



Newly Discovered *AqE* Gene is Highly Conserved in Non-tetrapod Vertebrates

Lyudmila V. Puzakova¹ · Mikhail V. Puzakov¹ · Olga L. Gostyukhina¹

Received: 20 July 2020 / Accepted: 21 January 2021 / Published online: 19 February 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC part of Springer Nature 2021

Abstract

Studying the diversity of energy production pathways is important for understanding the evolutionary relationships between metabolic pathways and their biochemical precursors. The lactate/malate dehydrogenase (LDH/MDH) superfamily has been a model system for structural and functional evolution for a long time. Recently, the type-2 family of LDH/MDH (or LDH2/MDH2 oxidoreductase) has been identified. The LDH2/MDH2 oxidoreductase family is now known to have functionally more diverse enzymes than the LDH/MDH superfamily. In channel catfish, the gene encoding the LDH2/MDH2 oxidoreductase has been found (and was provisionally termed *AqE*). Homologs of this enzyme are predominantly present in organisms living in an aquatic environment. In this work, we studied the *AqE* gene distribution among non-tetrapod vertebrates. It was found that the *AqE* gene is present in the genomes of bony and cartilaginous fish and in the genomes of hagfishes and lampreys. In addition, it has been confirmed that in representatives of Cypriniformes, the *AqE* gene has been lost. *AqE* in representatives of Salmoniformes underwent significant deletions, which most likely led to its pseudogenization. In most orders of non-Tetrapoda vertebrates, the *AqE* gene remains highly conserved, suggesting that the *AqE* gene in aquatic vertebrates is an essential gene and undergoes rigorous selection. The *AqE* gene has the highest sequence similarity with the archaeal *ComC* gene that encodes sulfolactate dehydrogenase (SLDH). Based on the similarity of substrates, the enzyme encoded by the *AqE* gene is likely involved in the malate–aspartate shuttle mechanism or the biosynthesis of the energy coenzyme M equivalent.

Keywords Gene evolution · Gene loss · Vertebrates · LDH2/MDH2 oxidoreductase · L-sulfolactate dehydrogenase

Abbreviations

bp	Base pair
aa	Amino acids
ORF	Open reading frame
WGS	Whole-genome shotgun sequences
TSA	Transcriptome shotgun assembly
SLDH	Sulfolactate dehydrogenase

Introduction

The lactate/malate dehydrogenase (LDH/MDH) superfamily has been a model system for structural and functional evolution for a long time (Madern 2002). These proteins catalyze

the reduction of 2-ketocarboxylic acids using the NAD(P)H cofactor. LDH provides a reversible conversion of lactate to pyruvate. MDH catalyzes interconversions between malate and oxaloacetate. LDH/MDH oxidoreductases have been characterized and sequenced in a wide variety of organisms representing the bacterial, archaeal, and eukaryotic domains of life. Based on structural and biochemical data, it has been found that LDH and MDH originate from a common ancestor (Madern 2002).

A new class of NAD(P)H-dependent MDH and LDH was later identified. The sequences of these enzymes do not have homology with classic MDH/LDH oxidoreductases (Honka et al. 1990; Jendrossek et al. 1993). These enzymes and their homologues have been annotated as type-2 LDH/MDH or LDH2/MDH2 oxidoreductases. Since the sequences of LDH2/MDH2 oxidoreductases were placed in databases, many proteins were subsequently annotated as type-2 LDH/MDH or LDH/MDH. However, subsequent biochemical and genetic studies have shown that putative proteins, showing significant sequence similarity with LDH2/MDH2 oxidoreductases, have catalytic activity other than MDH or LDH.

Handling editor: Kerry Geiler-Samerotte.

✉ Lyudmila V. Puzakova
kvluda@yandex.ru

¹ A.O. Kovalevsky Institute of Biology of the Southern Seas of RAS, Nakhimov av., 2, Sevastopol 299011, Russia

The LDH2/MDH2 oxidoreductase family is currently known to include functionally more diverse enzymes that are divided into eight clades (Muramatsu et al. 2005). They include L-sulfolactate dehydrogenase, Δ 1-piperidine-2-carboxylate/ Δ 1-pyrroline-2-carboxylate reductase, 2,3-diketo-L-gulonate reductase, ureidoglycolate dehydrogenase, and others. Functional studies of this group of enzymes have been carried out only in bacteria and archaea (Irimia et al. 2004; Denger and Cook 2010; Zhang et al. 2017).

