.39g63v2. https://datadryad.org/review?doi=doi:10.5061/dryad.39g63v2 [46].

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported by the PhD grant from University of Basel, Switzerland; postdoctoral funding from Southwest University, China; China Postdoctoral Science Foundation Grant (2018M633311); and National Natural Science Foundation of China (31802283) to LG.

Author's contributions

LG discovered the expansion of *ApoD* gene family, did data analysis and wrote the manuscript. LG did protein 3D structure comparison and cluster analyses with the help of CX.

Acknowledgements

We thank Prof. Walter Salzburger for his suggestions. We also thank Prof. Deshou Wang for his help of sampling of tilapia and laboratory support. We particularly thank Prof. Ziheng Yang for his valuable suggestions about positive selection detection using PAML. Many thanks to the kind computing support from Prof. Zhengwang Zhang. We appreciate the valuable suggestions and comments from the associate editor and two anonymous reviewers very much. Many thanks to Dario Moser, Heinz-Georg Belting, Jing Wei and Hua Ruan for their help of fish sampling. Many thanks for the help of fish dissection from Yang Zhao, Fabrizia Ronco, Attila Rüegg, Adrian Indermaur, Xianbo Zhang and He Ma. Many thanks to Lukas Zimmerman, Peter Fields, Yuchen Sun, Yanyan Xu, Zihui Zhang and De Chen for their help and fruitful discussions. We also would like to express our thanks to Prof. Xiangjiang Zhan in the Institute of Zoology, Chinese Academy of Sciences for the help of revising the rebuttal letter. Many thanks to Prof. Anders Moller in CNRS, Université Paris-Sud in France, Dr. Juan Felipe Ortiz in Vanderbilt University in America for the correction of English language.

Figure legends

Figure 1 (A) Cluster expansion of *ApoD* genes specific to teleost fishes after teleost-specific duplication (TGD). Each arrow represents a single gene copy. Genes with red and orange colours represent paralogs derived from one common ancestor. Genes with dark and light green colours represent paralogs derived from one common ancestor. Phylogeny reconstruction is based on a consensus fish phylogeny [22]. (B) Maximum likelihood phylogenetic tree reconstruction to infer gene duplication. Bootstrap values > 50% are marked on the branch. Lineage-specific duplication events were labelled. Note, although copy *B2b* of *Gadus morhua* is clustered within the *B2a* clade (labelled with a green star), its bootstrap value is low, which could be due to recent duplication. (C) A dynamic evolutionary pattern of *ApoD* genes across the entire phylogeny of teleost fishes. Highly

variable copy numbers are detected in different lineages, especially in the Paracanthopterygii lineage. *Stylephorus chordatus* lost all *ApoD* genes. Compared to copy *A1*, copy *A2* exhibits more variable lineage-specific duplicates in different fishes, with the highest numbers appearing in tilapia (four copies). Copy *B1* is absent in the whole clade of Gadiformes. Copy *B2* only shows up in species *Danio rerio*, *Osmerus eperlanus* and *Parasudis fraserbrunneri*. The co-existence of *B2a* and *B2b* is common in Percomorphaceae. The largest numbers of lineage-specific duplicated genes are found in tilapia (copy *A2*), medaka (copy *A2*) and stickleback (copy *B2b*) in Percomorphaceae.

Figure 2 (A) Syntenic analyses between teleost fishes and spotted gar (top) and chicken (bottom), respectively. The same colour between spotted gar and chicken represents orthologous chromosomes. Two paralogous duplicated segments in teleost fishes (chromosomes (chrs)/linkage groups (LG) indicated in red colour) can be traced back to one corresponding orthologous region in spotted gar and chicken (chrs/LGs named with black colour), linked by colourful lines. Arrows show the regions where *apolipoprotein D* (*ApoD*) clusters are located. (B) *ApoD* clusters as the breakpoints of genome rearrangements before teleost-specific genome duplication (TGD). The same colour between gar and chicken represents orthologous chromosomes. Neighbouring genes are labelled. The red arrows show breakpoints, and the black arrows show gene direction. Neighbouring genes are named according to Ensembl database. (C) Inversion with *ApoD* clusters as the breakpoints occurred again in cichlid fishes. The haplochromine lineage, the most species-rich lineage of cichlid fishes, is labelled. (D) ApoD domains and protein-protein association predictions. a. Conserved domains include a single peptide of approx. 20 amino acids (AA) and a lipocalin domain of approx. 144 AA. b. Different paralogs exhibited differential associations. The common association is with MAPK genes. One class (copies *A2*, *B2a* and *B1*) is associated with multiple forkhead transcription factors. The other class (copies *A1* and *B2b*) lost this association; instead, it

is associated with the lipoprotein-related gene *apoa1*. The single *ApoD* gene in coelacanth possesses both associations.

Figure 3 (A) Protein 3D structure modelling. Different ApoDs among species exhibited a conserved 3D structure, including a cup-like central part made by eight antiparallel β-sheet strands and two ends connected by loops (a wide opening part formed by loops 1, 3, 5 and 7; and a narrow, closed bottom formed by loops 2, 4 and 6). Unlike the very conserved cup-like central part, the loops at the two ends are highly variable. Note that most sites under positive selection are located on the loops or on the connections between the loops and the cup-like central part. (B) Cluster analyses based on the whole protein 3D structure. Cluster analyses can clearly segregate different *ApoD* duplicates, including lineage-specific duplicated genes. (C) Cluster analyses based on only four loops (loops 1, 3, 5 and 7) at the opening side. Cluster analyses can segregate different *ApoD* duplicates. (D) Cluster analyses based on only three loops (loops 2, 4 and 6) at the bottom side. Cluster analyses cannot segregate with different duplicates.

Figure 4 Positive selection detection of *ApoD* genes in different fishes under the branch-site model. Many duplicated *ApoD* genes are under positive selection with the value of ω significantly larger than 1 after Bonferroni correction, especially for lineage-specific duplicated genes. Note that if the whole clade is designated as the foreground branch when detecting positive selection, each branch of this clade will be labelled if its ω is significantly larger than 1. The unrooted tree was used for PAML analyses, and the rooted tree is only for presentation. More details can be found in Additional file 4.

Figure 5 Amino acids alignment of *ApoD* duplicates among species with annotation. The purple arrow represents the β-strand, and the yellow arrow represents the loop. Sites under positive selection under Bayes empirical Bayes (BEB) with posterior probability > 95% are labelled with the red colour. Most sites under positive selection are distributed on the loops or on the connections between the loop and the β-strand. More details can be found in Additional file 4. Note, the positions of the sites that under positive selection in Additional file 4 are based on the codon alignment used for PAML analyses (Additional file 4) instead of the amino acids alignment here. But the amino acids that under positive selection are the same in Figure 5 and Additional file 4.

Figure 6 Gene expression profiles of different *ApoD* genes. (a) Schematic figure showing the gene expression profiles of different *ApoD* genes in two clusters after teleost-specific duplication (TGD). Each block represents a single gene copy. Phylogeny reconstruction was based on the consensus fish phylogeny [22]. (b) Detailed gene expression profiles of *ApoD* genes among fishes. The red colour represents tissues with a high expression level. Dark grey, data unavailable (either because the tissue does not exist in the species, or is undetected in this study). Light grey represents tissues with a low expression level. LPJ, the lower pharyngeal jaw in cichlid fishes. (c) Gene expression profiles of *ApoD* genes in different developmental stages of gonads in tilapia based on available transcriptomic data [115]. RPKM, reads per kilobase per million mapped reads; dah, days after hatching.

Additional Files

Additional file 1 Sequence alignment for ML tree reconstruction.

Additional file 2 Predicted ApoD genes across the phylogeny.

Additional file 3 Protein PDB files and cluster analyses results.

Additional file 4 Sequence alignment for PAML test and statistics results.

Additional file 5 Raw transcriptomic data information and gene expression profiles.

Additional file 6 qPCR primer sequences.

References

1. Ohno S. Evolution by gene duplication. Springer Berlin Heidelb. 1970;

2. Innan H, Kondrashov F. The evolution of gene duplications: classifying and distinguishing between models. Nat. Rev. Genet. 2010;11:97–108.

3. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. Genetics. 1999;151:1531–45.

4. Rastogi S, Liberles DA. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. BMC Evol. Biol. 2005;5:28.

5. Turetzek N, Pechmann M, Schomburg C, Schneider J, Prpic N-M. Neofunctionalization of a Duplicate dachshund Gene Underlies the Evolution of a Novel Leg Segment in Arachnids. Mol. Biol. Evol. 2016;33:109–21.

6. Chen L, DeVries AL, Cheng C-HC. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish. Proc. Natl. Acad. Sci. 1997;94:3811–6.

7. Dulai KS, von Dornum M, Mollon JD, Hunt DM. The Evolution of Trichromatic Color Vision by Opsin Gene Duplication in New World and Old World Primates. Genome Res. 1999;9:629–38.

8. Tang Y-C, Amon A. Gene copy-number alterations: a cost-benefit analysis. Cell. 2013;152:394–405.

9. Lin Z, Li W-H. Expansion of hexose transporter genes was associated with the evolution of aerobic fermentation in yeasts. Mol. Biol. Evol. 2011;28:131–42.

10. Glasauer SMK, Neuhauss SCF. Whole-genome duplication in teleost fishes and its evolutionary consequences. Mol. Genet. Genomics. 2014;289:1045–60.

11. Baalsrud HT, Voje KL, Tørresen OK, Solbakken MH, Matschiner M, Malmstrøm M, et al. Evolution of Hemoglobin Genes in Codfishes Influenced by Ocean Depth. Sci. Rep. 2017;7:7956.

12. Zhang J, Zhang Y, Rosenberg HF. Adaptive evolution of a duplicated pancreatic ribonuclease gene in a leaf-eating monkey. Nat. Genet. 2002;30:411–5.

13. Storz JF, Opazo JC, Hoffmann FG. Gene duplication, genome duplication, and the functional diversification of vertebrate globins. Mol. Phylogenet. Evol. 2013;66:469–78.

14. Zhang J. Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. Nat. Genet. 2006;38:819–23.

15. Jiménez-Delgado S, Pascual-Anaya J, Garcia-Fernàndez J. Implications of duplicated cis-regulatory elements in the evolution of metazoans: the DDI model or how simplicity begets novelty. Brief. Funct. Genomic. Proteomic. 2009;8:266–75.

16. Moriyama Y, Ito F, Takeda H, Yano T, Okabe M, Kuraku S, et al. Evolution of the fish heart by sub/neofunctionalization of an elastin gene. Nat. Commun. 2016;7:10397.

17. Santos ME, Braasch I, Boileau N, Meyer BS, Sauteur L, Böhne A, et al. The evolution of cichlid fish egg-spots is linked with a cis-regulatory change. Nat. Commun. 2014;5:5149.

