The expansion of apolipoprotein D genes in cluster 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

Gene and genome duplication play an important role in the evolution of gene functions. Compared to an individual duplicated gene, gene clusters attract more attention, especially regarding their associations with innovation and adaptation. Here, we report, for the first time, the expansion of a gene family specific to teleost fishes, apolipoprotein D (*ApoD*) gene family. The only *ApoD* gene in the ancestor was expanded in two clusters via genome duplication and tandem gene duplication in teleost fishes, with a rather dynamic evolutionary pattern. Based on comparative genomic and transcriptomic analyses, protein 3D structure simulation, evolutionary rate detection and genome structure detection, subfunctionalization and neofunctionalization after duplication were observed both at the protein and expression levels, especially for lineage-specific duplicated genes that were under positive selection. Orthologous genes in the same physical order exhibited conserved expression patterns but became more specialized with the increasing number of duplicates. Different *ApoD* genes were expressed in tissues related to sexual selection and adaptation. This was particularly true for cichlid fishes, whose paralogues in different clusters showed high expression in anal fin pigmentation patterns (sexual selection related traits) and the lower pharyngeal jaw (related to feeding strategy), the two novelties famous for adaptive radiation of cichlid fishes. Interestingly, ApoD clusters are located at the breaking point of genome rearrangement. Since genome rearrangement can capture locally adapted genes or antagonous sex determining genes to protect them from introgression by reducing recombination, it can promote divergence and reproductive isolation. This further suggests the importance of the expansion of *ApoD* genes for speciation and adaptation in teleost fishes, especially for cichlid fishes.

Key words

apolipoprotein D, gene cluster, positive selection, breaking point, teleost fishes

Introduction

Gene and genome duplication play an important role 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 might happen after duplication [2]. In many instances, one of the duplicates becomes silenced via the accumulation of deleterious mutations (i.e. pesudogenization or nonfunctionalization [1]). Alternatively, the original pre-duplication function might be subdivided between the duplicates (i.e. subfunctionalization) [3], or one of the duplicates might gain a new function (i.e. neofunctionalization) [4]. Since the probability to accumulate beneficial substitutions is relatively low, examples for neofunctionalization are sparse. There are, nevertheless, examples for neofunctionalization. For example, the duplication of *dachshund (dac)* 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 a functional antifreeze glycoproteins in Antarctic notothenioid fish [6]; and the duplication of *opsin* genes is implicated with trichromatic vision in primates [7]. Another selective advantage of gene duplication can be attributed to the increased number of gene copies themselves, e.g. in the form of gene dosage effects [8][9].

Gene functional changes after duplication can occur at the protein level [6,10,11]. For example, the physiological division of labour between the oxygen-carrier function of haemoglobin and oxygen-storage function of myoglobin in vertebrate [12]; the acquired enhanced digestive efficiencies of duplicated gene encoding pancreatic ribonuclease in leaf monkey [13]. However, the chance to accumulate beneficial alleles is rather low, and thus the functional changes after duplication in protein level are sparse. Instead, changes in the expression level are more tolerable and efficient, since it does not require the modification of coding sequences and can immediately

75 offer phenotypic consequences. Many examples have provided evidence that duplicated genes
76 acquiring new expression domains are linked to the evolution of novel traits (e.g., *dac2*, a novel leg
77 segment in Arachnid [5]; *elnb*, bulbus arteriosus in teleost fishes [14]; *fhl2b*, egg-spots in cichlid
78 fishes [15]).
79
80 In some cases, gene duplication resulted in so-called gene clusters: genes from the same
81 gene family physically closely linked in one chromosome [16] have attracted considerable attention,
82 such as Hox gene clusters [17], globin gene clusters [18], paraHox gene clusters [19], MHC clusters
83 [20] and opsin gene clusters [21]. Duplicated genes in clusters are usually related to innovations and
84 adaptation [10,16,21], suggesting their advantageous roles during evolution. The expansion of gene
85 clusters can be traced back to WGD and/or tandem duplication [12,22], and they are suggested to be
86 causally linked to genome instability [23,24]. Actually, if genome rearrangement can capture
87 locally adapted genes or antagonous sex determining genes to protect them from introgression by
88 reducing recombination, it can promote divergence and reproductive isolation [25] and thus
89 contribute to speciation and adaptation, such as in butterfly [26], fish [27], mosquitoes [28] and
90 Atlantic cod [29]. However, few studies investigated the roles of gene clusters at the breaking point
91 of genome rearrangement in speciation and adaptation.
92
93 Here, we report, for the first time, the expansion of a gene family, apolipoprotein D (*ApoD*),
94 in teleost fishes. *ApoD* gene belongs to the lipocalin superfamily of lipid transport proteins [30,31].
95 In humans, ApoD was suggested to function as a multi-ligand, multifunctional transporter (e.g.,
96 hormone and pheromone) [31,32], which is important in homeostasis and housekeeping functions in
97 most organs [32]. Tetraodons possess only a single *ApoD* gene, which is expressed in multiple
98 tissues, most notably in brain and testis (see e.g. [31,33,34]) and have been suggested to be involved
99 in the central and peripheral nervous systems [31]. Interestingly, teleost fishes possess varying

100 numbers of duplicates located in two chromosomes (http://www.ensembl.org/). However, no
101 detailed analysis regarding this gene family in fishes has been reported. Recently, we found that the
102 orthologous *ApoD* gene was highly expressed in convergent evolved innovative anal fin
103 pigmentation patterns in cichlid fishes [35], which inspired us to further investigate the expansion
104 of *ApoD* genes in teleost fishes and their roles in speciation and adaptation.
105
106 **Results**
107
108 **1. The expansion of *ApoD* genes in two clusters in teleost fishes**
109
110 Based on phylogenetic reconstruction of all *ApoD* genes with available assembly genome
111 data and *in sillico* screen in teleost fishes, we traced the evolutionary history of *ApoD* gene family
112 (Figure 1). There is only one *ApoD* gene in coelacanth (*Latimeria chalumnae*). Tandem duplication
113 of the ancestral *ApoD* gene gave rise to two copies (*A* and *B*) in spotted gar (*Lepisosteus oculatus*)
114 located in one cluster. In the course of teleost-specific genome duplication (TGD) and tandem
115 duplication in a lineage specific manner, *ApoD* genes expanded in teleost fishes with different
116 numbers and are located in two clusters, i.e. with two copies in cavefish (*Astyanax mexicanus*) (*A1*
117 and *A2*) and pufferfish (*Tetraodon nigroviridis*) (*B2a* and *A2*), three copies (*A1*, *A2*, *B1*) in
118 zebrafish (*Danio rerio*) and cod (*Gadus morhua*) (*A1, A2* and *B2a*), four copies in platyfish
119 (*Xiphophorus maculatus*) (*A1*, *A2*, *B2a* and *B2b*), five copies in Amazon molly (*Poecilia formosa*)
120 (*A1*, *A2*, *B2a*, *B2ba1*, *B2ba2*) and fugu (*Takifugu rubripes*) (*A1*, *A2*, *B1*, *B2a*, *B2b*), six copies in
121 medaka (*Oryzias latipes*) (*A1*, *A2m1*, *A2m2*, *A2m3*, *B2a*, *B2b*), and eight copies in stickleback
122 (*Gasterosteus aculeatus*) (*A1*, *A2s1*, *A2s2*, *B1*, *B2a*, *B2bs1*, *B2bs2*, *B2bs3*) and tilapia (*Oreochromis*
123 *niloticus*) (*A1*, *A2t1*, *A2t2*, *A2t3*, *A2t4*, *B1*, *B2a*, *B2b*). Orthologous genes in different species are

124 located in the same physical order and lineage-specific duplicated genes are located next to each
125 other (except copy *B1*) (Figure 1).
126
127 Based on *in sillico* screen from draft genomes, the expansion of *ApoD* genes showed a very
128 dynamic evolutionary pattern across the phylogeny (Figure 2). 1) Highly variable copy numbers
129 were found in different lineages, especially in the Paracanthopterygii lineage, in which gene gain
130 and lost frequently occurred. Noticeably, *Stylephorus chordatus* lost all of the *ApoD* genes.
131 Compared to the Paracanthopterygii lineage, the Acanthopterygii lineage possesses fewer tandem
132 duplicates, except in tilapia, medaka and stickleback. 2) Different *ApoD* duplicates showed variable
133 evolutionary conservation. For example, copies *A1* and *A2* present almost across the whole
134 phylogeny. Copy *A1* is more conservative, since no lineage-specific duplicated genes were detected,
135 but copy *A2* exhibited variable lineage-specific copies in different fishes, with the most numbers in
136 tilapia (four copies). Copy *B1* could be evolved in the ancestor of Acanthomorphata lineage or
137 convergently evolved in clade Apercomorphaceae and the two species in the Paracanthopterygii
138 lineage (*Percopsis transmontana* and *Zeus faber*). Noticeably, copy *B1* was absent in the whole
139 clade of Gadiformes but was kept in the clade Acanthopterygii. Only species in the basal lineages
140 (*Danio rerio*, *Osmerus eperlanus* and *Parasudis fraserbrunneri*) possess copy *B*. Copies *B2a* and
141 *B2b* both showed varied copies in different species. However, the co-existence of *B2a* and *B2b* was
142 common in Percomorphaceae. 3) Interestingly, the largest numbers of lineage specific duplicated
143 genes were found in tilapia (copy *A2*), medaka (copy *A2*) and stickleback (copy *B2b*) in
144 Percomorphaceae. The evolutionary rates detection, considering lineage effect under branch-site
145 model, showed that all these lineage-specific duplicated genes were under positive selection (Figure
146 3 and Table 1).
147
148 **2. Neofunctionalization and subfunctionalization of *ApoD* genes at the protein level**

149

150 The protein-protein interaction database analysis showed that the protein domain of *ApoD*
151 genes in different fishes is very conservative, including ca. 20 amino acids as the signal peptide and
152 ca. 144 amino acids as the lipocalin domain (Figure 1). The common associations of different *ApoD*
153 genes are with immunity-related gene *pla2g15* [36,37], high-density lipoprotein biogenesis-related
154 gene *lcat* [38], different copies of *zg16* related to pathogenic fungi recognition [39] and MAPK
155 family involved in multiple signaling pathways and pattern formation [40,41]. Different paralogous
156 *ApoD* genes in teleost fishes also exhibited subdivided associations. One class (copy *A2* and copy
157 *B2a*) is associated with multiple forkhead transcription factors, which play a central role during
158 embryo and adult development [42]. The other class (copy *A1*, copy *B2b* and copy *B1*) lost this
159 association, but is associated with a lipoprotein-related gene *apoa1* [43]. However, the *ApoD* gene
160 at the basal lineage before the duplication in coelacanth possesses both associations, i.e., with
161 forkhead transcription factors and *apoa1*. In addition, *ApoD* in coelacanth is also associated with
162 the genes encoding ligands that can activate Notch signaling pathway (*JAG1*, *JAG2*) [44] and gene
163 belonging to the annexin family (*ANXAII*) (Figure 1). Noticeably, more members of forkhead
164 transcription factors and MAP kinase family are associated with *ApoD* genes after duplication.
165 Unlike in other fishes, copy *A1* in zebrafish has unique interactions with bone resorption-related
166 duplicated genes (*ostf1a*, *ostf1b*, *ostf1c*) [45], cell growth and division-related gene *ppp2cb* [46],
167 neurodevelopment-related gene *rab3gap2* [47] and guanylate-binding protein *gbp1* [48] (Figure 1).
168

169 Homologous protein simulation of different *ApoD* genes showed a conservative structure,
170 including a cup-like central part made by eight antiparallel β-sheet strands and two ends connected
171 by loops (a wide opening part formed by four loops and a narrow closed bottom formed by three
172 loops) (Figure 1). The topology of the neighbor-joining (NJ) tree constructed based on the amino
173 acids composing the cup-like central part is clearly segregated with different duplicates (Figure 1).