In silico studies have identified NAD(P)H-dependent LDH2/MDH2 oxidoreductase in the channel catfish *Ictalurus punctatus* genome. This protein (NP_001188075) has been erroneously annotated as MDH. Further analysis showed that homologues of LDH2/MDH2 oxidoreductases of channel catfish are predominantly present in organisms that live in aquatic environments (Puzakova et al. 2019). In this regard, the gene encoding this LDH2/MDH2 oxidoreductase has been provisionally termed *AqE* (from aquatic enzyme). It has been suggested that this gene is involved in anaerobic respiration. Hence, gene loss in terrestrial animals and plants can be associated with transition to land and a related restructuring of metabolic pathways owing to environmental oxygen saturation and the absence of the natural hypoxia characteristic of aquatic organisms (Puzakova et al. 2019).

As of now, the primary sequence of genomic DNA has been determined in more than 30 thousand species. In previous study, we operated on large taxa to cover all groups of organisms when searching for the *AqE* gene (Puzakova et al. 2019). As a result, the distribution of *AqE* gene could be observed only in such taxa as phylum or class. The main goal of the previous work was to show that this gene encodes a protein that is fundamentally different from the cytoplasmic MDH. In addition, the fact of the presence of *AqE* in one or another large taxon was investigated without analyzing the exon–intron structure. In addition, it was shown that it is absent in the predominant number of terrestrial (atmospheric) organisms. In animals, the loss of the *AqE* gene was observed in four classes: amphibians, reptiles, birds, and mammals, belonging to the Vertebrata subtype. In this study, the distribution and structural features of the *AqE* gene were investigated in detail in non-tetrapod vertebrates with taxonomic resolution up to the order. It has been shown that the gene is retained in most of the studied species and is under the influence to purifying selection. It was also found that *AqE* is a homologue of the archaeal *ComC* gene.

Results

AqE Gene Landscape in Genomes of Non-tetrapod Vertebrates

In the study of the *AqE* gene distribution among vertebrates, 86 orders of vertebrates not belonging to the Tetrapoda group were analyzed. At the time of the study, the NCBI database contained whole-genome shotgun sequences (WGSs) only for representatives of 58 orders out of 86 analyzed. According to the analysis of non-redundant protein sequences and WGS databases using BLASTp and tBLASTn, the genomes of 118 species were analyzed (Additional file 1, Fig. 1).

As a reference protein (query), we used the amino acid sequence encoded by the *AqE* gene of the channel catfish, *Ictalurus punctatus*. The protein of this organism was chosen because it was the first used to study this enzyme (Puzakova et al. 2019). In addition, the channel catfish *AqE* has an exon–intron structure that has been derived by automated computational analysis using gene prediction method BestRefSeq relies on transcript data. (GeneID: 100528876, NCBI). This made it possible to more accurately determine the boundaries of the *AqE* gene exons in species that lacked transcriptomic data.

Overall, nucleotide sequences homologous to *AqE* were revealed in 101 species (56 orders). In 16 organisms belonging to two orders (Rajiformes and Cypriniformes), the gene was not detected. In another species, *Hucho hucho* (Salmoniformes), the detected coding sequence is a bacterial gene, which apparently contaminated the samples during sequencing. This was confirmed by the intron absence of the *H. hucho* gene and the high sequence similarity of the *H. hucho* protein to bacteria *Vogesella perlucida AqE* (80.24%), whereas there was a low identity of the *H. hucho* protein to the *AqE* of *I. punctatus* (26.97%).

In the three species of *Tachysurus fulvidraco*, *Amphiprion ocellaris*, and *Acipenser ruthenus* (orders Siluriformes, Pomacentridae, and Acipenseriformes, respectively), the *AqE* gene was found in two copies. This type of duplication of the gene is the exception rather than the trend because in the vast majority of the studied organisms, there is only one copy of the gene. In detail, in superclass ray-finned fishes (Actinopterygii), sequences homologous to *AqE* were detected in members of 50 orders. In members of 19 orders of superclass Actinopterygii, WGSs were absent; thus, there are no data regarding the presence of *AqE* in organisms of these taxa. Only in representatives of one order (Cypriniformes) no homologies to the *AqE* gene were found. Since the WGSs of 15 species of Cypriniformes were analyzed, this result can be considered reliable, and accordingly, the *AqE* gene was lost in this group (Additional file 1, Fig. 1).