18. Carrasco AE, McGinnis W, Gehring WJ, De Robertis EM. Cloning of an X. laevis gene expressed during early embryogenesis coding for a peptide region homologous to Drosophila homeotic genes. Cell. 1984;37:409–14.

19. Proudfoot NJ, Shander MH, Manley JL, Gefter ML, Maniatis T. Structure and in vitro transcription of human globin genes. Science. 1980;209:1329–36.

20. Brooke NM, Garcia-Fernàndez J, Holland PWH. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. Nature. 1998;392:920–2.

21. Malmstrøm M, Matschiner M, Tørresen OK, Star B, Snipen LG, Hansen TF, et al. Evolution of the immune system influences speciation rates in teleost fishes. Nat. Genet. 2016;48:1204–10.

22. Cortesi F, Musilová Z, Stieb SM, Hart NS, Siebeck UE, Malmstrøm M, et al. Ancestral

duplications and highly dynamic opsin gene evolution in percomorph fishes. Proc. Natl. Acad. Sci. 2015;112:1493–8.

23. Garcia-Fernàndez J. The genesis and evolution of homeobox gene clusters. Nat. Rev. Genet. 2005;6:881–92.

24. Hardison RC. Evolution of hemoglobin and its genes. Cold Spring Harb. Perspect. Med. 2012;2:a011627.

25. Robinson-Rechavi M, Boussau B, Laudet V. Phylogenetic dating and characterization of gene duplications in vertebrates: the cartilaginous fish reference. Mol. Biol. Evol. 2004;21:580–6.

26. Sato Y, Nishida M. Teleost fish with specific genome duplication as unique models of vertebrate evolution. Environ. Biol. Fishes.2010;88:169–88.

27. Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manousaki T, Pasquier J, et al. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. Nat. Genet. 2016;48:427–37.

28. Brunet FG, Volff J-N, Schartl M. Whole Genome Duplications Shaped the Receptor Tyrosine Kinase Repertoire of Jawed Vertebrates. Genome Biol. Evol. 2016;8:1600–13.

29. Gillis WQ, St John J, Bowerman B, Schneider SQ. Whole genome duplications and expansion of the vertebrate GATA transcription factor gene family. BMC Evol. Biol. 2009;9:207.

30. Voldoire E, Brunet F, Naville M, Volff J-N, Galiana D. Expansion by whole genome duplication and evolution of the sox gene family in teleost fish. Vaudry H, editor. PLoS One. 2017;12:e0180936.

31. Sémon M, Wolfe KH. Rearrangement rate following the whole-genome duplication in teleosts. Mol. Biol. Evol. 2007;24:860–7.

32. Kirkpatrick M, Barton N. Chromosome inversions, local adaptation and speciation. Genetics. 2006;173:419–34.

33. Farré M, Micheletti D, Ruiz-Herrera A. Recombination Rates and Genomic Shuffling in Human

and Chimpanzee—A New Twist in the Chromosomal Speciation Theory. Mol. Biol. Evol. 2013;30:853–64.

34. Joron M, Frezal L, Jones RT, Chamberlain NL, Lee SF, Haag CR, et al. Chromosomal rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature. 2011;477:203–6.

35. Ayala D, Fontaine MC, Cohuet A, Fontenille D, Vitalis R, Simard F. Chromosomal inversions, natural selection and adaptation in the malaria vector Anopheles funestus. Mol. Biol. Evol. 2011;28:745–58.

36. Barth JMI, Berg PR, Jonsson PR, Bonanomi S, Corell H, Hemmer-Hansen J, et al. Genome architecture enables local adaptation of Atlantic cod despite high connectivity. Mol. Ecol. 2017;26:4452–66.

37. Stevison LS, Hoehn KB, Noor MAF. Effects of Inversions on Within- and Between-Species Recombination and Divergence. Genome Biol. Evol. 2011;3:830–41.

38. Corbett-Detig RB. Selection on Inversion Breakpoints Favors Proximity to Pairing Sensitive Sites in *Drosophila melanogaster*. Genetics. 2016;204:259–65.

39. Otto SP. The Evolutionary Consequences of Polyploidy. Cell. 2007;131:452–62.

40. Hufton AL, Groth D, Vingron M, Lehrach H, Poustka AJ, Panopoulou G. Early vertebrate whole genome duplications were predated by a period of intense genome rearrangement. Genome Res. 2008;18:1582–91.

41. Ayrault Jarrier M, Levy G, Polonovski J. ETUDE DES ALPHA-LIPOPROT'EINES S'ERIQUES HUMAINES PAR. Bull. Soc. Chim. Biol. (Paris). 1963;45:703–13.

42. Rassart E, Bedirian A, Do Carmo S, Guinard O, Sirois J, Terrisse L, et al. Apolipoprotein D. Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol. 2000;1482:185–98.

43. Weech P, Provost P, Tremblay N, Camato R, Milne R, Marcel Y, et al. Apolipoprotein D—An atypical apolipoprotein. Prog. Lipid Res. 1991;30:259–66.

44. Drayna D, Fielding C, McLean J, Baer B, Castro G, Chen E, et al. Cloning and expression of human apolipoprotein D cDNA. J. Biol. Chem. 1986;261:16535–9.

45. Provost PR, Villeneuve L, Weech PK, Milne RW, Marcel YL, Rassart E. Localization of the major sites of rabbit apolipoprotein D gene transcription by in situ hybridization. J. Lipid Res. 1991;32:1959–70.

46. Gu L, Xia C. Data from: Cluster expansion of apolipoprotein D (ApoD) genes in teleost fishes. Dryad Digit. Repos. 2018. p. https://doi.org/10.5061/dryad.39g63v2.

47. Gilleron M, Lepore M, Layre E, Cala-De Paepe D, Mebarek N, Shayman JA, et al. Lysosomal Lipases PLRP2 and LPLA2 Process Mycobacterial Multi-acylated Lipids and Generate T Cell Stimulatory Antigens. Cell Chem. Biol. 2016;23:1147–56.

48. Bailey SD, Xie C, Do R, Montpetit A, Diaz R, Mohan V, et al. Variation at the NFATC2 Locus Increases the Risk of Thiazolidinedione-Induced Edema in the Diabetes REduction Assessment with ramipril and rosiglitazone Medication (DREAM) Study. Diabetes Care. 2010;33:2250–3.

49. Fotakis P, Kuivenhoven JA, Dafnis E, Kardassis D, Zannis VI. The Effect of Natural LCAT Mutations on the Biogenesis of HDL. Biochemistry. 2015;54:3348–59.

50. Tateno H, Yabe R, Sato T, Shibazaki A, Shikanai T, Gonoi T, et al. Human ZG16p recognizes pathogenic fungi through non-self polyvalent mannose in the digestive system. Glycobiology. 2012;22:210–20.

51. Plestant C, Anton ES. Scaling the MAPK Signaling Threshold during CNS Patterning. Dev. Cell. 2013;25:221–2.

52. Shvartsman SY, Coppey M, Berezhkovskii AM. MAPK signaling in equations and embryos. Fly. 2009;3:62–7.

53. Mangaraj M, Nanda R, Panda S. Apolipoprotein A-I: A Molecule of Diverse Function. Indian J. Clin. Biochem. 2016;31:253–9.

54. Fiaschetti G, Schroeder C, Castelletti D, Arcaro A, Westermann F, Baumgartner M, et al.

NOTCH ligands JAG1 and JAG2 as critical pro-survival factors in childhood medulloblastoma. Acta Neuropathol. Commun. 2014;2:39.

55. Reddy S, Devlin R, Menaa C, Nishimura R, Choi SJ, Dallas M, et al. Isolation and characterization of a cDNA clone encoding a novel peptide (OSF) that enhances osteoclast formation and bone resorption. J. Cell. Physiol. 1998;177:636–45.

56. Wong C-H, Fung Y-WW, Ng EK-O, Lee SM-Y, Waye MM-Y, Tsui SK-W. LIM domain protein FHL1B interacts with PP2A catalytic β subunit - A novel cell cycle regulatory pathway. FEBS Lett. 2010;584:4511–6.

57. Ng EL, Tang BL. Rab GTPases and their roles in brain neurons and glia. Brain Res. Rev. 2008;58:236–46.

58. Pandita E, Rajan S, Rahman S, Mullick R, Das S, Sau AK. Tetrameric assembly of hGBP1 is crucial for both stimulated GMP formation and antiviral activity. Biochem. J. 2016;473:1745–57.

59. Gu L, Xia C. Revelation of the Genetic Basis for Convergent Innovative Anal Fin Pigmentation Patterns in Cichlid Fishes. bioRxiv. 2017; doi:https://doi.org/10.1101/165217.

60. Hofberger JA, Nsibo DL, Govers F, Bouwmeester K, Schranz ME. A Complex Interplay of Tandem- and Whole-Genome Duplication Drives Expansion of the L-Type Lectin Receptor Kinase Gene Family in the Brassicaceae. Genome Biol. Evol. 2015;7:720–34.

61. Bellieny-Rabelo D, Oliveira AEA, Venancio TM. Impact of Whole-Genome and Tandem Duplications in the Expansion and Functional Diversification of the F-Box Family in Legumes (Fabaceae). Nikolaidis N, editor. PLoS One. 2013;8:e55127.

62. Hammoudi V, Vlachakis G, Schranz ME, van den Burg HA. Whole-genome duplications followed by tandem duplications drive diversification of the protein modifier SUMO in Angiosperms. New Phytol. 2016;211:172–85.

63. Kondrashov FA, Kondrashov AS. Role of selection in fixation of gene duplications. J. Theor. Biol. 2006;239:141–51.

64. Schranz ME, Mohammadin S, Edger PP. Ancient whole genome duplications, novelty and diversification: the WGD Radiation Lag-Time Model. Curr. Opin. Plant Biol. 2012;15:147–53.

65. Engel A, Gaub HE. Structure and mechanics of membrane proteins. Annu. Rev. Biochem. 2008;77:127–48.

66. Flower DR. The lipocalin protein family: structure and function. Biochem. J. 1996;318:1–14.

67. McLachlan AD. Protein Structure and Function. Annu. Rev. Phys. Chem. 1972;23:165–92.

68. Klingenberg M. Membrane protein oligomeric structure and transport function. Nature. 1981;290:449–54.

69. Friedman JM. Structure, dynamics, and reactivity in hemoglobin. Science. 1985;228:1273–80.

70. Ewart K V., Lin Q, Hew CL. Structure, function and evolution of antifreeze proteins. Cell. Mol. Life Sci. C. 1999;55:271–83.

71. Arévalo-Pinzón G, Curtidor H, Muñoz M, Patarroyo MA, Bermudez A, Patarroyo ME. A single amino acid change in the Plasmodium falciparum RH5 (PfRH5) human RBC binding sequence modifies its structure and determines species-specific binding activity. Vaccine. 2012;30:637–46.