174 Unlike the cup-like central part, the loops at the two ends are highly variable. A morphometric data
175 analysis of protein 3D structure of all the loops can segregate different copies, although not for copy
176 *B1*. Further morphometric analysis of 3D data of the loops only composing the opening part of the
177 cup showed clear pattern segregating with different copies, especially for lineage-specific
178 duplicates, although for different loops in different species (Figure 4).
179

180 **3. Neofunctionalization and subfunctionalization of *ApoD* genes at the expression level**

181

182 The gene expression profile analysis in different teleost fishes revealed that orthologous
183 genes showed conserved expression patterns, but lineage-specific duplicated genes evolved more
184 specific and even new expression profiles. For example, copy *A1* is mainly expressed in skin and
185 eye in different teleost fishes, and it was also reported to be highly expressed in the convergent
186 innovative anal fin pigmentation patterns in cichlid fishes [35]. Copy *A2* is mainly expressed in gills
187 and eyes, but its lineage-specific duplicated genes (*A2a*, *A2b*, *A2c*) in a representative
188 haplochromine species, *A. burtoni,* evolved a new expression profile in a key innovative tissue, the
189 lower pharyngeal jaw. Copies *B2a* and *B2b* did not show high expression in adult tissues of we
190 tested here in tilapia; instead, they showed specific high expression in the early developmental stage
191 (5day) of gonads in tilapia (Figure 5). The three lineage-specific duplicated genes, *B2bs1*, *B2bs2*,
192 and *B2bs3*, and their paralogous gene *B2a* are highly expressed in the spleen and liver in
193 stickleback. In medaka, the lineage-specific duplicated genes, *A2m1, A2m2* and *A2m3*, were highly
194 expressed in gills. Gene copy *B1* showed highly specific expression in the liver. Unlike in other
195 teleost fishes, the copies in zebrafish and cavefish showed variable and overlapping expression
196 profiles, including skin, eyes, gills and gonad tissues (Figure 5).

197 **4. *ApoD* genes at the breaking point of genome rearrangement, especially for cichlid**
198 **fishes**
199
200 Syntenic analyses revealed that the two regions containing ApoD clusters in different teleost
201 fishes are syntenic with the linkage group (LG)s/chromosomes containing *ApoD* in gar (LG14) and
202 chicken (chr9), respectively. In addition, genome rearrangement occurred in the chromosomes (or
203 LGs) containing ApoD clusters, involving syntenic regions with chromosomes LG10, LG9, LG19
204 and LG3 in gar, and chr8, chr2, chr28 and chr1 in chicken (Figure 3). Interestingly, ApoD gene
205 clusters were just located at the breaking point, with conserved neighbour genes across different
206 species (Figure 3). A further analysis of available cichlid genomes showed that genome
207 rearrangement occurred again in the derived lineages of cichlid fishes, with ApoD clusters at the
208 breaking point but not in the basal species (Figure 6).
209
210 **Discussion**
211
212 This study is the first one that report the expansion of *ApoD* genes in teleost fishes.
213 Followed by TGD, lineage-specific duplication and genome rearrangement, different numbers of
214 duplicated *ApoD* genes are located in two clusters in different teleost fishes. Protein 3D structure
215 simulation and gene expression profile analyses revealed that subfunctionalization and
216 neofunctionalization occurred at both the protein and expression levels after duplication. Lineage-
217 specific duplicated *ApoD* genes were under positive selection, suggesting their advantageous role
218 during evolution. Genome rearrangement with ApoD clusters at the breaking points further suggest
219 the importance of the expansion of *ApoD* genes for speciation and adaptation in teleost fishes,
220 especially for cichlid fishes.

221 **Dynamic evolutionary pattern of *ApoD* genes in teleost fishes**

222

223 Based on *in sillico* screen, phylogenetic reconstruction and syntenic analysis, *ApoD* gene
224 clusters showed highly dynamic evolutionary patterns in teleost fishes. Copy *A1* showed the most
225 conservative pattern. Considering its shortest branch in the phylogeny, conserved number and
226 expression pattern, copy *A1* resembles the ancestral status of the *ApoD* gene. Unlike the parental
227 copy *B2*, which only showed up in the basal lineages, paralogues *B2a* and *B2b*, with tandem
228 duplicates, evolved in the derived lineages. Interestingly, the largest number of lineage-specific
229 duplicated genes evolved in stickleback, medaka and cichlid fishes were under positive selection,
230 indicating their advantageous roles. In addition, Copy *B1* is absent in the basal lineage and the
231 whole Gadiformes lineage, but it was kept in most species of Euacanthomorphacea. Although we
232 do not know its function yet, the expression profile showed that this gene was highly specifically
233 expressed in the liver.

234

235 It has been suggested that the additional TGD help fuel the phenotypic diversification by
236 providing raw genetic materials and may be related to adaptive radiation in teleost fishes [1,49,50].
237 However, the expansion of teleost fishes actually is after TGD, compatible with the "*time lag model*"
238 [51]. The role of tandem duplication should not be ignored. For example, most gene families
239 occurred independently in different lineages, and only a small fraction of gene families with
240 duplications arose with a common ancestor [52]. In our study, new gene expression profiles and
241 protein structures also occurred in lineage-specific genes. However, noticeably, it is after TGD that
242 one of the two clusters evolve relatively freely. It has been suggested that fixation of duplications is
243 much more common in genome regions where rates of mutations are elevated due to the presence of
244 already-fixed duplication, such as the so called "snow-ball" effect [53]. Therefore, TGD could be a

245 crucial prerequisite to prompt the following tandem duplication, which can explain the expansion of
246 the *ApoD* gene family specific to teleost fishes.
247
248 **Subfunctionalization and neofunctionalization of *ApoD* genes at the protein level**
249
250 Functional divergence occurred in *ApoD* genes at the protein level after duplication. The
251 common associations with multiple members of MAP kinase family involved in multiple signaling
252 pathways and pattern formation [40,41], indicating the realted roles in *ApoD* genes. The *ApoD* gene
253 at the basal lineage, before the duplication in coelacanth, possesses both associations with forkhead
254 transcription factors and *apoa1* compared to the subdivided associations in the paralogues after
255 duplicaiton in teleost fishes, indicating subfunctionalization. The association with forkhead
256 transcription factors suggests the important roles of *ApoD* genes during embryo and adult
257 development [42]. Noticeably, more members of forkhead transcription factors and MAPK family
258 associated with *ApoD* genes after duplication; therefore, neofunctionalization could also occur in
259 this case, for example, but not limited to, the dosage effect.
260
261 Detailed analyses of simulated protein 3D structures further supported functional divergence
262 after duplication. Different ApoDs showed very conservative backbone conformation. Indeed, one
263 feature of lipocalin family is their conservative protein structure in spite of sequence divergence
264 [54]. The cup-like central part is used to bind and transport large varieties of ligands, such as
265 hormones, pheromones, etc [55]. The NJ tree constructed based on the amino acids composing this
266 part are clearly segregated with different duplicates, indicating a ligand-binding capabilities
267 subdivision. Noticeably, unlike the conservative cup-like central core, the loops are highly variable.
268 Actually, reshaping the loop region can efficiently generate novel ligand specificities, similar to the
269 binding functions of antibodies [56]. Indeed, as we showed here, morphomotric analyses of the

270 loops can subdivide different duplicates, especially for lineage-specific duplicates. Interestingly,
271 most of the amino acid sites that are under positive selection are located on the loops in cichlid
272 fishes (Figure 6). Reshaping different parts of the protein structure can be an efficient mechansim
273 for functional divergence at the protein level in a short time period. Indeed, this is one of the
274 reasons why lipocalin genes are well adapted to the recognition of individual ligands [55].
275
276 **Subfunctionalization and neofunctionalization of *ApoD* genes at the expression level**
277
278 Functional divergence of *ApoD* genes also occurred at the expression level. The relaxed
279 environment induced by changing expression profiles can further prompt the following mutations
280 accumulation at the protein level. In our study, different copies of *ApoD* genes at the basal lineages
281 (zebrafish and cavefish) exhibited overlapping expression domains but became much more specific
282 as the numbers of tandem duplicates increased, indicating subfunctionalization. For examples, copy
283 *A1* showed a conserved expression pattern in skin and eye in different fishes. Copies *B2a* and *B2b*
284 are highly specifically expressed at the early developmental stage of gonad tissues in tilapia, which
285 are good candidates of sex determination [57]. Copy *B1* is highly specifically expressed in the liver,
286 and copy *A2* is mainly expressed in the gills. Lineage-specific duplicated genes showed redundancy
287 and even new expression profiles. Noticeably, copy *A2s* in *A. burtoni*, a representative of
288 haplochromine lineage that is famous for species richness, showed highly specific expression in the
289 lower pharyngeal jaw. The lower pharyngeal jaw is an apparatus related to feeding stragegy and one
290 of the key innovations related to adaptive radiation in cichlid fishes [58]. Acquiring new expression
291 profiles indicates the neofunctionalization of copy *A2* in *A. burtoni*. Indeed, it has been suggested
292 that with expression changes, tandem duplication can be an important source for novelty emergence
293 [59], especially if they are adaptive.
294

295 **The roles of *ApoD* genes in speciation and adaptation, especially in cichlid fishes**

296

297 Noticeably, the tissues that the *ApoD* genes expressed, such as eye, skin, gill, lower
298 pharyngeal jaw and spleen are related to neural crest cells , a key innovation in vertebrates [60–63].
299 Combining the associations with the MAPK signaling pathway and forkhead transcription factors,
300 as well as their function as pheromone and hormone transporters, the expansion of *ApoD* genes
301 specific to teleost fishes can be important for the key development of fishes. Another important
302 feature of ApoD clusters is that the tissues they expressed are related to sexual selection and natural
303 selection. For example, in cluster II, *ApoD* duplicates were expressed in the tissues that were
304 supposed to be related to ecological adaptation, such as gills, which are important for freshwater-
305 marine water transition [64,65], and the lower pharyngeal jaw, which is related to feeding strategy
306 [58,66]. Thus, we named cluster II as the "adaptation cluster" (Figure 5). In cluster I, *ApoD* genes
307 are mainly expressed in tissues related to sexual selection, e.g., skin [67], eye [68,69], anal fin
308 pigmentation patterns [35] and gonads. This is especially true for the innovative anal fin
309 pigmentation patterns in cichlid fishes, which is one of the key innovations related to adaptive
310 radiation [70]. Thus, we named cluster I as the "sexy cluster" (Figure 5).