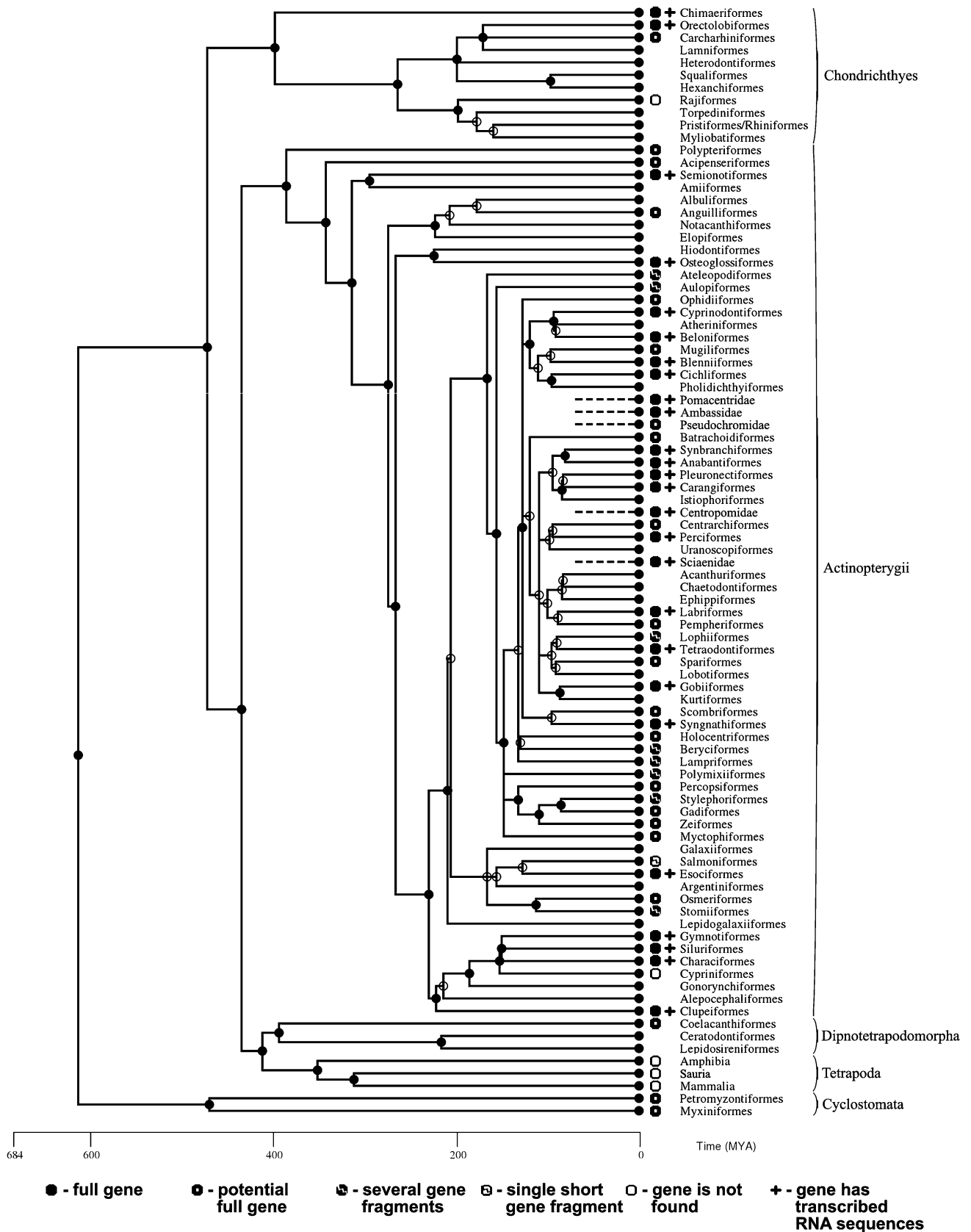


Fig. 1 The distribution of *AqE* gene among vertebrates. The phylogenetic tree from <http://www.timetree.org/> was taken as a basis

In the cartilaginous fishes (Chondrichthyes) class, WGSs were available only for representatives of four groups out of eleven Chimaeriformes (subclass Holocephali, chimeras), Orectolobiformes and Carcharhiniformes (infraclass Selachii, sharks), and Rajiformes (superorder Batoidea, rays).