72. Schaefer C, Rost B. Predict impact of single amino acid change upon protein structure. BMC Genomics. 2012;13:S4.

73. Flower DR. Multiple molecular recognition properties of the lipocalin protein family. J. Mol. Recognit. 1995;8:185–95.

74. Skerra A. Engineered protein scaffolds for molecular recognition. J. Mol. Recognit. 2000;13:167–87.

75. Salzburger W. The interaction of sexually and naturally selected traits in the adaptive radiations of cichlid fishes. Mol. Ecol. 2009;18:169–85.

76. Rogers RL, Shao L, Thornton KR. Tandem duplications lead to novel expression patterns through exon shuffling in Drosophila yakuba. Begun DJ, editor. PLOS Genet. 2017;13:e1006795.

77. Chen Y, Jia L, Wei C, Wang F, Lv H, Jia J. Association between polymorphisms in the

apolipoprotein D gene and sporadic Alzheimer's disease. Brain Res. 2008;1233:196–202.

78. Helisalmi S, Hiltunen M, Vepsäläinen S, Iivonen S, Corder EH, Lehtovirta M, et al. Genetic variation in apolipoprotein D and Alzheimer's disease. J. Neurol. 2004;251:951–7.

79. Waldner A, Dassati S, Redl B, Smania N, Gandolfi M. Apolipoprotein D Concentration in Human Plasma during Aging and in Parkinson's Disease: A Cross-Sectional Study. Parkinsons. Dis. 2018;2018:1–7.

80. Wellenreuther M, Svensson EI, Hansson B. Sexual selection and genetic colour polymorphisms in animals. Mol. Ecol. 2014;23:5398–414.

81. CARLETON KL, PARRY JWL, BOWMAKER JK, HUNT DM, SEEHAUSEN O. Colour vision and speciation in Lake Victoria cichlids of the genus Pundamilia. Mol. Ecol. 2005;14:4341–53.

82. Flamarique IN, Bergstrom C, Cheng CL, Reimchen TE. Role of the iridescent eye in stickleback female mate choice. J. Exp. Biol. 2013;216:2806–12.

83. Bystriansky JS, Schulte PM. Changes in gill H+-ATPase and Na+/K+-ATPase expression and activity during freshwater acclimation of Atlantic salmon (Salmo salar). J. Exp. Biol. 2011;214:2435–42.

84. McCormick SD, Bradshaw D. Hormonal control of salt and water balance in vertebrates. Gen. Comp. Endocrinol. 2006;147:3–8.

85. Sakamoto T. Growth hormone and prolactin in environmental adaptation. Zoolog. Sci. 2003;20:1497–8.

86. Foskett JK, Bern HA, Machen TE, Conner M. Chloride cells and the hormonal control of teleost fish osmoregulation. J. Exp. Biol. 1983;106:255–81.

87. Papetti C, Harms L, Windisch HS, Frickenhaus S, Sandersfeld T, Jürgens J, et al. A first insight into the spleen transcriptome of the notothenioid fish Lepidonotothen nudifrons: Resource

description and functional overview. Mar. Genomics. 2015;24:237–9.

88. Huang Y, Chain FJJ, Panchal M, Eizaguirre C, Kalbe M, Lenz TL, et al. Transcriptome profiling of immune tissues reveals habitat-specific gene expression between lake and river sticklebacks. Mol. Ecol. 2016;25:943–58.

89. Muschick M, Indermaur A, Salzburger W. Convergent evolution within an adaptive radiation of cichlid fishes. Curr. Biol. 2012;22:2362–8.

90. Green SA, Simoes-Costa M, Bronner ME. Evolution of vertebrates as viewed from the crest. Nature. 2015;520:474–82.

91. Barlow-Anacker AJ, Fu M, Erickson CS, Bertocchini F, Gosain A. Neural Crest Cells Contribute an Astrocyte-like Glial Population to the Spleen. Sci. Rep. 2017;7:45645.

92. Bailey AP, Bhattacharyya S, Bronner-Fraser M, Streit A. Lens Specification Is the Ground State of All Sensory Placodes, from which FGF Promotes Olfactory Identity. Dev. Cell. 2006;11:505–17.

93. Salzburger W, Mack T, Verheyen E, Meyer A. Out of Tanganyika: genesis, explosive speciation, key-innovations and phylogeography of the haplochromine cichlid fishes. BMC Evol. Biol. 2005;5:17.

94. Charlesworth D. Evolution of recombination rates between sex chromosomes. Philos. Trans. R. Soc. B Biol. Sci. 2017;372:20160456.

95. Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2014. Nucleic Acids Res. 2014;42:D749–55.

96. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics. 2008;24:637–44.

97. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014;30:1312–3.

98. Soderlund C, Bomhoff M, Nelson W. SyMAP v3.4: a turnkey synteny system with application to plant genomes. Nucleic Acids Res. 2011;39:e68.

99. Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, et al. The genomic substrate for adaptive radiation in African cichlid fish. Nature. 2014;513:375–81.

100. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012;28:i333–9.

101. Letunic I, Doerks T, Bork P. SMART: recent updates, new developments and status in 2015. Nucleic Acids Res. 2015;43:D257–60.

102. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. Nucleic Acids Res. 2018;46:D493–6.

103. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Res. 2017;45:D362–8.

104. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 2014;42:W252–8.

105. Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T. Protein structure homology modeling using SWISS-MODEL workspace. Nat. Protoc. 2008;4:1–13.

106. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. Bioinformatics. 2006;22:195–201.

107. Guex N, Peitsch MC, Schwede T. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: A historical perspective. Electrophoresis. 2009;30:S162–73.

108. Birzele F, Gewehr JE, Csaba G, Zimmer R. Vorolign--fast structural alignment using Voronoi contacts. Bioinformatics. 2007;23:e205–11.

109. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 2007;24:1586–91.

110. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput. Appl. Biosci. 1997;13:555–6.

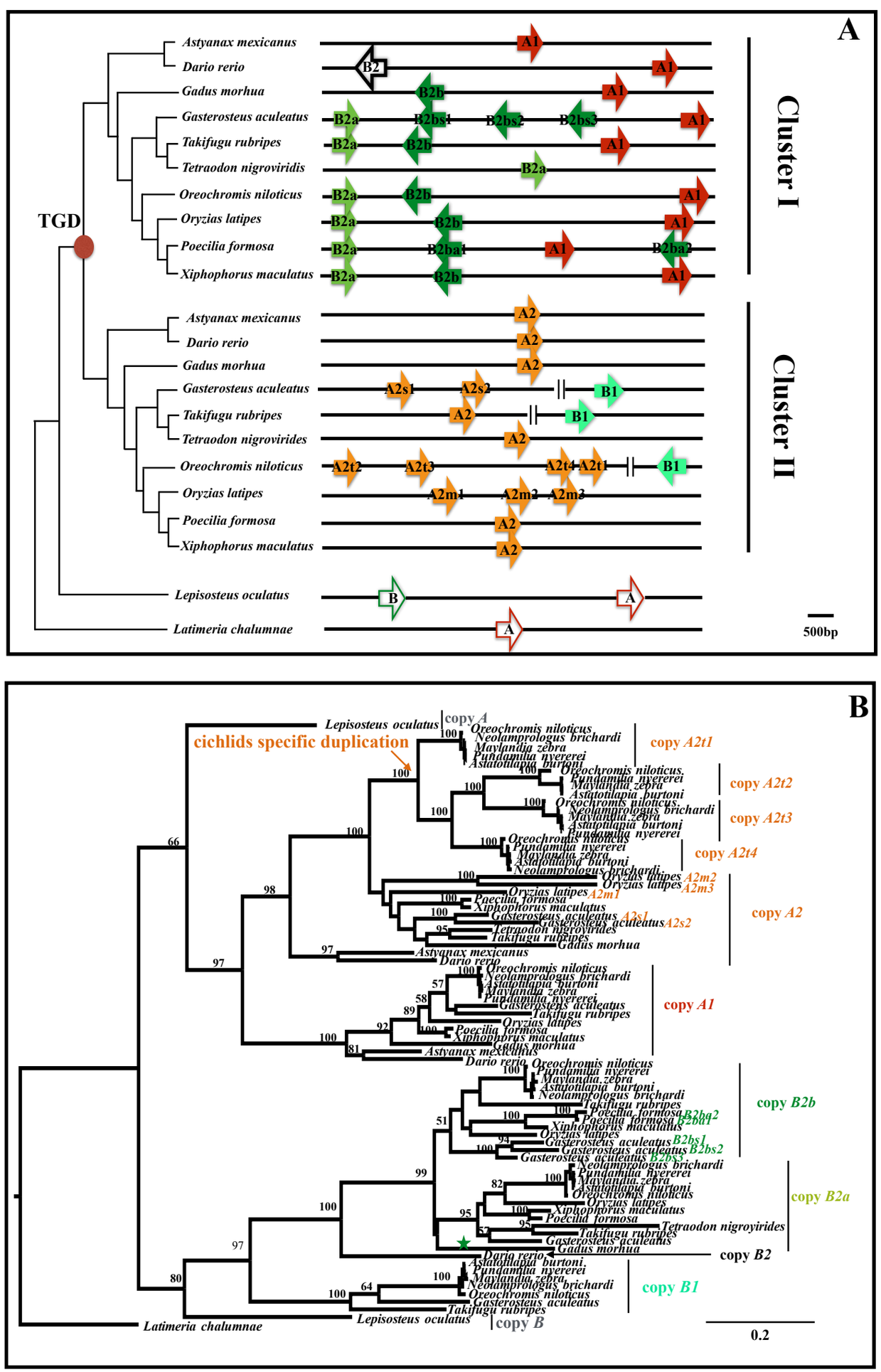
111. Anisimova M, Yang Z. Multiple hypothesis testing to detect lineages under positive selection that affects only a few sites. Mol. Biol. Evol. 2007;24:1219–28.