311

312 It has been assumed that if genome rearrangement can capture local adaptive genes or genes
313 related to sexual antagonism, it will accelerate divergence by reducing recombination rates, and
314 thus prompt speciation and adaptation, even evolve neo-sex chromosome [71]. ApoD clusters
315 satisfy all of these requirements. The genes in the two clusters under positive selection and related
316 to sexual antagonism are just located at the breaking point of genome rearrangement. Especially for
317 cichlid fishes, inversion occurred again in the derived lineages but not for the basal lineage. As we
318 mentioned above, different paralogues in two clusters (copy *A1* in cluster I and copy *A2* in cluster II)

319 are expressed in anal fin pigmentation patterns and the lower pharyngeal jaw in cichlid fishes.
320 These are two well-known evolutionary novelties related to adaptive radiation [61]. The Location at
321 the breaking point of genome rearrangement can protect these genes from recombination and
322 prompt divergence and speciation. Further detection across the whole phylogeny of cichlid fishes
323 when the 500-genome data come out and functional experiments will provide detailed evidence
324 about the roles of ApoD clusters in teleost fishes, which are our ongoing projects.
325
326 **Conclusion**
327
328 Our study is the first one that report the expansion of *ApoD* genes in teleost fishes. Different
329 duplicated genes are located in two clusters with neo-function and sub-function evolved. Lineage-
330 specific duplicated genes locating at the breaking point of genome rearrangement were under
331 positive selection. This indicates their evolutionary advantages and could speed up speciation.
332 Especially in cichlid fishes, whose paralogous in different clusters were expressed in anal fin
333 pigmentation patterns (related to sexual selection) and the lower pharyngeal jaw (related to feeding
334 strategy), respectively, the two novelties that were supposed to be related to cichlid fish adaptive
335 radiation. The expansion of *ApoD* gene family in teleost fishes thus provides an ideal model to
336 study gene duplication, cluster maintenance, as well as speciation and adaptation.
337
338 **Methods**
339
340 ***In silico* screening and phylogeny reconstruction to infer gene duplication**
341
342 To retrieve *ApoD* gene duplication in teleost fishes, we first extracted orthologues and
343 paralogues of fishes with available genome data from the Ensembl Release 84 [72] and NCBI

344 database (https://www.ncbi.nlm.nih.gov/genome/). To confirm gene copy numbers, all these
345 orthologues and paralogues were used as a query in a tblastx search against the corresponding
346 genomes. For all unannotated positive hits, a region spanning ca. 2 kb were extracted, and open
347 reading frames (ORF) were predicted using Augustus (http://augustus.gobics.de/) [73]. The
348 predicted coding sequence were then BLAST to the existing transcriptome database to retrieve the
349 corresponding cDNA sequences. The cDNAs were then re-mapped to the corresponding predicted
350 genes to recheck the predicted exon-intron boundary. Coding sequence of genes from human and
351 gar and their neighbour genes were used to perform tblastx searches against the genomes of
352 Amphioxus and Lamprey, respectively, to check whether there are unannotated genes or gene losses
353 in Amphioxus and Lamprey. To infer gene duplication, a maximum likelihood (ML) tree was built
354 first with *ApoD* genes retrieved from available assembled genomes using PAUP4.0 [74] with
355 coelacanth as the outgroup. The best-fitting model of nucleotide substitution was determined in
356 jModeltest v2.1.4 [75,76] applying Akaike Information Criterion with bootstrap value of 200.
357
358 To further retrieve *ApoD* genes in other fishes across the whole phylogeny with draft
359 genomes [20], all the sequences retrieved above were used as a query in a tblastx search with
360 threshold *e* value 0.001. The hit scaffolds were retrieved, and genes within scaffolds were predicted
361 using Augustus. The predicted *ApoD* genes were then translated and re-aligned with known *ApoD*
362 ORF to further confirm exon-intron boundary. All the predicted orthlogues and paralogues were
363 used as a query again in a tblastx search against the corresponding genome data until no more *ApoD*
364 genes were predicted. NJ trees were built with Mega7 [77] to infer gene duplication.
365
366 **Positive selection detection**
367

368 To examine whether *ApoD* duplicates underwent adaptive sequence evolution in a specific
369 branch, branch-site model was used to test positive selection affecting a few sites along particular
370 lineages (foreground branches) in codeml within PAML [78,79]. Rates of non-synonymous to
371 synonymous substitutions (ω or dN/dS) with a priori partitions for foreground branches (see PAML
372 manual). All the model comparisons in PAML were performed with fixed branch lengths
373 (fix_blength=2) derived under M0 model in PAML. Alignment gaps and ambiguity characters were
374 removed (Cleandata=1). A likelihood ratio test was used to calculate a *chi*-square approximation.
375 The Bayes empirical Bayes (BEB) was used to identify sites that are under positive selection.
376
377 **Protein-protein interaction prediction, protein 3D structure simulation and morphometric**
378 **analyses**
379
380 Protein domains of *ApoD* genes and protein-protein interactions were predicted with Simple
381 Modular Architecture Research Tool (SMART) [80,81] and stringdb database http://string-db.org/
382 [82]. Protein 3D structures were predicted with Swiss-model [83–85] using human ApoD crystal
383 protein structure (PDB ID 2hzr) as the template. The results were further visualized, evaluated and
384 analyzed with Swiss-PdbViewer [86]. The amino acids composing the cup-like central part of each
385 ApoD protein structure were extracted to build the NJ tree. To get morphometric 3D data of
386 different ApoD loops, the protein structures were first imposed to human ApoD as the reference.
387 Then, the locations of α-carbon were extracted and further analyzed using hcluster function in R
388 package [87].
389
390 **Gene expression profile analysis**
391

392 Raw transcriptome data from zebrafish, cavefish, cod, medaka, tilapia and *A. burtoni* were
393 retrieved from NCBI https://www.ncbi.nlm.nih.gov/ (Additional file 1). Raw reads were mapped to
394 the corresponding cDNAs (http://www.ensembl.org/) to calculate RPKM (Reads Per Kilobase per
395 Million mapped reads) value. Quantitative polymerase chain reaction (qPCR) was used to detect
396 expression profiles of different *ApoD* duplicates in the tissues that are not included in the existing
397 transcriptomes (Additional file 1). Prior to tissue dissection, specimens were euthanized with MS
398 222 (Sigma-Aldrich, USA) following an approved procedure (permit nr. 2317 issued by the
399 cantonal veterinary office, Switzerland; Guidelines for Care and Use of Laboratory Animals
400 prescribed by the Regulation of Animal Experimentation of Chongqing, China). RNA isolation was
401 performed according to the TRIzol protocol (Invitrogen, USA). DNase treatment was performed
402 with DNA-free™ Kit (Ambion, Life Technologies). RNA quantity and quality was determined with
403 a NanoDrop1000 spectrophotometer (Thermo Scientific, USA). cDNA was produced using the
404 High Capacity RNA-to-cDNA Kit (Applied Biosystems, USA). Housekeeping gene elongation
405 factor 1 alpha *(elfa1)* [88], ubiquitin (*ubc*) [89] and ribosomal protein L7 (*rpl7*) [15] [90] were used
406 as endogenous control. qPCR were performed on a StepOnePlusTM Real-Time PCR system
407 (Applied Biosystems, Life Technologies) using the SYBR Green master mix (Roche, Switzerland)
408 with an annealing temperature of 58°C and following the manufacture's protocols. Primers are
409 available in Additional file 2.
410
411 **Syntenic analyses and inversion detection**
412
413 To further confirm that gene duplication is resulted from TGD, gene region adjacent to the
414 duplicates, as well as outgroup species that did not experience TGD, such as gar and chicken were
415 retrieved. To this end, a window of 5Mb around ApoD clusters in teleost fishes as well as the
416 corresponding chromosomes in gar and chicken were retrieved from Ensemble database and NCBI,

417 respectively. Syntenic analyses in these chromosomes were performed using SyMap [91]. To
418 further detect the structural variation around *ApoD* genes in cichlid fishes, we retrieved the
419 available cichlid genomes raw data from [92]. Delly [93] was used to detect the inversion and the
420 corresponding breaking point around ApoD clusters in tilapia as the reference.
421

422 **Author's contributions**

423 LG discovered the *ApoD* gene family, did the data analysis and wrote the manuscript. LG
424 did cluster analysis of morphometric 3D loop data with the help of CX. This work was supported by
425 the PhD grant from University of Basel, Switzerland and Postdoctoral funding from Southwest
426 University, China.
427

428 **Acknowledgements**

429 We particularly thank Prof. Walter Salzburger for the valuable suggestions and support. We
430 also thank Prof. Deshou Wang for the sampling of tilapia and laboratory support. Many thanks to
431 Dario Moser, Heinz-Georg Belting, Jing Wei and Hua Ruan for the sampling support. Many thanks
432 for the help from Yang Zhao, Fabrizia Ronco, Attila Rüegg, Adrian Indermaur, Xianbo Zhang and
433 He Ma. Many thanks for the discussion and help from Lukas Zimmerman, Peter Fields, Yuchen
434 Sun, Yanyan Xu, Zihui Zhang and De Chen.