In the studied representatives of chimeras and sharks, the *AqE* gene was detected (Additional file 1, Fig. 1). In *Raja erinacia* (rays), no homologies to the *AqE* gene were found. Since this is the only species with WGSs from superorder Batoidea, the gene absence in *R. erinacia* can be considered both as evidence of *AqE* loss in all skates (like in tetrapods) and as a species-specific exception. It is possible that the *AqE* gene was not detected due to the low assembly level (contig) and low genome coverage (26×). In the no rank group of Cyclostomata, the *AqE* gene was detected in representatives of both orders of jawless vertebrates, Petromyzontiformes (lampreys) and Myxiniiformes (hagfishes).

Structure Variations of *AqE* Genes

To better understand the evolution of the *AqE* gene in vertebrates, the exon–intron structure was analyzed. It was first defined for 50 species. The *AqE* gene of a predominant number of non-tetrapod vertebrates is quite conserved and on average consists of 11 exons. Due to the different lengths of the introns, the *AqE* lengths varies from 3 to 134 kbp (Additional file 2). In organisms of 26 orders (Fig. 1), the *AqE* gene encoding a full-sized protein *AqE* was detected. In representatives of another 21 orders, the gene is characterized as hypothetically whole. These species do not have a transcriptome assembly (TSA); therefore, their exon–intron structure was determined exceptionally by homology with the amino acid sequence of *I. punctatus*. This approach did not allow the detection of exon 1 because the coding sequence length in this exon is extremely small (on average 3 aa). For the same reason, we could not detect exon 11 (the last) owing to its significant variability in some species as mentioned above. Nevertheless, these organisms likely have the full-sized *AqE* gene. Thus, representatives of 47 orders have a full-sized *AqE* gene with some species-specific variations. For example, in *Lamprogrammus exutus* (Ophidiiformes), exon 3 duplication was detected. In *Xiphophorus maculatus* (Cyprinodontiformes), only part of exon 2 was preserved, but exons 7 and 8 were absent. In *Cyprinodon variegatus* (Cyprinodontiformes), a fusion of exons 3 and 4 and a fusion of exons 5 and 6 were revealed. In *Xiphophorus couchianus* (Cyprinodontiformes), exons 2 and 6 were only partially present, and exons 8 and 9 were fused. However, in the other seven representatives of Cyprinodontiformes, the gene was found to be similar to the “classical” one. In representatives of eight orders (Ateleopodiformes, Aulopiformes, Lophiiformes, Beryciformes, Lampriformes,

Polymixiiformes, Stylephoriformes, and Stomiiformes), several fragments homologous to the query amino acid sequence were found in the genomic sequences (Additional file 1). They were mainly localized on short scaffolds from 2033 to 18,322 bp. The assemblies of species from these orders had a low N50 values (< 20,000 bp), a rather low genome coverage (< 25×) and a low assembly level (> 50,000 number of scaffolds) (Additional file 3). This could be the reason that some fragments of the *AqE* gene were absent in the genomic sequences of representatives of these orders. Therefore, we cannot consider these data as evidence of deletions or pseudogenization, and we believe that the *AqE* gene in these orders is full-sized.

Only short fragments of the *AqE* gene were found in the order Salmoniformes. However, in contrast to the previous case, almost all assemblies had a genome coverage of more than 100× and an assembly level up to the chromosome (Additional file 3). Seven species were examined, and in five of these, the *AqE* gene was represented only by exons 5 and 6; one species had only fragments of exons 2, 3, and 6, and one had *AqE*, which was most likely a bacterial gene that contaminated the samples (Additional file 1).