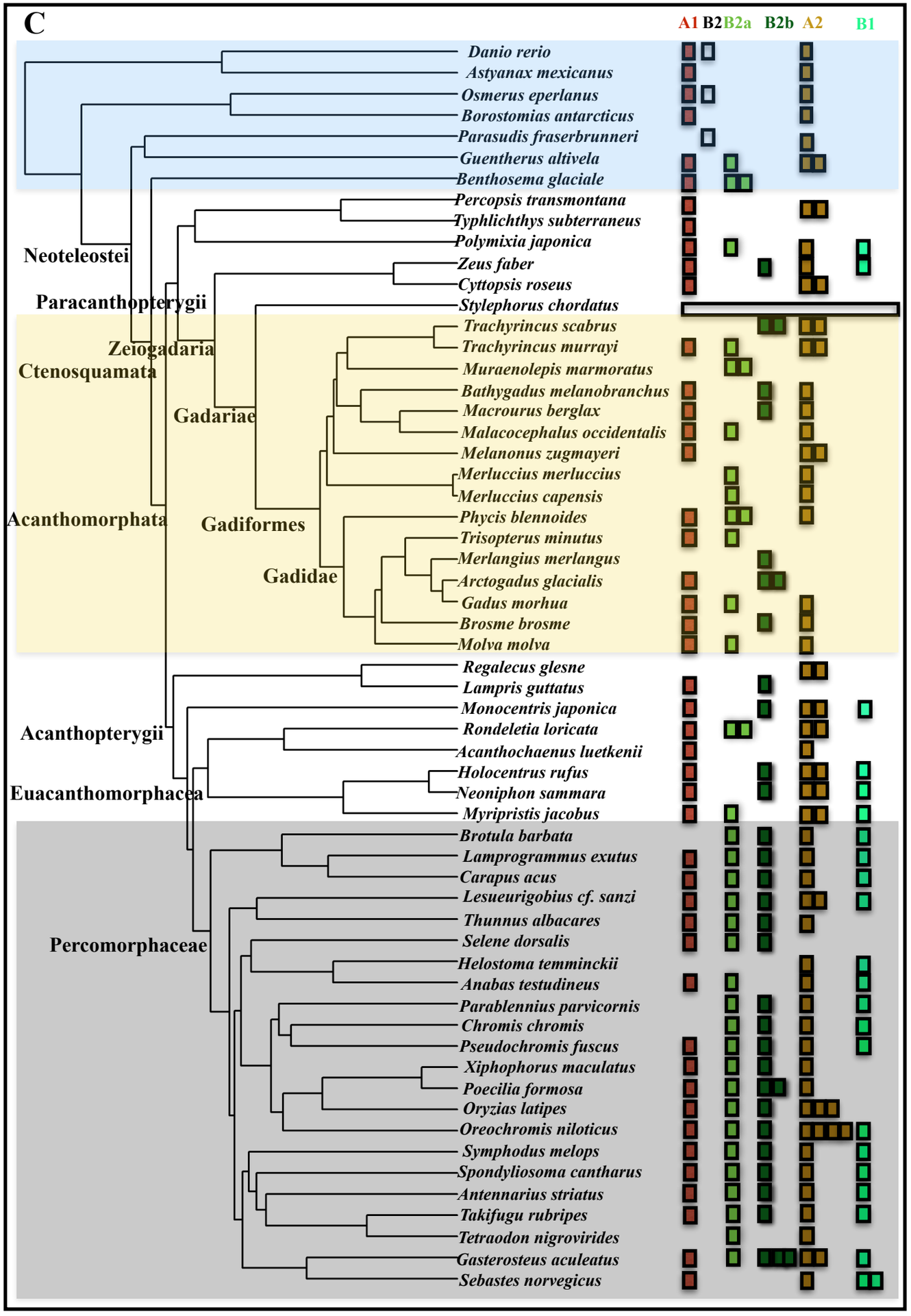
112. McCurley AT, Callard G V. Characterization of housekeeping genes in zebrafish: male-female differences and effects of tissue type, developmental stage and chemical treatment. BMC Mol. Biol. 2008;9:102.

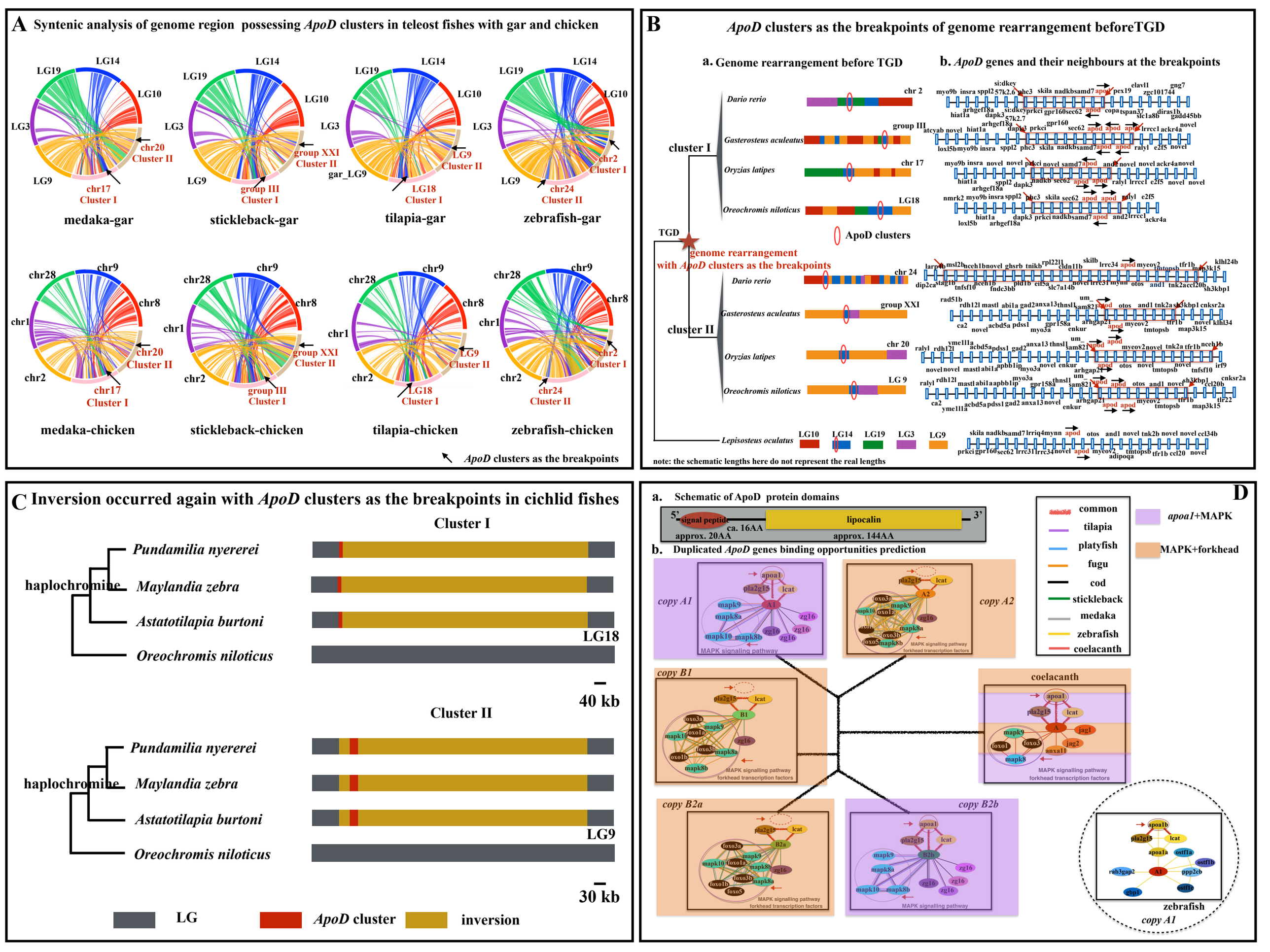
113. Hibbeler S, Scharsack JP, Becker S. Housekeeping genes for quantitative expression studies in the three-spined stickleback Gasterosteus aculeatus. BMC Mol. Biol. 2008;9:18.

114. Zhang Z, Hu J. Development and validation of endogenous reference genes for expression profiling of medaka (Oryzias latipes) exposed to endocrine disrupting chemicals by quantitative real-time RT-PCR. Toxicol. Sci. 2007;95:356–68.

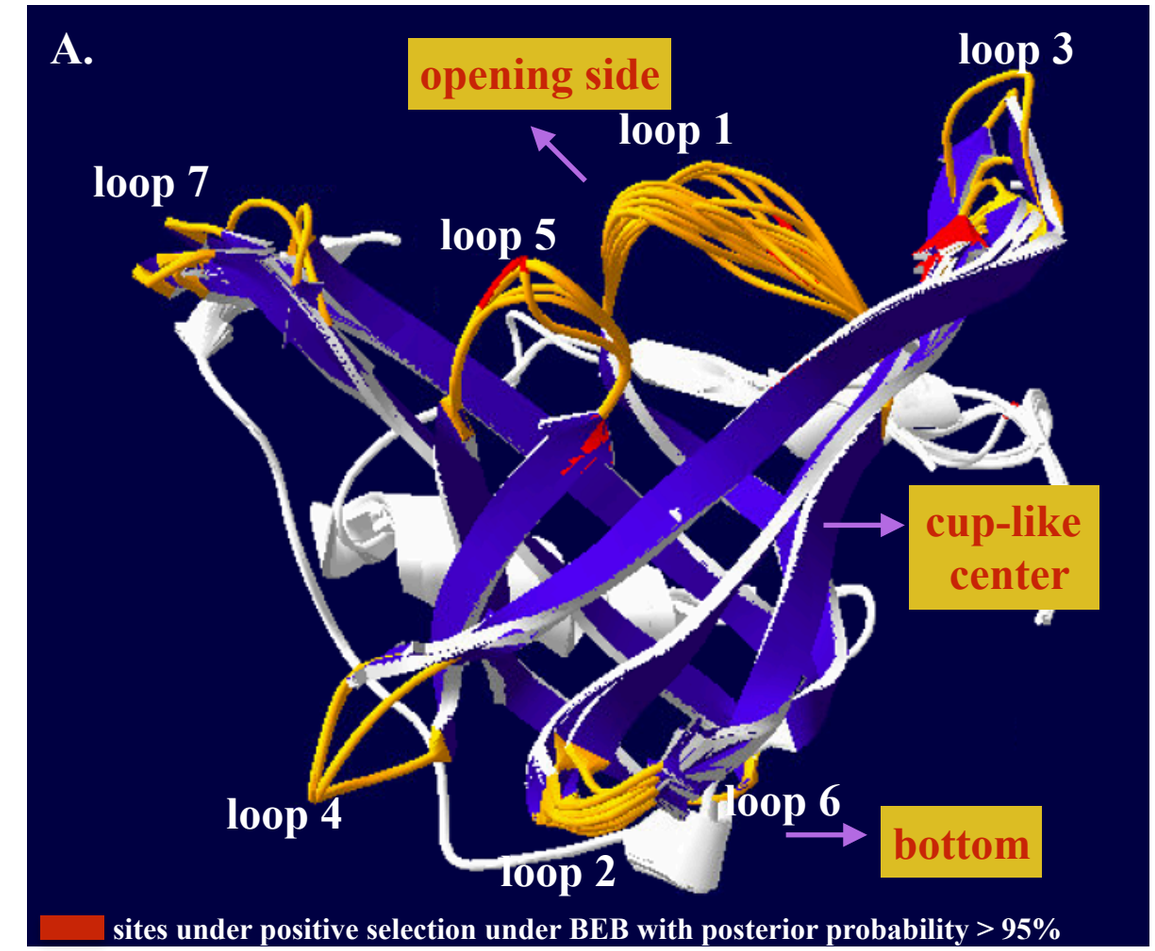
115. Tao W, Sun L, Shi H, Cheng Y, Jiang D, Fu B, et al. Integrated analysis of miRNA and mRNA expression profiles in tilapia gonads at an early stage of sex differentiation. BMC Genomics. 2016;17:328.