435 **Figures**

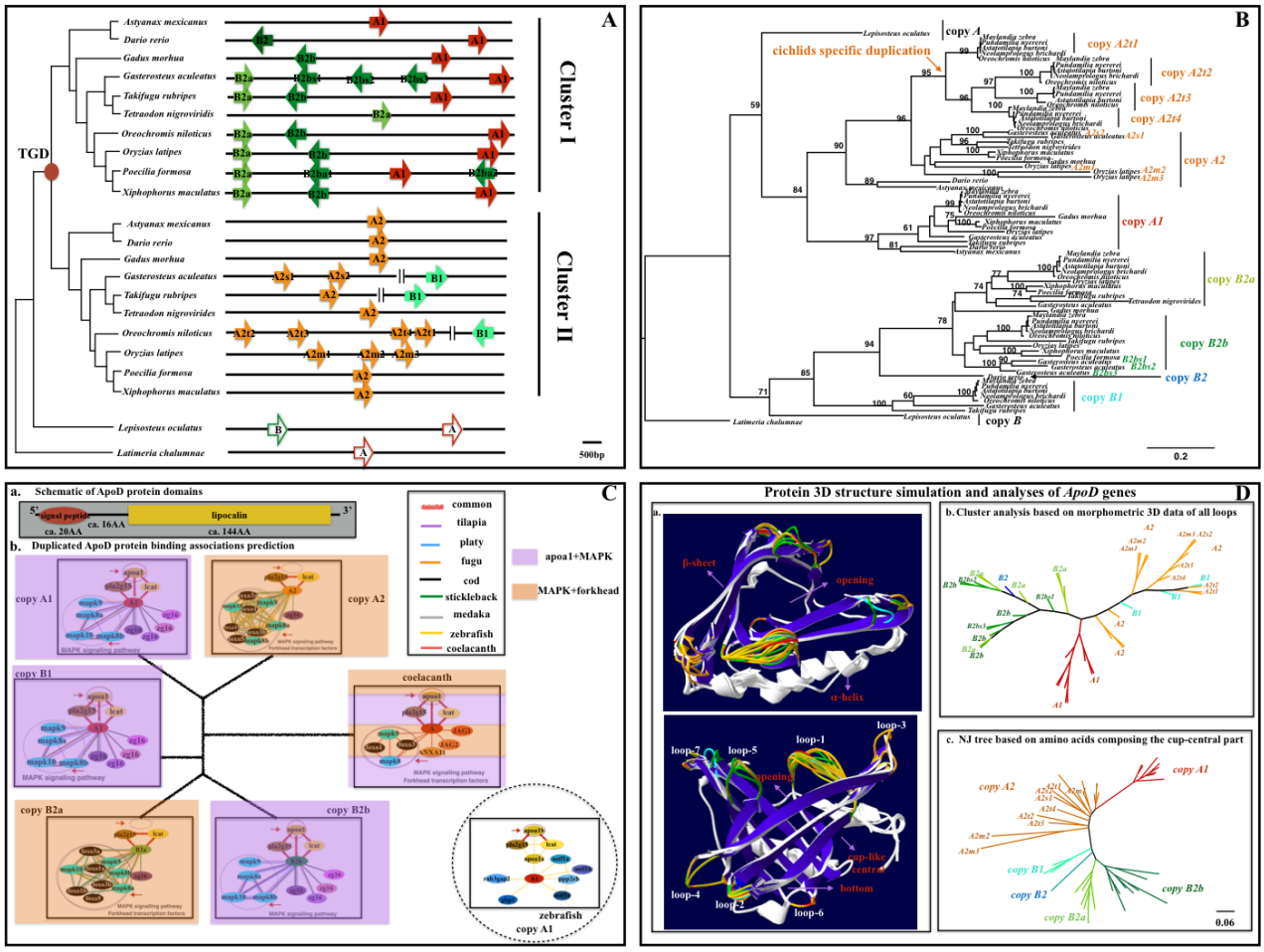
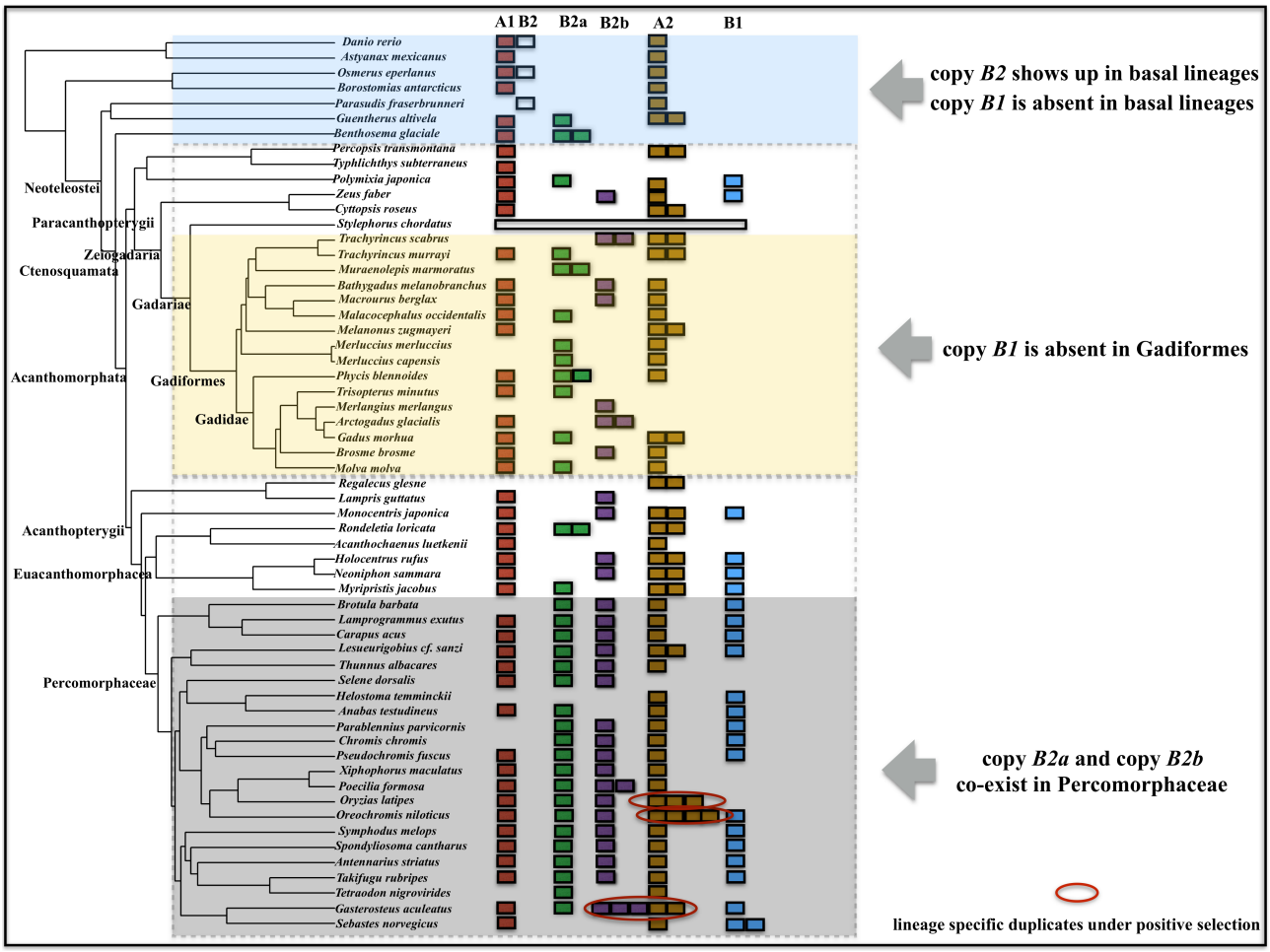


Figure 1 (A) Gene expansion of *ApoD* genes in different teleost fishes in two cluster after teleost-specific duplication (TGD). Each block represents a single gene copy. Phylogeny reconstruction was based on the consensus fish phylogenies [21]. (B) Maximum likelihood phylogenetic tree reconstruction to infer gene duplication. Bootstrap value larger than 50 were marked on the branch. (C) *ApoD* gene structure and association prediction. a. All *ApoD* genes showed conserved domains, with ca. 20 amino acids (AA) as the signal peptide and ca. 144 AA as the lipocalin domain. b. Different paralogous exhibited subdivided associations. The common association is with MAP kinase family members. One class (copy *A2* and copy *B2a*) is associated with multiple forkhead transcription factors. The other class (copy *A1*, copy *B2b* and copy *B1*) lost this association, instead, is associated with a lipoprotein related gene *apoa1*. The *ApoD* gene in coelacanth possesses both associations. (D) Protein structure simulation and analyses. a. 3D structure simulation of ApoD

448 proteins. b. Cluster analysis based on morphometric 3D data of all the loops. c. Neighbour-joining

449 (NJ) tree construction based on amino acids composing the cup-like central part.

450



451

452 **Figure 2** A dynamic evolutionary pattern of Apolipoprotein D (*ApoD*) gene family across the

453 whole phylogeny in teleost fishes. The expansion of *ApoD* genes showed a very dynamic

454 evolutionary pattern across the phylogeny. Highly variable copy numbers were found in different

455 lineages, especially in Paracanthopterygii lineage. Noticeably, species *Stylephorus chordatus* lost

456 all the *ApoD* genes. Different *ApoD* duplicates showed variable evolutionary conservation.

457 Compared to copy *A1*, copy *A2* exhibited more variable lineage-specific duplicates in different

458 fishes, with the most numbers appears in tilapia (four copies). Copy *B1* was absent in the whole

459 clade of Gadiformes. Copy *B* showed up only in species (*Danio rerio*, *Osmerus eperlanus* and

460 *Parasudis fraserbrunneri*) in the basal lineages. The co-existence of *B2a* and *B2b* was common in

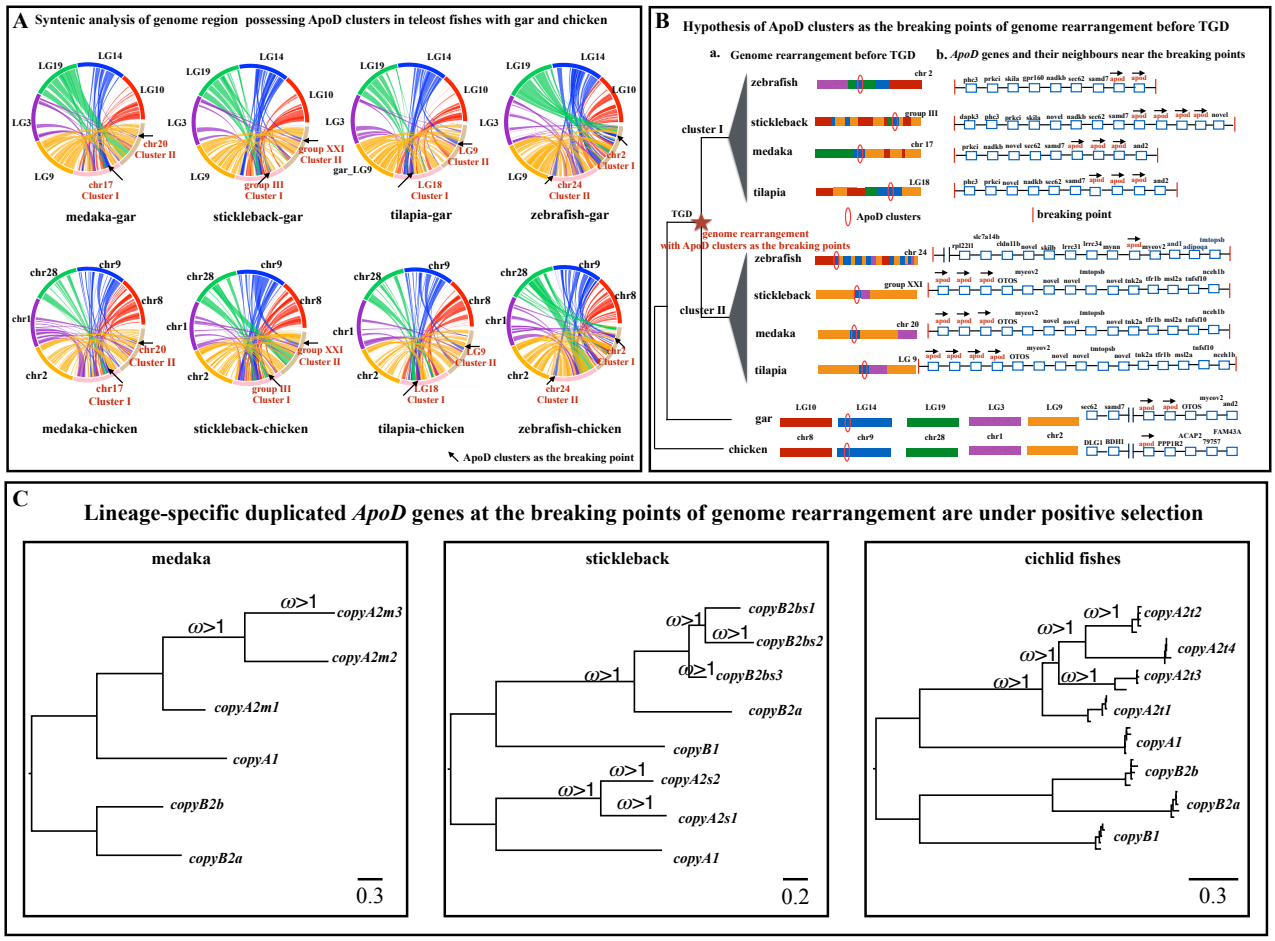
461 Percomorphaceae. The largest numbers of lineage-specific duplicated genes were found in species

462 tilapia (copy *A2*), medaka (copy *A2*) and stickleback (copy *B2b*) in Percomorphaceae. Evolutionary

463 rates detection considering lineage effect under branch-site model showed that these lineage

464 specific duplicated genes were under positive selection.