Accordingly, in this order, the *AqE* gene is pseudogenized (deleted) and, most likely, not functional. Analysis of the transcriptome databases (TSA) showed that the *AqE* gene was active in all organisms that had this gene and transcriptome assemblies (Fig. 1). The presence of alternative transcripts was found for a number of organisms (Additional file 1). No dependence between the presence of alternative transcripts and the taxonomic position of organisms was found.

Phylogenetic Relationships Among the *AqE* Genes in Vertebrates

To better understand the evolutionary relationships between *AqE* genes in vertebrates, a phylogenetic analysis was performed based on the maximum likelihood (ML) method in which 97 identified *AqE* proteins were included.

Amino acid sequences with a length of less than 50% of the length of query sequences (orders Aulopiformes and Salmoniformes) were excluded from the analysis. In general, the distribution on clades correlated to the taxonomic division. Three orders, Beryciformes, Pleuronectiformes, and Blenniiformes, were exceptions, whose representatives were distributed on different branches. Nevertheless, taking into consideration the low bootstrap values within the a clade (Additional file 4) and the associated polytomy, the overall picture of the phylogenetic relationships was determined to still be classical. The dendrogram also shows that the increase in copies in individual species is associated exclusively with intraspecific duplications.

Sequence Modification of AQE Proteins

To identify conserved motifs (CM) and define their location in the proteins encoded by the studied *AqE* genes, the MEME online server was used. This analysis included 113 amino acid sequences: full-sized, potentially full-sized, deleted in less than 50% of the total enzyme length, and isoforms resulting from alternative transcription. Fifteen supposed conserved motifs were identified (Additional file 5), the length of which was from 10 to 50 amino acids. Ten conserved motifs were frequently found (94–111 proteins), and the other five were rare (3–14 proteins). The number of conserved motifs in each AQE protein ranged from five to ten. The most AQE proteins had ten conserved motifs, which represent over 90% of the entire amino acid sequence.

A detailed analysis of conserved motifs showed that CM-15 is an alternative (modification) of CM-10 in representatives of Chondrichthyes. CM-12 was found in deleted proteins and isoforms produced from transcripts with an alternative start of translation in exon 2, which is actually a fragment of CM-2. CM-13 and CM-12 are shortened versions of CM-2, resulting from alternative transcription. CM-14 is a deleted version of CM-4. CM-11 in six cases (*Xiphophorus couchianus*, *Xiphophorus maculatus*, *Acanthochaenus luetkenii*, *Polymixia japonica*, *Borostomias antarcticus*, and *Rondeletia loricata*) is located at the site of CM-5 (exons 5 and 6), and in three cases (the chimeric protein of *Acipenser ruthenus* and two isoforms of *Paramormyrops kingsleyae*), it is located after CM-10 (the end of exon 11). Thus, only 11 conserved motifs can be considered unique (CM-1 to CM-11). Therefore, *AqE* genes demonstrate rather high conservatism during the evolution of vertebrates.

Discussion

Possible Reasons for the Variation of *AqE* Distribution in Non-Tetrapoda Vertebrates

The distribution and structure of the *AqE* gene in vertebrates not belonging to the Tetrapoda group (with taxonomic resolution up to the order) were analyzed in this study for the first time. It was found that representatives of most (54 of 56) of the studied non-tetrapod vertebrate orders have the *AqE* gene (Additional file 1, Fig. 1). In three species (*Tachysurus fulvidraco*, *Amphiprion ocellaris*, and *Acipenser ruthenus*), duplication of the *AqE* gene was revealed, which was most likely associated with exclusive intraspecific duplications. The presence of *AqE* duplicates in certain teleost species is not surprising since a gene duplication arise at a very high rate (on average 0.01 per gene per million years) (Lynch and Conery 2000).