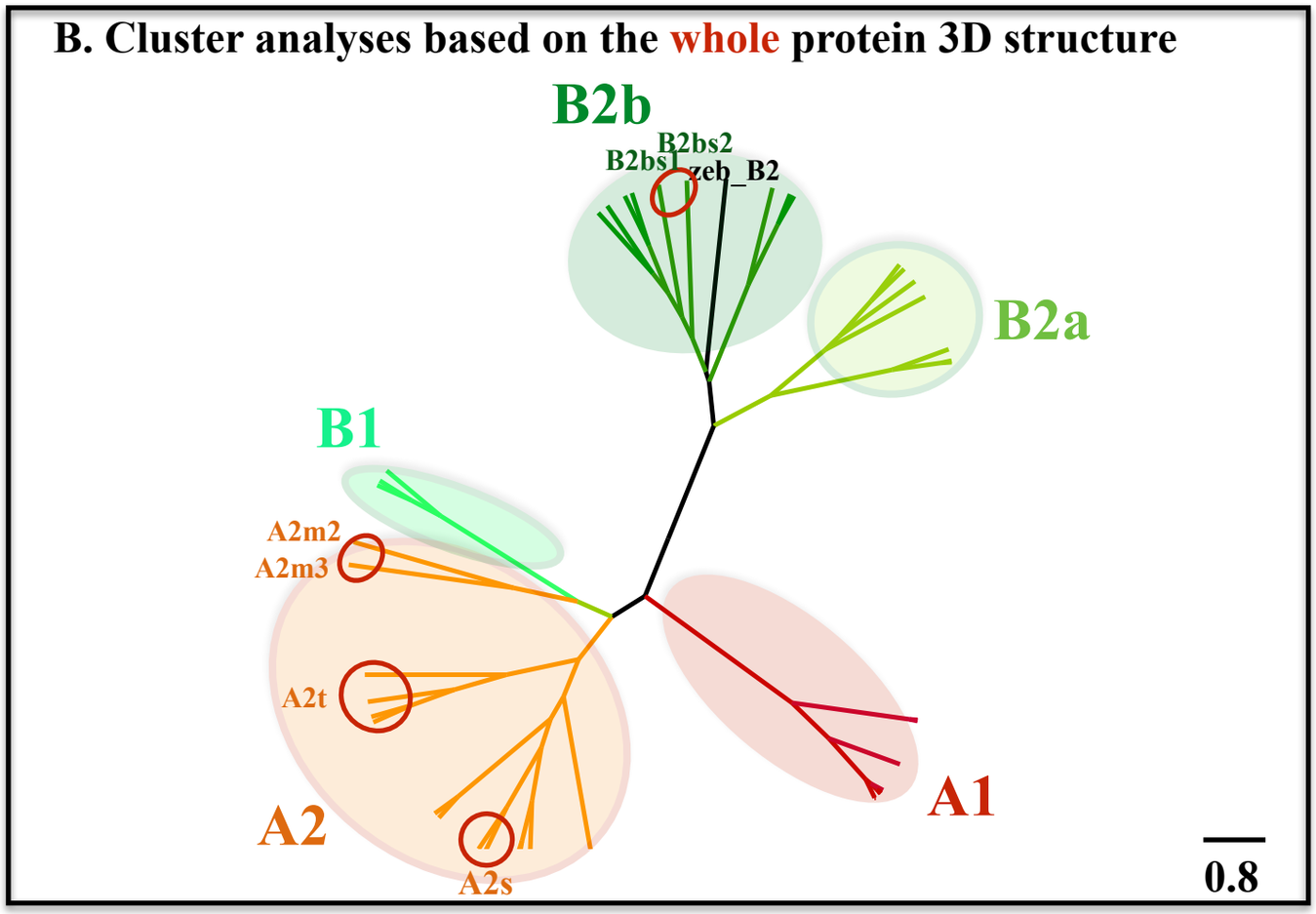

A
Astyanax mexicanus
Dario rerio
Gadus morhua
Gasterosteus aculeatus
Takifugu rubripes
Tetraodon nigroviridis
Oreochromis niloticus
Oryzias latipes
Poecilia formosa
Xiphophorus maculatus
TGD
Cluster I
Cluster II
Lepisosteus oculatus
Latimeria chalumnae
500bp
B
cichlids specific duplication
copy A
copy A2t1
copy A2t2
copy A2t3
copy A2t4
copy A2
copy A1
copy B2b
copy B2a
copy B2
copy B1
copy B
Latimeria chalumnae
Lepisosteus oculatus
0.2

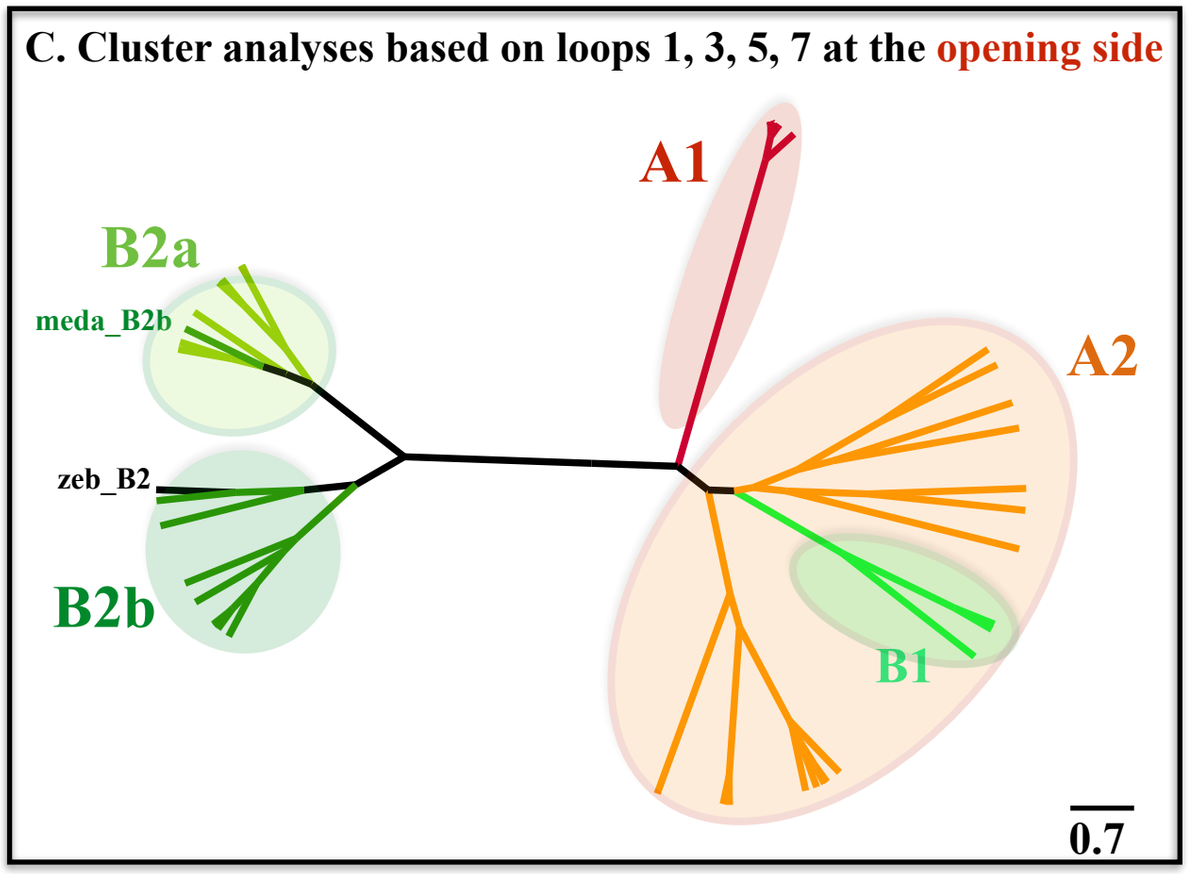

C
A1 B2 B2a B2b A2 B1
Danio rerio
Astyanax mexicanus
Osmerus eperlanus
Borostomias antarcticus
Parasudis fraserbrunneri
Guentherus altivela
Benthosema glaciale
Percopsis transmontana
Typhlichthys subterraneus
Polymixia japonica
Zeus faber
Cyttopsis roseus
Stylephorus chordatus
Trachyrincus scabrus
Trachyrincus murrayi
Muraenolepis marmoratus
Bathygadus melanobranchus
Macrourus berglax
Malacocephalus occidentalis
Melanonus zugmayeri
Merluccius merluccius
Merluccius capensis
Phycis blennoides
Trisopterus minutus
Merlangius merlangus
Arctogadus glacialis
Gadus morhua
Brosme brosme
Molva molva
Regalecus glesne
Lampris guttatus
Monocentris japonica
Rondeletia loricata
Acanthochaenus luetkenii
Holocentrus rufus
Neoniphon sammara
Myripristis jacobus
Brotula barbata
Lamprogrammus exutus
Carapus acus
Lesueurigobius cf. sanzi
Thunnus albacares
Selene dorsalis
Helostoma temminckii
Anabas testudineus
Parablennius parvicornis
Chromis chromis
Pseudochromis fuscus
Xiphophorus maculatus
Poecilia formosa
Oryzias latipes
Oreochromis niloticus
Symphodus melops
Spondyliosoma cantharus
Antennarius striatus
Takifugu rubripes
Tetraodon nigrovirides
Gasterosteus aculeatus
Sebastes norvegicus
Neoteleostei
Paracanthopterygii
Zeiogadaria
Ctenosquamata
Gadariae
Acanthomorphata
Gadiformes
Gadidae
Acanthopterygii
Euacanthomorphacea
Percomorphaceae