465



466

467 **Figure 3** (A) Syntenic analyses among teleost fishes, gar and chicken to infer genome duplication.

468 Top is the analyses between teleost fishes and gar. Bottom is the analyses between teleost fishes and

469 chicken. Same color between gar and chicken represents the orthologous chromosome. (B)

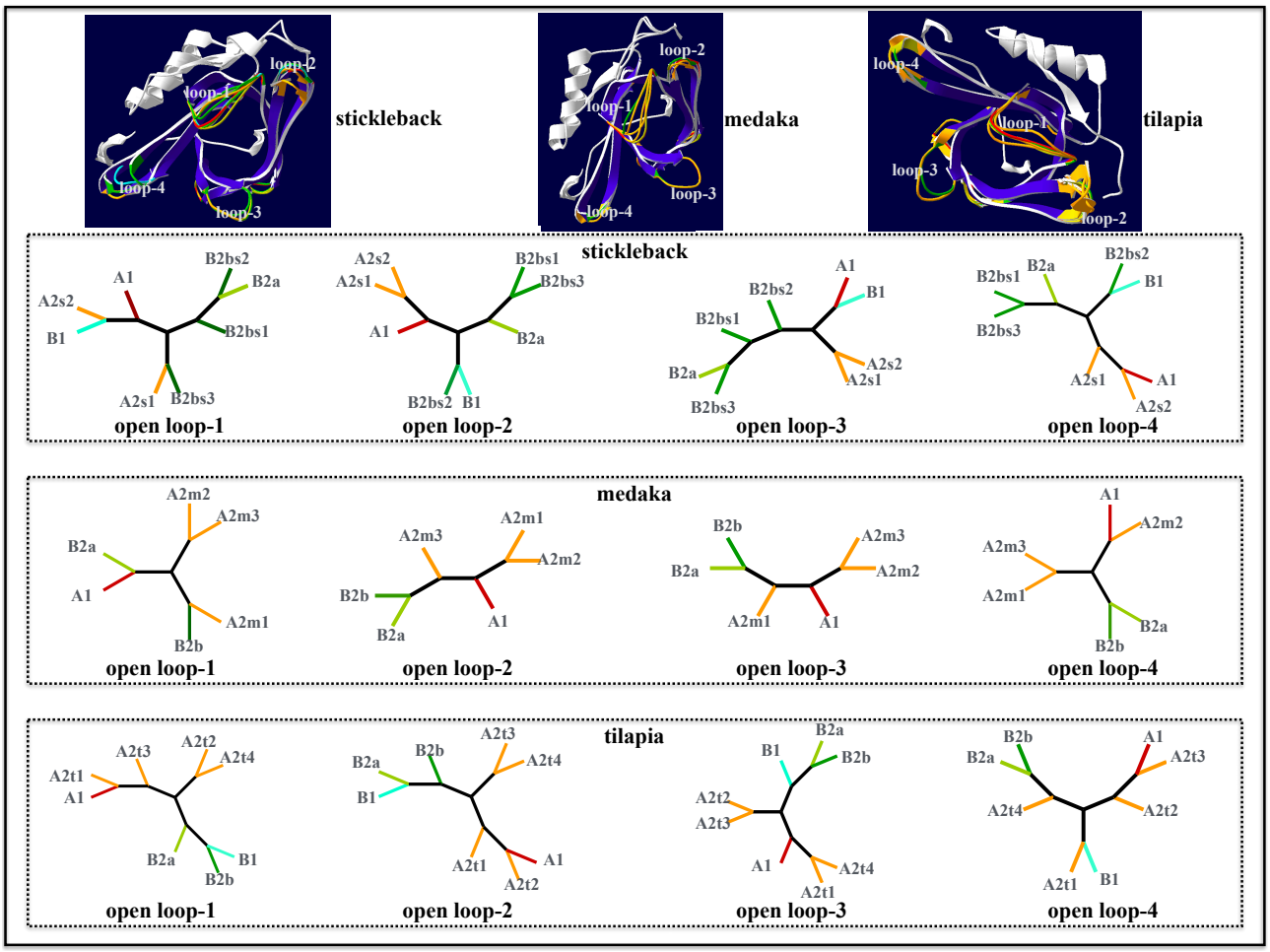
470 Hypothesis of ApoD clusters as the breaking points of genome rearrangement before teleost-

471 specific genome duplication (TGD). (C) Lineage-specific duplicated *ApoD* genes at the breaking

472 points of genome rearrangement are under positive selection under branch-site model within

473 codeml in PAML.

474



475

476 **Figure 4** Morphometric 3D data analysis of protein loops of Apolipoprotein D (ApoD).

477 Morphometric analysis of 3D data of the loops only composing the opening part of the cup showed

478 clear pattern segregating with different copies, especially for lineage-specific duplicates.

479

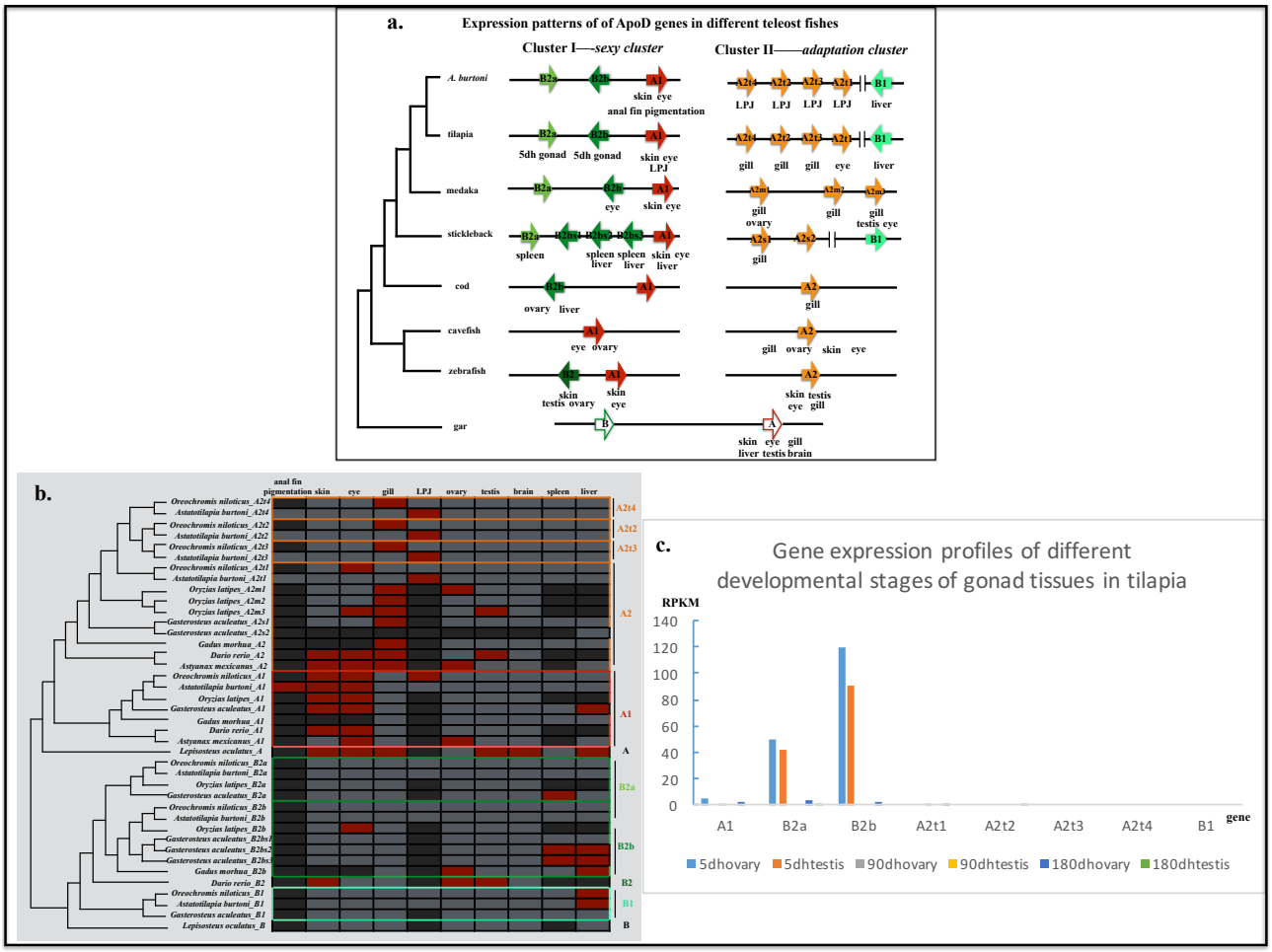
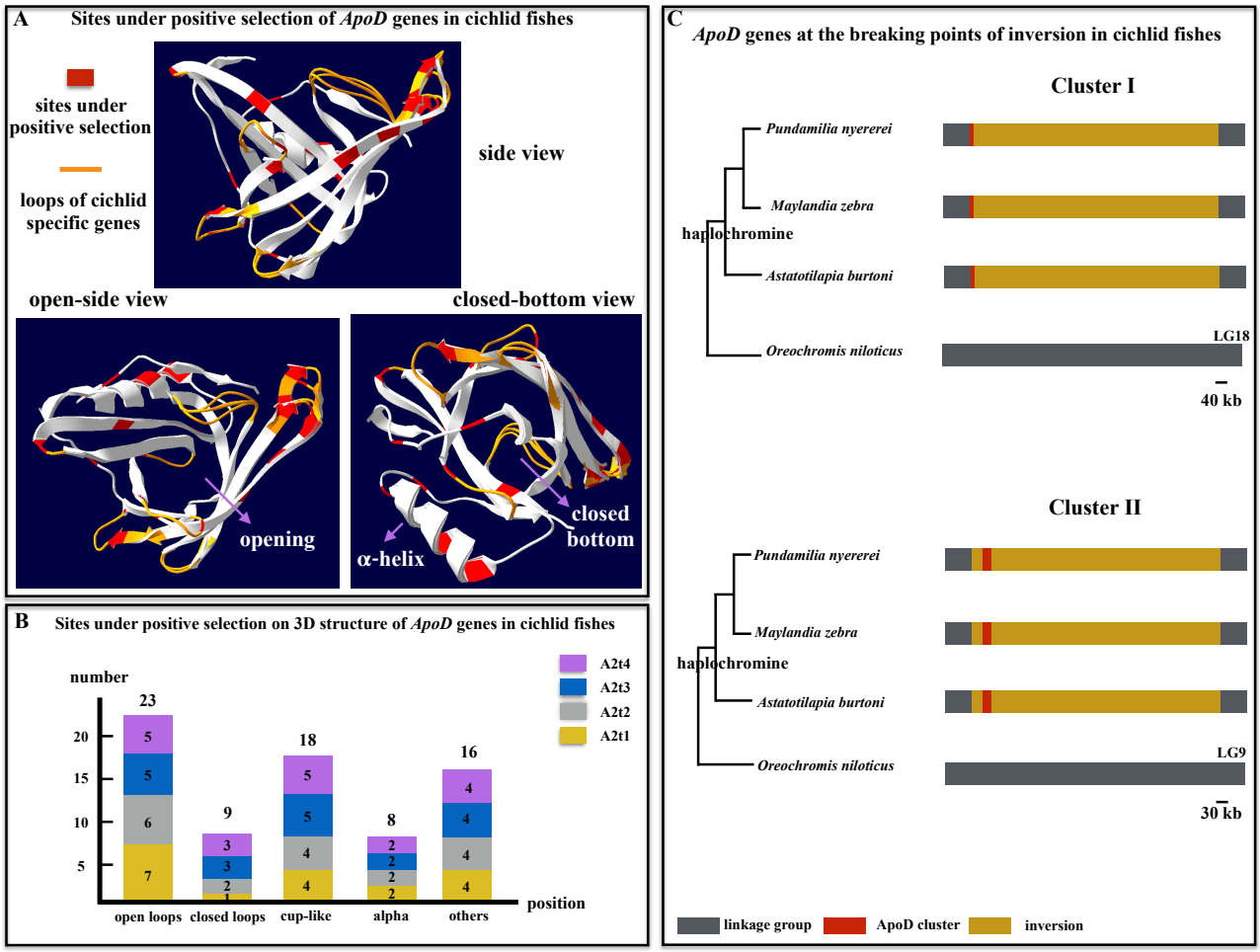


Figure 5 Gene expression profiles for different duplicates in teleost fishes. (a) Schematic figure showing gene expression profiles of different ApoD duplicates in different teleost fishes in two cluster after teleost-specific duplication (TGD). Each block represents a single gene copy. Phylogeny reconstruction was based on the consensus fish phylogenies [21]. (b) Detailed results of gene expression profiles of different ApoD duplicates in different tissues among teleost fishes. Red color, tissues with high expression level. Dark grey, data unavailable (either because the tissue does not exist in the species, or undetected in this study). Light grey, tissues with low expression level. The high expression level in anal fin pigmentation patterns were based on the study from [35]. LPJ, lower pharyngeal jaw in cichlid fishes. (c) Gene expression profiles of ApoD genes in different developmental stages of gonad tissues in tilapia based on available transcriptomic data from [94]. RPKM, Reads Per Kilobase per Million mapped reads ; dh, days.

492



493

494 **Figure 6** (A) Most amino acid sites that are under positive selection of Apolipoprotein D (ApoD)

495 composed the loops in cichlid fishes. (B) Inversion happened in derived lineages of cichlid fishes

496 with ApoD clusters at the breaking point.

497

Table 1 Statistics of branch-site model comparisons of ApoD genes in different fishes

foreground branch	branch-site model		
	model A	model B	p/2
A2m2,A2m3	-3429.42	-3431.09	0.034
A2m3	-3435.38	-3436.83	0.045
A2s1, A2s2	-3829.49	-3834.59	<0.01
A2s1	-3831.05	-3836.99	<0.01
A2s2	-3834.94	-3837.34	0.014
B2a,B2bs1,B2bs2,B2bs3	-3822.26	-3825	0.01
B2bs1,B2bs2,B2bs3	-3833.06	-3834.86	0.029
B2bs1,B2bs2	-3831.39	-3834.1	0.01
B2bs2	-3835.96	-3838.07	0.02
A2t1,A2t2,A2t3,A2t4	-5340.2	-5342.63	0.014
A2t2,A2t3,A2t4	-5345.24	-5349.93	<0.01
A2t2,A2t4	-5347.45	-5353.7	<0.01
A2t3	-5351.84	-5354.38	0.012
A2t2	-5353.18	-5356.22	<0.01

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. [Internet]. 2010 [cited 2016 Jan 3];11:97–108. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20051986

3. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J. Preservation of duplicate genes by complementary, degenerative mutations. Genetics [Internet]. 1999 [cited 2016 Apr 13];151:1531–45. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1460548&tool=pmcentrez&rendertype=abstract

4. Rastogi S, Liberles DA. Subfunctionalization of duplicated genes as a transition state to neofunctionalization. BMC Evol. Biol. [Internet]. 2005 [cited 2016 Apr 13];5:28. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15831095

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.

515 Biol. Evol. [Internet]. 2016 [cited 2016 May 28];33:109–21. Available from:
516 http://www.ncbi.nlm.nih.gov/pubmed/26443673
517 6. Chen L, DeVries AL, Cheng C-HC. Evolution of antifreeze glycoprotein gene from a
518 trypsinogen gene in Antarctic notothenioid fish. Proc. Natl. Acad. Sci. [Internet]. 1997 [cited 2016
519 Apr 13];94:3811–6. Available from: http://www.pnas.org/content/94/8/3811.full
520 7. Dulai KS, von Dornum M, Mollon JD, Hunt DM. The Evolution of Trichromatic Color Vision
521 by Opsin Gene Duplication in New World and Old World Primates. Genome Res. [Internet]. 1999
522 [cited 2016 Apr 13];9:629–38. Available from: http://genome.cshlp.org/content/9/7/629.long
523 8. Tang Y-C, Amon A. Gene copy-number alterations: a cost-benefit analysis. Cell [Internet]. 2013
524 [cited 2016 Feb 22];152:394–405. Available from:
525 http://www.sciencedirect.com/science/article/pii/S0092867412014316
526 9. Lin Z, Li W-H. Expansion of hexose transporter genes was associated with the evolution of
527 aerobic fermentation in yeasts. Mol. Biol. Evol. [Internet]. 2011 [cited 2016 May 4];28:131–42.
528 Available from:
529 http://mbe.oxfordjournals.org/content/28/1/131?ijkey=95bcc58fb94a2349718a4dc5c2802d9a66f3d
530 05e&keytype2=tf_ipsecsha
531 10. Baalsrud HT, Voje KL, Tørresen OK, Solbakken MH, Matschiner M, Malmstrøm M, et al.
532 Evolution of Hemoglobin Genes in Codfishes Influenced by Ocean Depth. Sci. Rep. [Internet].
533 2017 [cited 2017 Nov 24];7:7956. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28801564
534 11. Zhang J, Zhang Y, Rosenberg HF. Adaptive evolution of a duplicated pancreatic ribonuclease
535 gene in a leaf-eating monkey. Nat. Genet. [Internet]. 2002 [cited 2016 Mar 9];30:411–5. Available
536 from: http://www.ncbi.nlm.nih.gov/pubmed/11925567
537 12. Storz JF, Opazo JC, Hoffmann FG. Gene duplication, genome duplication, and the functional
538 diversification of vertebrate globins. Mol. Phylogenet. Evol. [Internet]. 2013 [cited 2017 Dec
539 23];66:469–78. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22846683

13. Zhang J. Parallel adaptive origins of digestive RNases in Asian and African leaf monkeys. Nat. Genet. [Internet]. 2006 [cited 2016 Oct 15];38:819–23. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16767103

14. 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. [Internet]. 2016 [cited 2016 May 28];7:10397. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4735684&tool=pmcentrez&rendertype=abstract

15. 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. [Internet]. Nature Publishing Group; 2014 [cited 2016 Apr 12];5:5149. Available from: http://www.nature.com/ncomms/2014/141009/ncomms6149/full/ncomms6149.html

16. Garcia-Fernàndez J. The genesis and evolution of homeobox gene clusters. Nat. Rev. Genet. [Internet]. Nature Publishing Group; 2005 [cited 2016 Apr 27];6:881–92. Available from: http://dx.doi.org/10.1038/nrg1723

17. 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 [Internet]. 1984 [cited 2017 Mar 4];37:409–14. Available from: http://www.ncbi.nlm.nih.gov/pubmed/6327066

18. Proudfoot NJ, Shander MH, Manley JL, Gefter ML, Maniatis T. Structure and in vitro transcription of human globin genes. Science [Internet]. 1980 [cited 2016 Apr 27];209:1329–36. Available from: http://www.ncbi.nlm.nih.gov/pubmed/6158093

19. Brooke NM, Garcia-Fernàndez J, Holland PWH. The ParaHox gene cluster is an evolutionary sister of the Hox gene cluster. Nature [Internet]. 1998 [cited 2017 Mar 4];392:920–2. Available from: http://www.ncbi.nlm.nih.gov/pubmed/9582071

565 20. Malmstrøm M, Matschiner M, Tørresen OK, Star B, Snipen LG, Hansen TF, et al. Evolution of
566 the immune system influences speciation rates in teleost fishes. Nat. Genet. [Internet]. 2016 [cited
567 2017 Mar 28];48:1204–10. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27548311
568 21. Cortesi F, Musilová Z, Stieb SM, Hart NS, Siebeck UE, Malmstrøm M, et al. Ancestral
569 duplications and highly dynamic opsin gene evolution in percomorph fishes. Proc. Natl. Acad. Sci.
570 [Internet]. 2015 [cited 2017 Mar 28];112:1493–8. Available from:
571 http://www.ncbi.nlm.nih.gov/pubmed/25548152
572 22. Rennison DJ, Owens GL, Taylor JS. Opsin gene duplication and divergence in ray-finned fish.
573 Mol. Phylogenet. Evol. [Internet]. 2012 [cited 2018 Jan 14];62:986–1008. Available from:
574 http://www.ncbi.nlm.nih.gov/pubmed/22178363
575 23. Otto SP. The Evolutionary Consequences of Polyploidy. Cell [Internet]. 2007 [cited 2016 Nov
576 25];131:452–62. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0092867407013402
577 24. Hufton AL, Groth D, Vingron M, Lehrach H, Poustka AJ, Panopoulou G. Early vertebrate
578 whole genome duplications were predated by a period of intense genome rearrangement. Genome
579 Res. [Internet]. Cold Spring Harbor Laboratory Press; 2008 [cited 2016 Nov 25];18:1582–91.
580 Available from: http://www.ncbi.nlm.nih.gov/pubmed/18625908
581 25. Kirkpatrick M, Barton N. Chromosome inversions, local adaptation and speciation. Genetics
582 [Internet]. 2006 [cited 2016 Sep 24];173:419–34. Available from:
583 http://www.ncbi.nlm.nih.gov/pubmed/16204214
584 26. Joron M, Frezal L, Jones RT, Chamberlain NL, Lee SF, Haag CR, et al. Chromosomal
585 rearrangements maintain a polymorphic supergene controlling butterfly mimicry. Nature [Internet].
586 2011 [cited 2016 Sep 24];477:203–6. Available from:
587 http://www.ncbi.nlm.nih.gov/pubmed/21841803
588 27. Jones RT, Salazar PA, ffrench-Constant RH, Jiggins CD, Joron M. Evolution of a mimicry
589 supergene from a multilocus architecture. Proc. Biol. Sci. [Internet]. 2012 [cited 2016 Sep
28