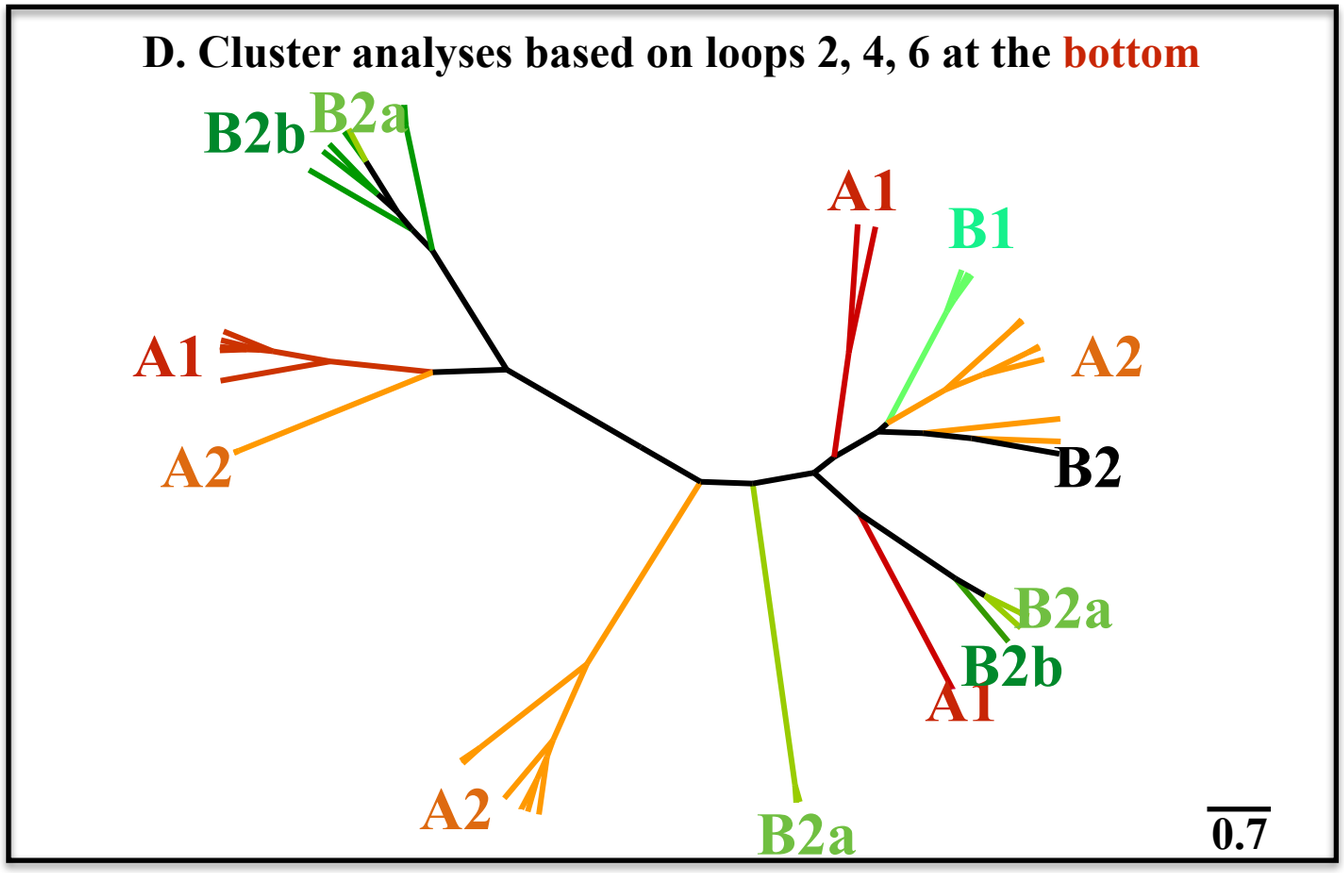

A Syntenic analysis of genome region possessing ApoD clusters in teleost fishes with gar and chicken
medaka-gar
stickleback-gar
tilapia-gar
zebrafish-gar
medaka-chicken
stickleback-chicken
tilapia-chicken
zebrafish-chicken
ApoD clusters as the breakpoints
B ApoD clusters as the breakpoints of genome rearrangement beforeTGD
a. Genome rearrangement before TGD
b. ApoD genes and their neighbours at the breakpoints
cluster I
cluster II
TGD
genome rearrangement with ApoD clusters as the breakpoints
ApoD clusters
Lepisosteus oculatus
note: the schematic lengths here do not represent the real lengths
C Inversion occurred again with ApoD clusters as the breakpoints in cichlid fishes
Cluster I
Cluster II
haplochromine
Pundamilia nyererei
Maylandia zebra
Astatotilapia burtoni
Oreochromis niloticus
LG18
LG9
40 kb
30 kb
LG
ApoD cluster
inversion
D
a. Schematic of ApoD protein domains
signal peptide
lipocalin
ca. 16AA
approx. 20AA
approx. 144AA
b. Duplicated ApoD genes binding opportunities prediction
copy A1
copy A2
copy B1
copy B2a
copy B2b
coelacanth
zebrafish
copy A1
MAPK signaling pathway
MAPK signalling pathway
forkhead transcription factors
common
tilapia
platyfish
fugu
cod
stickleback
medaka
zebrafish
coelacanth
apoa1+MAPK
MAPK+forkhead

Protein 3D structure modelling and cluster analyses of *ApoD* genes

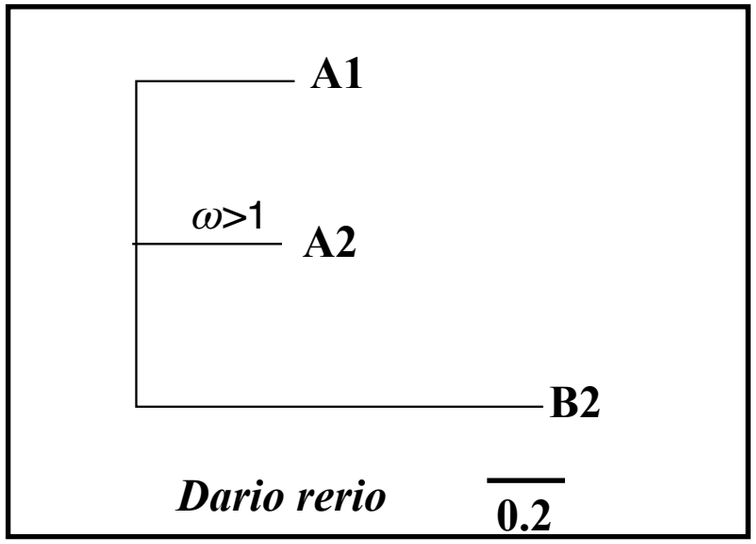


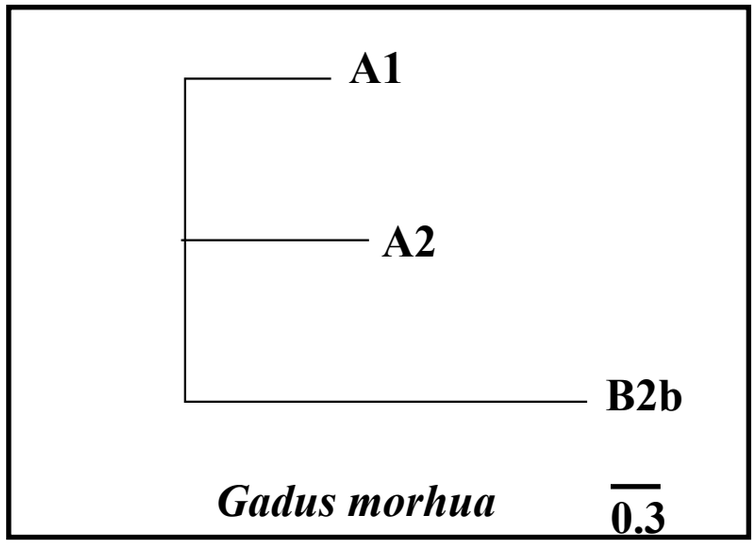


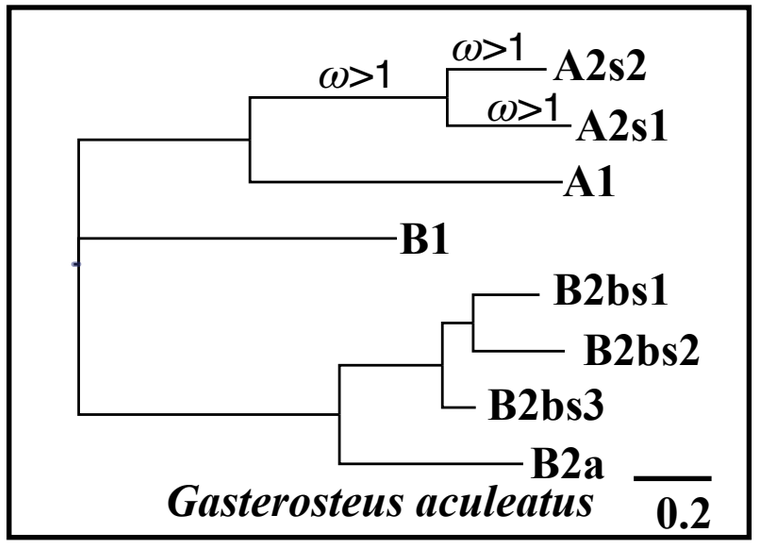


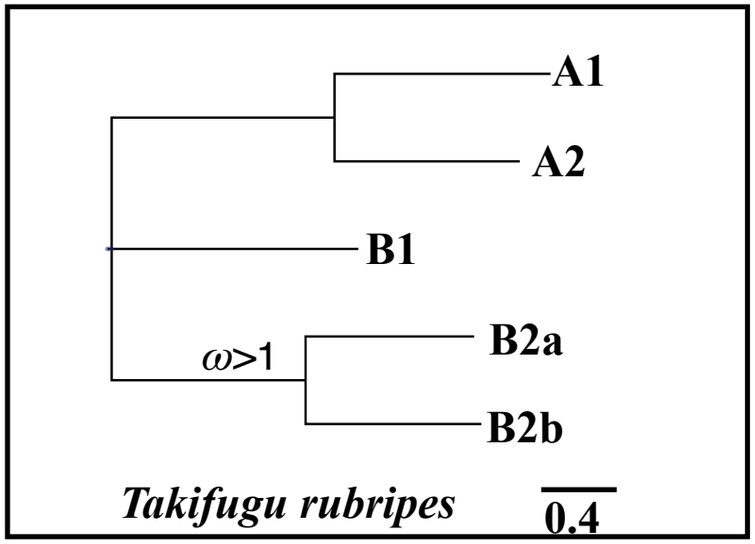


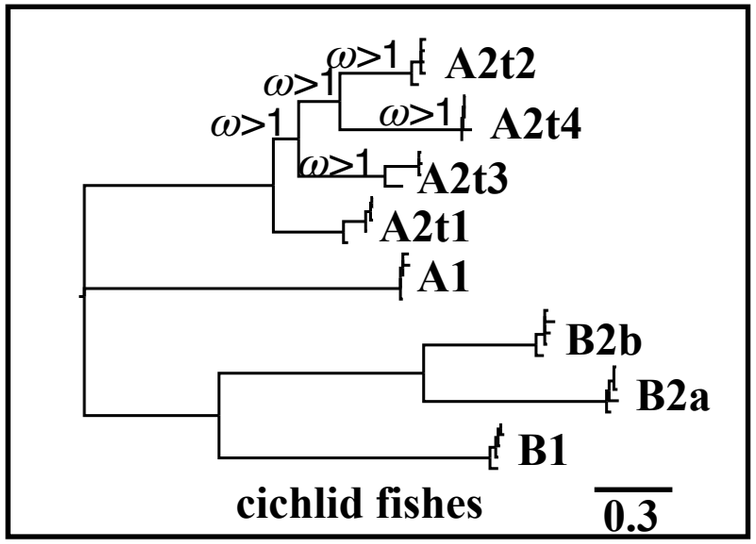
Positive selection detection under branch-site model within codeml in PAML