590 24];279:316–25. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21676976
591 28. Ayala D, Fontaine MC, Cohuet A, Fontenille D, Vitalis R, Simard F. Chromosomal inversions,
592 natural selection and adaptation in the malaria vector Anopheles funestus. Mol. Biol. Evol.
593 [Internet]. 2011 [cited 2016 Sep 24];28:745–58. Available from:
594 http://www.ncbi.nlm.nih.gov/pubmed/20837604
595 29. Barth JMI, Berg PR, Jonsson PR, Bonanomi S, Corell H, Hemmer-Hansen J, et al. Genome
596 architecture enables local adaptation of Atlantic cod despite high connectivity. Mol. Ecol.
597 [Internet]. 2017 [cited 2017 Nov 24];26:4452–66. Available from:
598 http://www.ncbi.nlm.nih.gov/pubmed/28626905
599 30. Ayrault Jarrier M, Levy G, Polonovski J. ETUDE DES ALPHA-LIPOPROT'EINES
600 S'ERIQUES HUMAINES PAR. Bull. Soc. Chim. Biol. (Paris). [Internet]. 1963;45:703–13.
601 Available from: http://www.scopus.com/inward/record.url?eid=2-s2.0-
602 0000602532&partnerID=tZOtx3y1
603 31. Rassart E, Bedirian A, Do Carmo S, Guinard O, Sirois J, Terrisse L, et al. Apolipoprotein D.
604 Biochim. Biophys. Acta - Protein Struct. Mol. Enzymol. [Internet]. 2000 [cited 2016 May
605 3];1482:185–98. Available from:
606 http://www.sciencedirect.com/science/article/pii/S016748380000162X
607 32. Weech P, Provost P, Tremblay N, Camato R, Milne R, Marcel Y, et al. Apolipoprotein D—An
608 atypical apolipoprotein. Prog. Lipid Res. [Internet]. 1991 [cited 2016 May 3];30:259–66. Available
609 from: http://www.sciencedirect.com/science/article/pii/016378279190023X
610 33. Drayna D, Fielding C, McLean J, Baer B, Castro G, Chen E, et al. Cloning and expression of
611 human apolipoprotein D cDNA. J. Biol. Chem. [Internet]. 1986;261:16535–9. Available from:
612 http://www.scopus.com/inward/record.url?eid=2-s2.0-0022869267&partnerID=tZOtx3y1
613 34. Provost PR, Villeneuve L, Weech PK, Milne RW, Marcel YL, Rassart E. Localization of the
614 major sites of rabbit apolipoprotein D gene transcription by in situ hybridization. J. Lipid Res.

615 [Internet]. 1991;32:1959–70. Available from: http://www.scopus.com/inward/record.url?eid=2-
616 s2.0-0026321849&partnerID=tZOtx3y1
617 35. Gu L, Xia C. Revelation of the Genetic Basis for Convergent Innovative Anal Fin Pigmentation
618 Patterns in Cichlid Fishes. bioRxiv [Internet]. 2017; Available from:
619 http://biorxiv.org/content/early/2017/12/24/165217.abstract
620 36. Gilleron M, Lepore M, Layre E, Cala-De Paepe D, Mebarek N, Shayman JA, et al. Lysosomal
621 Lipases PLRP2 and LPLA2 Process Mycobacterial Multi-acylated Lipids and Generate T Cell
622 Stimulatory Antigens. Cell Chem. Biol. [Internet]. 2016 [cited 2017 Nov 25];23:1147–56.
623 Available from: http://www.ncbi.nlm.nih.gov/pubmed/27662254
624 37. Bailey SD, Xie C, Do R, Montpetit A, Diaz R, Mohan V, et al. Variation at the NFATC2 Locus
625 Increases the Risk of Thiazolidinedione-Induced Edema in the Diabetes REduction Assessment
626 with ramipril and rosiglitazone Medication (DREAM) Study. Diabetes Care [Internet]. 2010 [cited
627 2017 Nov 25];33:2250–3. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20628086
628 38. Fotakis P, Kuivenhoven JA, Dafnis E, Kardassis D, Zannis VI. The Effect of Natural LCAT
629 Mutations on the Biogenesis of HDL. Biochemistry [Internet]. 2015 [cited 2017 Nov 25];54:3348–
630 59. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25948084
631 39. Tateno H, Yabe R, Sato T, Shibazaki A, Shikanai T, Gonoi T, et al. Human ZG16p recognizes
632 pathogenic fungi through non-self polyvalent mannose in the digestive system. Glycobiology
633 [Internet]. 2012 [cited 2017 Nov 25];22:210–20. Available from:
634 http://www.ncbi.nlm.nih.gov/pubmed/21893569
635 40. Plestant C, Anton ES. Scaling the MAPK Signaling Threshold during CNS Patterning. Dev.
636 Cell [Internet]. 2013 [cited 2017 Jun 28];25:221–2. Available from:
637 http://www.ncbi.nlm.nih.gov/pubmed/23673327
638 41. Shvartsman SY, Coppey M, Berezhkovskii AM. MAPK signaling in equations and embryos.
639 Fly (Austin). [Internet]. [cited 2018 Jan 14];3:62–7. Available from:

640 http://www.ncbi.nlm.nih.gov/pubmed/19182542
641 42. Ho KK, Myatt SS, Lam EW-F. Many forks in the path: cycling with FoxO. Oncogene
642 [Internet]. 2008 [cited 2018 Jan 14];27:2300–11. Available from:
643 http://www.ncbi.nlm.nih.gov/pubmed/18391972
644 43. Mangaraj M, Nanda R, Panda S. Apolipoprotein A-I: A Molecule of Diverse Function. Indian J.
645 Clin. Biochem. [Internet]. 2016 [cited 2017 Nov 25];31:253–9. Available from:
646 http://www.ncbi.nlm.nih.gov/pubmed/27382195
647 44. Fiaschetti G, Schroeder C, Castelletti D, Arcaro A, Westermann F, Baumgartner M, et al.
648 NOTCH ligands JAG1 and JAG2 as critical pro-survival factors in childhood medulloblastoma.
649 Acta Neuropathol. Commun. [Internet]. 2014 [cited 2017 Nov 25];2:39. Available from:
650 http://www.ncbi.nlm.nih.gov/pubmed/24708907
651 45. Reddy S, Devlin R, Menaa C, Nishimura R, Choi SJ, Dallas M, et al. Isolation and
652 characterization of a cDNA clone encoding a novel peptide (OSF) that enhances osteoclast
653 formation and bone resorption. J. Cell. Physiol. [Internet]. 1998 [cited 2017 Nov 25];177:636–45.
654 Available from: http://www.ncbi.nlm.nih.gov/pubmed/10092216
655 46. Wong C-H, Fung Y-WW, Ng EK-O, Lee SM-Y, Waye MM-Y, Tsui SK-W. LIM domain
656 protein FHL1B interacts with PP2A catalytic β subunit - A novel cell cycle regulatory pathway.
657 FEBS Lett. [Internet]. 2010 [cited 2017 Nov 25];584:4511–6. Available from:
658 http://www.ncbi.nlm.nih.gov/pubmed/20969868
659 47. Ng EL, Tang BL. Rab GTPases and their roles in brain neurons and glia. Brain Res. Rev.
660 [Internet]. 2008 [cited 2017 Nov 25];58:236–46. Available from:
661 http://www.ncbi.nlm.nih.gov/pubmed/18485483
662 48. Pandita E, Rajan S, Rahman S, Mullick R, Das S, Sau AK. Tetrameric assembly of hGBP1 is
663 crucial for both stimulated GMP formation and antiviral activity. Biochem. J. [Internet]. 2016 [cited
664 2017 Nov 25];473:1745–57. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27071416

665 49. Venkatesh B. Evolution and diversity of fish genomes. Curr. Opin. Genet. Dev. [Internet]. 2003
666 [cited 2016 Apr 26];13:588–92. Available from: http://www.ncbi.nlm.nih.gov/pubmed/14638319
667 50. Postlethwait J, Amores A, Cresko W, Singer A, Yan Y-L. Subfunction partitioning, the teleost
668 radiation and the annotation of the human genome. Trends Genet. [Internet]. 2004 [cited 2016 Apr
669 13];20:481–90. Available from:
670 http://www.sciencedirect.com/science/article/pii/S0168952504002136
671 51. Schranz ME, Mohammadin S, Edger PP. Ancient whole genome duplications, novelty and
672 diversification: the WGD Radiation Lag-Time Model. Curr. Opin. Plant Biol. [Internet]. 2012 [cited
673 2016 Apr 13];15:147–53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22480429
674 52. Robinson-Rechavi M, Marchand O, Escriva H, Laudet V. An ancestral whole-genome
675 duplication may not have been responsible for the abundance of duplicated fish genes. Curr. Biol.
676 [Internet]. 2001 [cited 2016 Apr 13];11:R458–9. Available from:
677 http://www.sciencedirect.com/science/article/pii/S0960982201002809
678 53. Kondrashov FA, Kondrashov AS. Role of selection in fixation of gene duplications. J. Theor.
679 Biol. [Internet]. 2006 [cited 2016 May 27];239:141–51. Available from:
680 http://www.ncbi.nlm.nih.gov/pubmed/16242725
681 54. Flower DR. The lipocalin protein family: structure and function. Biochem. J. [Internet].
682 Portland Press Ltd; 1996 [cited 2018 Feb 5];318 (Pt 1):1–14. Available from:
683 http://www.ncbi.nlm.nih.gov/pubmed/8761444
684 55. Flower DR. Multiple molecular recognition properties of the lipocalin protein family. J. Mol.
685 Recognit. [Internet]. John Wiley & Sons, Ltd.; 1995 [cited 2018 Feb 5];8:185–95. Available from:
686 http://doi.wiley.com/10.1002/jmr.300080304
687 56. Skerra A. Engineered protein scaffolds for molecular recognition. J. Mol. Recognit. [Internet].
688 John Wiley & Sons, Ltd.; 2000 [cited 2018 Feb 5];13:167–87. Available from:
689 http://doi.wiley.com/10.1002/1099-1352%28200007/08%2913%3A4%3C167%3A%3AAID-
32

690 JMR502%3E3.0.CO%3B2-9
691 57. Tao W, Sun L, Shi H, Cheng Y, Jiang D, Fu B, et al. Integrated analysis of miRNA and mRNA
692 expression profiles in tilapia gonads at an early stage of sex differentiation. BMC Genomics
693 [Internet]. 2016 [cited 2017 Mar 12];17:328. Available from:
694 http://www.ncbi.nlm.nih.gov/pubmed/27142172
695 58. Muschick M, Barluenga M, Salzburger W, Meyer A. Adaptive phenotypic plasticity in the
696 Midas cichlid fish pharyngeal jaw and its relevance in adaptive radiation. BMC Evol. Biol.
697 [Internet]. BioMed Central; 2011 [cited 2016 May 3];11:116. Available from:
698 http://bmcevolbiol.biomedcentral.com/articles/10.1186/1471-2148-11-116
699 59. Rogers RL, Shao L, Thornton KR. Tandem duplications lead to novel expression patterns
700 through exon shuffling in Drosophila yakuba. Begun DJ, editor. PLOS Genet. [Internet]. Public
701 Library of Science; 2017 [cited 2018 Feb 5];13:e1006795. Available from:
702 http://dx.plos.org/10.1371/journal.pgen.1006795
703 60. Green SA, Simoes-Costa M, Bronner ME. Evolution of vertebrates as viewed from the crest.
704 Nature [Internet]. 2015 [cited 2016 Jan 13];520:474–82. Available from:
705 http://www.ncbi.nlm.nih.gov/pubmed/25903629
706 61. Salzburger W. The interaction of sexually and naturally selected traits in the adaptive radiations
707 of cichlid fishes. Mol. Ecol. [Internet]. 2009 [cited 2016 Apr 12];18:169–85. Available from:
708 http://www.ncbi.nlm.nih.gov/pubmed/18992003
709 62. Barlow-Anacker AJ, Fu M, Erickson CS, Bertocchini F, Gosain A. Neural Crest Cells
710 Contribute an Astrocyte-like Glial Population to the Spleen. Sci. Rep. [Internet]. 2017 [cited 2018
711 Feb 14];7:45645. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28349968
712 63. Bailey AP, Bhattacharyya S, Bronner-Fraser M, Streit A. Lens Specification Is the Ground State
713 of All Sensory Placodes, from which FGF Promotes Olfactory Identity. Dev. Cell [Internet]. 2006
714 [cited 2018 Feb 14];11:505–17. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17011490

715 64. Kang C-K, Yang S-Y, Lin S-T, Lee T-H. The inner opercular membrane of the euryhaline
716 teleost: a useful surrogate model for comparisons of different characteristics of ionocytes between
717 seawater- and freshwater-acclimated medaka. Histochem. Cell Biol. [Internet]. 2015 [cited 2016
718 Nov 7];143:69–81. Available from: http://link.springer.com/10.1007/s00418-014-1266-2
719 65. Miyanishi H, Inokuchi M, Nobata S, Kaneko T. Past seawater experience enhances seawater
720 adaptability in medaka, Oryzias latipes. Zool. Lett. [Internet]. 2016 [cited 2016 Nov 7];2:12.
721 Available from: http://zoologicalletters.biomedcentral.com/articles/10.1186/s40851-016-0047-2
722 66. Muschick M, Indermaur A, Salzburger W. Convergent evolution within an adaptive radiation of
723 cichlid fishes. Curr. Biol. [Internet]. 2012 [cited 2016 Apr 13];22:2362–8. Available from:
724 http://www.sciencedirect.com/science/article/pii/S0960982212012699
725 67. Wellenreuther M, Svensson EI, Hansson B. Sexual selection and genetic colour polymorphisms
726 in animals. Mol. Ecol. [Internet]. 2014 [cited 2016 Apr 12];23:5398–414. Available from:
727 http://www.ncbi.nlm.nih.gov/pubmed/25251393
728 68. CARLETON KL, PARRY JWL, BOWMAKER JK, HUNT DM, SEEHAUSEN O. Colour
729 vision and speciation in Lake Victoria cichlids of the genus Pundamilia. Mol. Ecol. [Internet]. 2005
730 [cited 2018 Feb 5];14:4341–53. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16313597
731 69. Flamarique IN, Bergstrom C, Cheng CL, Reimchen TE. Role of the iridescent eye in
732 stickleback female mate choice. J. Exp. Biol. [Internet]. 2013 [cited 2018 Feb 5];216:2806–12.
733 Available from: http://www.ncbi.nlm.nih.gov/pubmed/23580716
734 70. Salzburger W, Braasch I, Meyer A. Adaptive sequence evolution in a color gene involved in the
735 formation of the characteristic egg-dummies of male haplochromine cichlid fishes. BMC Biol.
736 [Internet]. BioMed Central; 2007 [cited 2016 Apr 12];5:51. Available from:
737 http://bmcbiol.biomedcentral.com/articles/10.1186/1741-7007-5-51
738 71. Charlesworth D. Evolution of recombination rates between sex chromosomes. Philos. Trans. R.
739 Soc. B Biol. Sci. [Internet]. 2017 [cited 2018 Feb 5];372:20160456. Available from:

740 http://www.ncbi.nlm.nih.gov/pubmed/29109220
741 72. Flicek P, Amode MR, Barrell D, Beal K, Billis K, Brent S, et al. Ensembl 2014. Nucleic Acids
742 Res. [Internet]. 2014 [cited 2014 Jul 12];42:D749–55. Available from:
743 http://nar.oxfordjournals.org/content/42/D1/D749
744 73. Stanke M, Diekhans M, Baertsch R, Haussler D. Using native and syntenically mapped cDNA
745 alignments to improve de novo gene finding. Bioinformatics [Internet]. 2008 [cited 2018 Jan
746 14];24:637–44. Available from: https://academic.oup.com/bioinformatics/article-
747 lookup/doi/10.1093/bioinformatics/btn013
748 74. Swofford D. Phylogenetic Analysis Using Parsimony (*and Other Methods). Version 4. Sinauer
749 Assoc. Sunderland, Massachusetts. 2002;
750 75. Guindon S, Gascuel O. A simple, fast, and accurate algorithm to estimate large phylogenies by
751 maximum likelihood. Syst. Biol. [Internet]. 2003 [cited 2016 Apr 30];52:696–704. Available from:
752 http://www.ncbi.nlm.nih.gov/pubmed/14530136
753 76. Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and
754 parallel computing. Nat. Methods [Internet]. Nature Publishing Group, a division of Macmillan
755 Publishers Limited. All Rights Reserved.; 2012 [cited 2015 Apr 5];9:772. Available from:
756 http://dx.doi.org/10.1038/nmeth.2109
757 77. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version
758 7.0 for Bigger Datasets. Mol. Biol. Evol. [Internet]. 2016 [cited 2017 Mar 26];33:1870–4. Available
759 from: http://www.ncbi.nlm.nih.gov/pubmed/27004904
760 78. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. [Internet].
761 2007 [cited 2014 Jul 9];24:1586–91. Available from:
762 http://mbe.oxfordjournals.org/content/24/8/1586.abstract
763 79. Yang Z. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput.
764 Appl. Biosci. [Internet]. 1997 [cited 2016 Apr 16];13:555–6. Available from:

765 http://www.ncbi.nlm.nih.gov/pubmed/9367129
766 80. Letunic I, Doerks T, Bork P. SMART: recent updates, new developments and status in 2015.
767 Nucleic Acids Res. [Internet]. 2015 [cited 2018 Jan 14];43:D257–60. Available from:
768 http://www.ncbi.nlm.nih.gov/pubmed/25300481
769 81. Letunic I, Bork P. 20 years of the SMART protein domain annotation resource. Nucleic Acids
770 Res. [Internet]. Oxford University Press; 2018 [cited 2018 Jan 14];46:D493–6. Available from:
771 http://academic.oup.com/nar/article/46/D1/D493/4429069
772 82. Szklarczyk D, Morris JH, Cook H, Kuhn M, Wyder S, Simonovic M, et al. The STRING
773 database in 2017: quality-controlled protein–protein association networks, made broadly accessible.
774 Nucleic Acids Res. [Internet]. 2017 [cited 2018 Jan 14];45:D362–8. Available from:
775 http://www.ncbi.nlm.nih.gov/pubmed/27924014
776 83. Biasini M, Bienert S, Waterhouse A, Arnold K, Studer G, Schmidt T, et al. SWISS-MODEL:
777 modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids
778 Res. [Internet]. 2014 [cited 2017 Nov 3];42:W252–8. Available from:
779 http://www.ncbi.nlm.nih.gov/pubmed/24782522
780 84. Bordoli L, Kiefer F, Arnold K, Benkert P, Battey J, Schwede T. Protein structure homology
781 modeling using SWISS-MODEL workspace. Nat. Protoc. [Internet]. Nature Publishing Group;
782 2008 [cited 2017 Nov 3];4:1–13. Available from:
783 http://www.nature.com/doifinder/10.1038/nprot.2008.197
784 85. Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based
785 environment for protein structure homology modelling. Bioinformatics [Internet]. 2006 [cited 2017
786 Nov 3];22:195–201. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16301204
787 86. Guex N, Peitsch MC, Schwede T. Automated comparative protein structure modeling with
788 SWISS-MODEL and Swiss-PdbViewer: A historical perspective. Electrophoresis [Internet]. 2009
789 [cited 2017 Nov 3];30:S162–73. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19517507

790 87. R Developmental Core Team. R: a language and environment for statistical computing
791 [Internet]. Vienna, Austria. 2011. Available from: http://www.r-project.org/
792 88. McCurley AT, Callard G V. Characterization of housekeeping genes in zebrafish: male-female
793 differences and effects of tissue type, developmental stage and chemical treatment. BMC Mol. Biol.
794 [Internet]. BioMed Central; 2008 [cited 2016 Apr 12];9:102. Available from:
795 http://bmcmolbiol.biomedcentral.com/articles/10.1186/1471-2199-9-102
796 89. Hibbeler S, Scharsack JP, Becker S. Housekeeping genes for quantitative expression studies in
797 the three-spined stickleback Gasterosteus aculeatus. BMC Mol. Biol. [Internet]. BioMed Central;
798 2008 [cited 2016 May 3];9:18. Available from:
799 http://bmcmolbiol.biomedcentral.com/articles/10.1186/1471-2199-9-18
800 90. Zhang Z, Hu J. Development and validation of endogenous reference genes for expression
801 profiling of medaka (Oryzias latipes) exposed to endocrine disrupting chemicals by quantitative
802 real-time RT-PCR. Toxicol. Sci. [Internet]. 2007 [cited 2016 Oct 27];95:356–68. Available from:
803 http://www.ncbi.nlm.nih.gov/pubmed/17093204
804 91. Soderlund C, Bomhoff M, Nelson W. SyMAP v3.4: a turnkey synteny system with application
805 to plant genomes. Nucleic Acids Res. 2011;39:e68.
806 92. Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, et al. The genomic substrate for
807 adaptive radiation in African cichlid fish. Nature [Internet]. Nature Publishing Group, a division of
808 Macmillan Publishers Limited. All Rights Reserved.; 2014 [cited 2016 May 5];513:375–81.
809 Available from: http://dx.doi.org/10.1038/nature13726
810 93. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant
811 discovery by integrated paired-end and split-read analysis. Bioinformatics [Internet]. 2012 [cited
812 2018 Jan 14];28:i333–9. Available from: https://academic.oup.com/bioinformatics/article-
813 lookup/doi/10.1093/bioinformatics/bts378
814 94. Tao W, Sun L, Shi H, Cheng Y, Jiang D, Fu B, et al. Integrated analysis of miRNA and mRNA

815 expression profiles in tilapia gonads at an early stage of sex differentiation. BMC Genomics
816 [Internet]. 2016 [cited 2018 Feb 5];17:328. Available from:
817 http://bmcgenomics.biomedcentral.com/articles/10.1186/s12864-016-2636-z
818