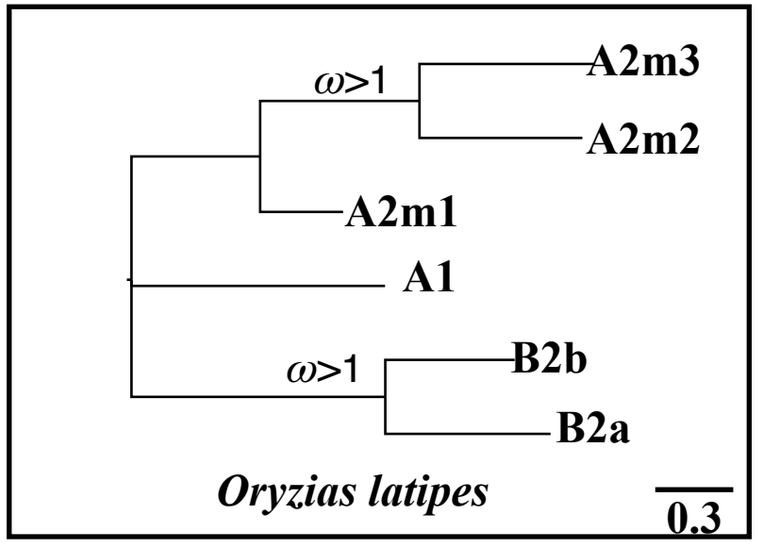


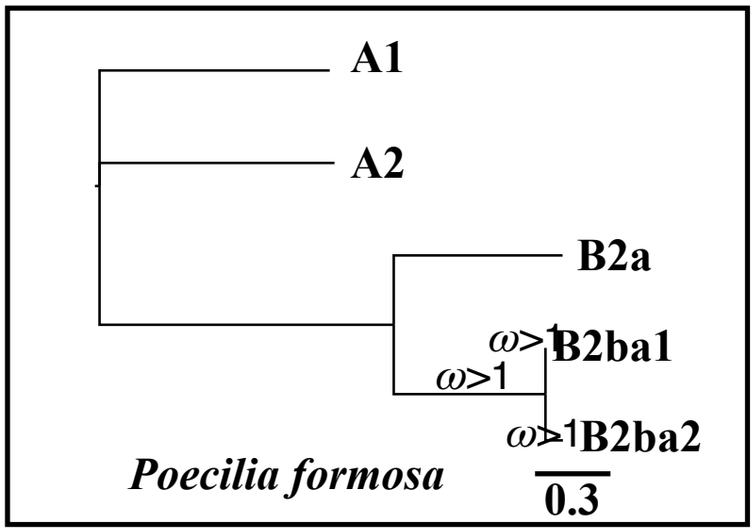


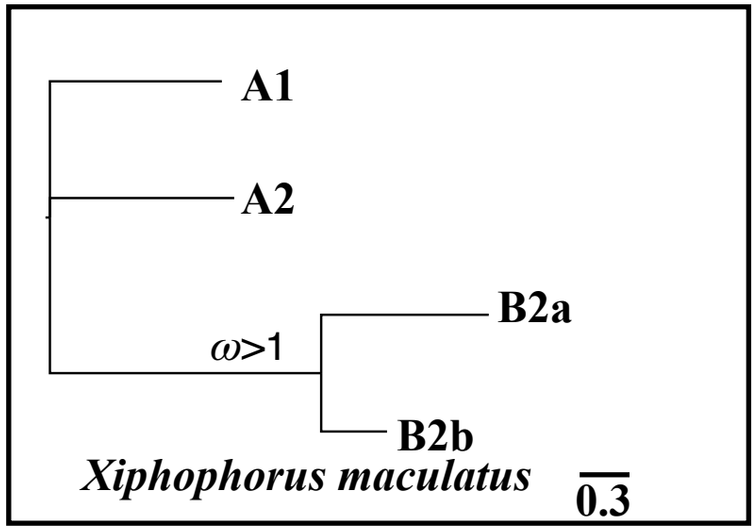




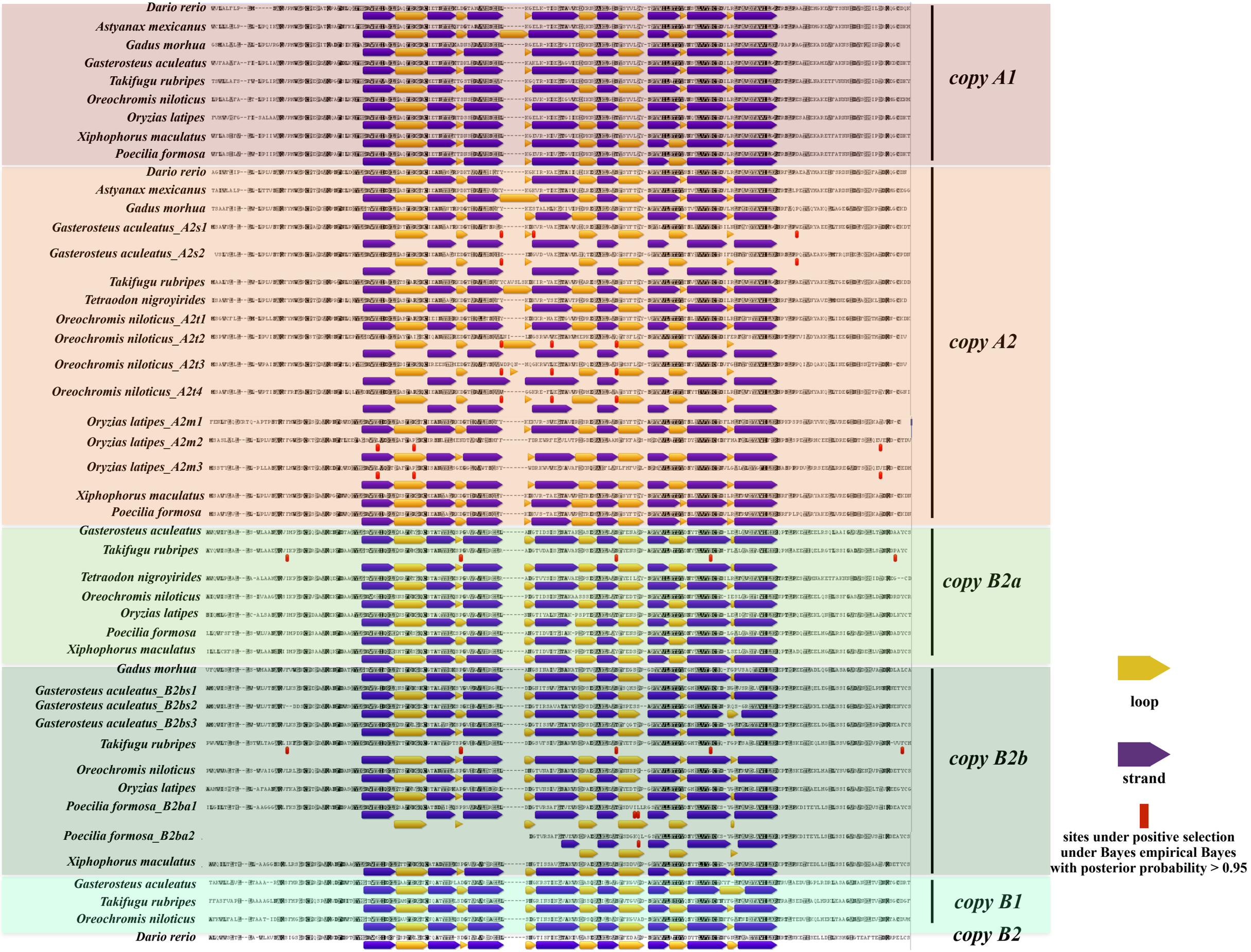


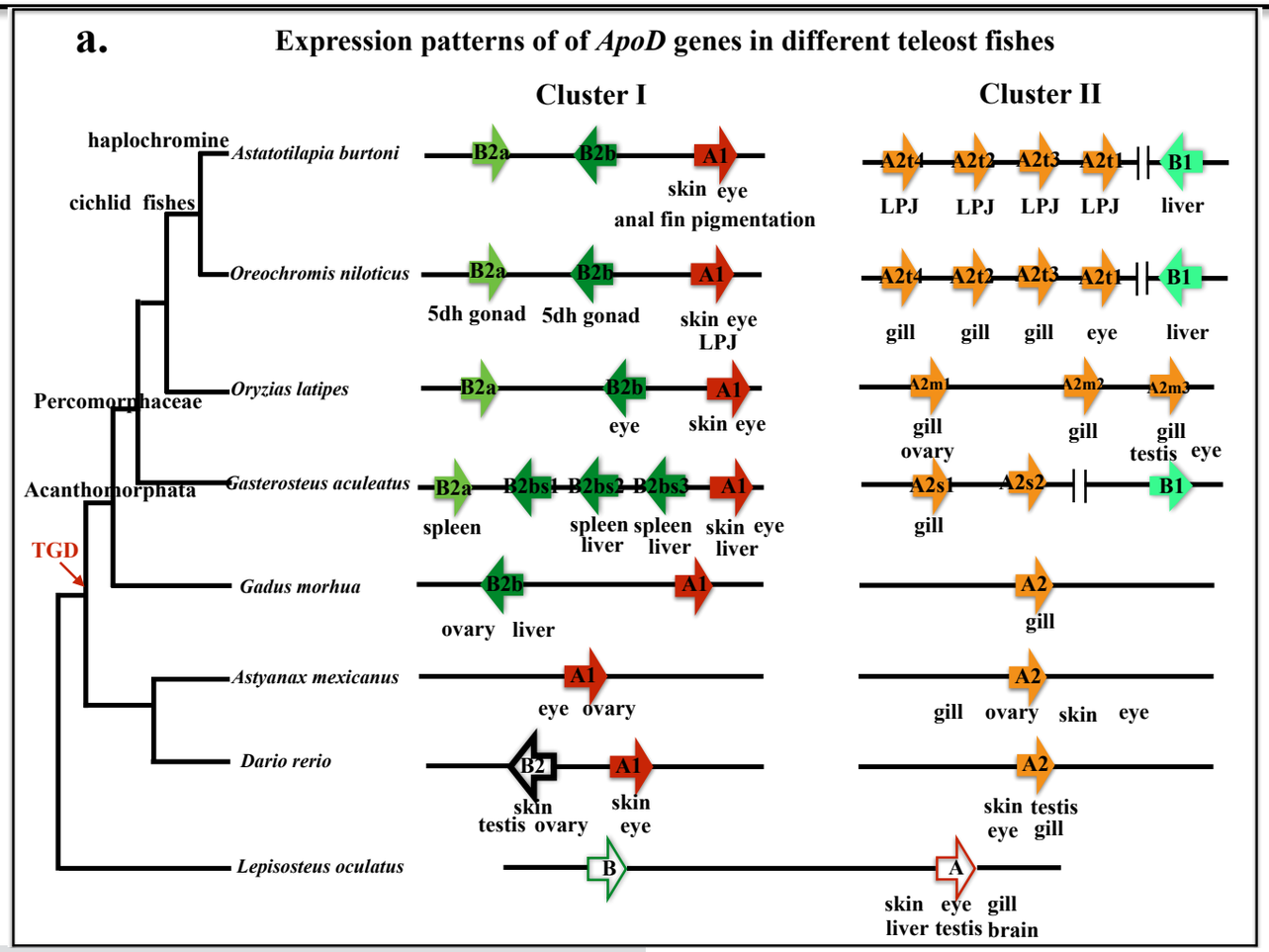


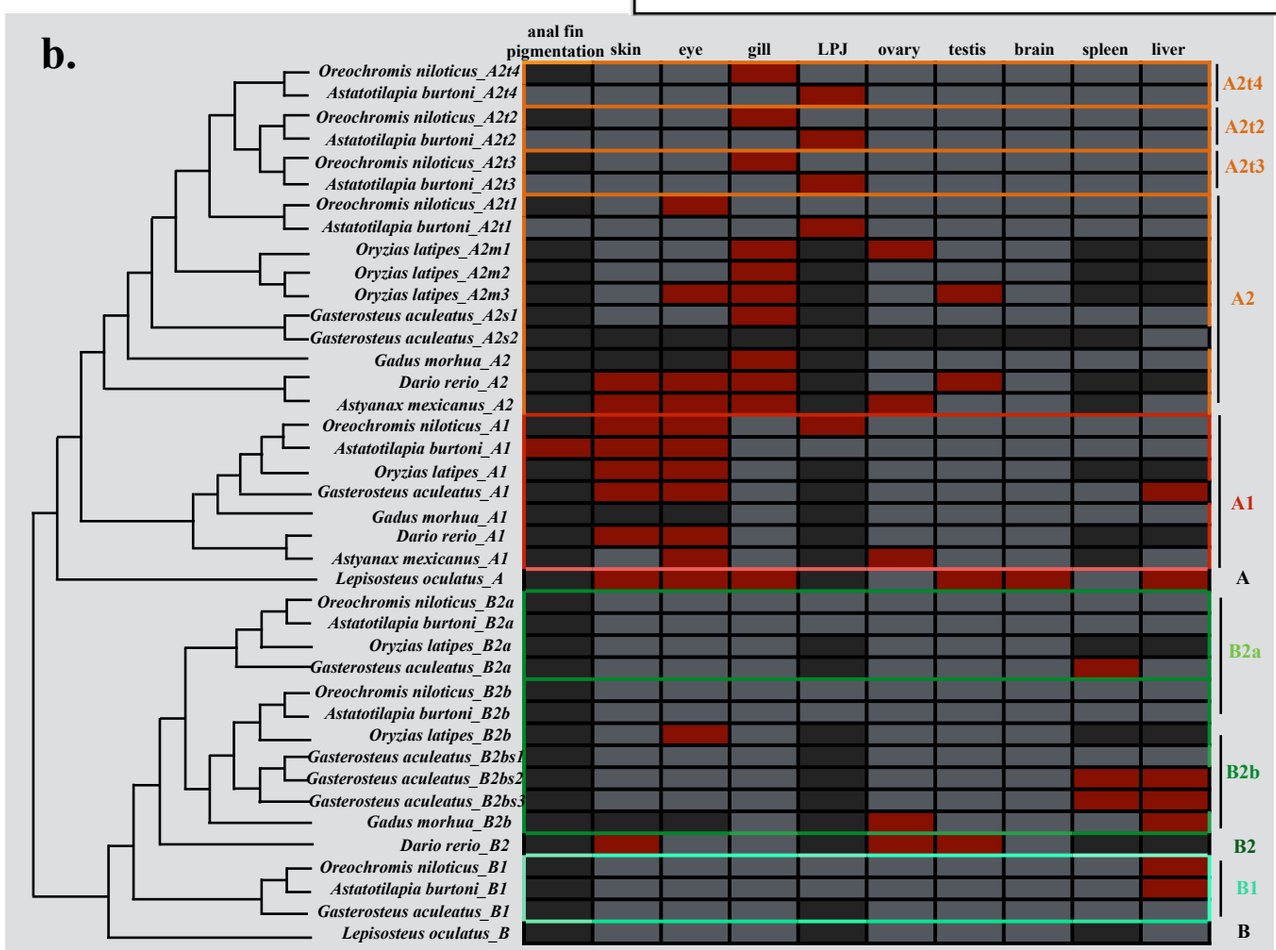


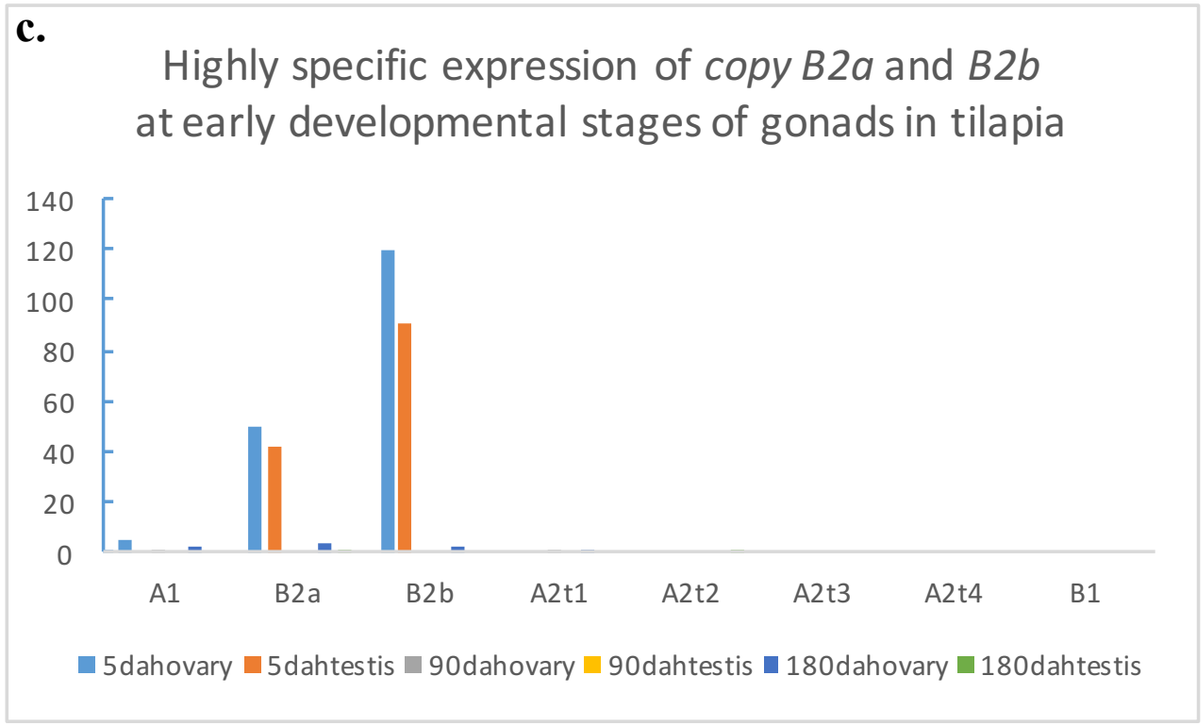


Amino acids alignment of *ApoD* genes among species with protein 3D structure annotation




a.
Expression patterns of of ApoD genes in different teleost fishes
Cluster I
Cluster II
haplochromine
cichlid fishes
Percomorphaceae
Acanthomorphata
TGD
Astatotilapia burtoni
Oreochromis niloticus
Oryzias latipes
Gasterosteus aculeatus
Gadus morhua
Astyanax mexicanus
Dario rerio
Lepisosteus oculatus
B2a
B2b
A1
skin eye
anal fin pigmentation
A2t4
A2t2
A2t3
A2t1
B1
LPJ
liver
5dh gonad
gill
A2m1
A2m2
A2m3
ovary
testis
B2bs1
B2bs2
B2bs3
spleen
A2s1
A2s2
A2
B
brain


b.
anal fin pigmentation
skin
eye
gill
LPJ
ovary
testis
brain
spleen
liver
Oreochromis niloticus_A2t4
Astatotilapia burtoni_A2t4
Oreochromis niloticus_A2t2
Astatotilapia burtoni_A2t2
Oreochromis niloticus_A2t3
Astatotilapia burtoni_A2t3
Oreochromis niloticus_A2t1
Astatotilapia burtoni_A2t1
Oryzias latipes_A2m1
Oryzias latipes_A2m2
Oryzias latipes_A2m3
Gasterosteus aculeatus_A2s1
Gasterosteus aculeatus_A2s2
Gadus morhua_A2
Dario rerio_A2
Astyanax mexicanus_A2
Oreochromis niloticus_A1
Astatotilapia burtoni_A1
Oryzias latipes_A1
Gasterosteus aculeatus_A1
Gadus morhua_A1
Dario rerio_A1
Astyanax mexicanus_A1
Lepisosteus oculatus_A
Oreochromis niloticus_B2a
Astatotilapia burtoni_B2a
Oryzias latipes_B2a
Gasterosteus aculeatus_B2a
Oreochromis niloticus_B2b
Astatotilapia burtoni_B2b
Oryzias latipes_B2b
Gasterosteus aculeatus_B2bs1
Gasterosteus aculeatus_B2bs2
Gasterosteus aculeatus_B2bs3
Gadus morhua_B2b
Dario rerio_B2
Oreochromis niloticus_B1
Astatotilapia burtoni_B1
Gasterosteus aculeatus_B1
Lepisosteus oculatus_B
A2t4
A2t2
A2t3
A2
A1
A
B2a
B2b
B2
B1
B


c.
Highly specific expression of copy B2a and B2b
at early developmental stages of gonads in tilapia
140
120
100
80
60
40
20
0
A1
B2a
B2b
A2t1
A2t2
A2t3
A2t4
B1
5dahovary
5dahtestis
90dahovary
90dahtestis
180dahovary
180dahtestis