No homologies to the *AqE* gene were found in genomes of two orders (Rajiformes or Cypriniformes) of representatives. Genome sequences of only one representative were available for analysis of the order Rajiformes. Therefore, we could not confirm that the *AqE* gene absence is a characteristic feature of this order but not a species-specific phenomenon or the result of insufficient high-quality sequencing or assembly. In the order of Cypriniformes, the gene is absent in all 15 species analyzed; thus, the *AqE* gene loss in this taxon is out of doubt. In species of the order of Salmoniformes, the *AqE* gene underwent substantial deleting. This was confirmed by the study of seven species of this taxon. An additional search for the *AqE* gene in the mitochondrial genomes of Cypriniformes and Salmoniformes gave a negative result. This ruled out the assumption of the possible transfer of the gene from the nucleus to mitochondria. Thus, the fact that organisms of Cypriniformes and Salmoniformes lost the *AqE* gene (entirely or partially) is reliable.

This result is quite unexpected because we initially considered the hypothesis in which all aquatic organisms have the *AqE* gene (Puzakova et al. 2019). In representatives of the remaining orders of non-tetrapod vertebrates, *AqE* is preserved. Moreover, it is conserved and transcribed. These data confirm our assumption that the *AqE* gene is required for aquatic organisms. In addition, this hypothesis is supported by the results of the codon-based Z-test of selection (Additional file 6). This analysis is based on comparing the numbers of synonymous and nonsynonymous substitutions per site (dS and dN, respectively). The Z-test shows that *AqE* is under the influence of purifying selection (dN < dS).

In this case, why did Cypriniformes and Salmoniformes lose the *AqE* gene, which is so necessary for other taxa? Gene loss is known to be a rather common phenomenon in evolution. Gene loss can have a neutral effect on vital activity (Moreau and Dabrowski 1998; Drouin et al. 2011) or significantly increase the adaptive potential of a species (Greenberg et al. 2003; Clark et al. 2007; McBride et al. 2007; Goldman-Huertas et al. 2015). Otherwise, the loss of the essential gene will be lethal and will not be fixed in the population. There are different scenarios for evolutionary gene inactivation and/or loss. It can be a slow accumulation of mutational changes in the gene with its transformation to the pseudogene and further gradual degradation (fragmentation). Another part of the process is sudden, and complete gene loss (deletion) can occur due to unequal crossing over during meiosis or mobile genetic element transposition. Gene pseudogenization occurs when the gene becomes redundant. When gene loss is sudden, an organism can survive only if the gene has already lost its significance for the organism or if there are analogues that can take over the function of the lost gene. We cannot exclude the idea that in Cypriniformes, in which we did not find even gene residues, a deletion could occur. It is

also possible that the gene could have been pseudogenized; however, the process resulted in such a considerable degradation that homologies could not be found. In Salmoniformes, only two exons were preserved. Such gene degradation is characteristic of pseudogenization.

The loss of the *AqE* gene in these bony fish orders may have occurred because of individual evolution of these taxa and restructuring of metabolic pathways. For example, alternative pathways could be formed to work with AQE substrates. It is also possible that other enzymes have taken over the function of the AQE enzyme. Examples of non-homologous gene replacement are known. For example, SLDH can utilize oxaloacetate as a substrate with relatively high efficiency.

This suggests that SLDH of methanogenic archaea may act as an analogue of MDH to compensate for the lack of a specific LDH-like MDH (Irimia et al. 2004) to act as analogous MalDHs to compensate for the lack of a specific orthologous (LDH-like) MalDH. In any case, this enzyme lost its significance.

It is known that an event of whole-genome duplication (WGD) occurred independently in Cypriniformes and Salmoniformes (the fourth vertebrate WGD) (Berthelot et al. 2014; Petit et al. 2017). An excess of oxidoreductases resulting from genome duplication may have become a reason for the “painless” *AqE* gene loss because a greater “stock” for non-homologous replacement or the formation of new pathways and new enzymes occurred. Table 1 presents data on the amount of some enzymes of energy metabolism in the genomes of organisms from the Otomorpha clade, including the order Cypriniformes, an unnamed clade including the order Salmoniformes and the order Perciformes (as an external group). The orders Cypriniformes and Salmoniformes obviously have a greater number of homologues to the indicated enzymes. These data confirm our assumption about the excess of enzymes in these orders. It is also known that polyploidy in Salmoniformes resulted in the presence of at least 30 aldehyde dehydrogenase genes, which is more than in other higher vertebrates (Holmes 2019). A trend toward an increase in the number of genes has been observed for

many enzymes, including lactate dehydrogenase, creatine kinase, and glucose-6-phosphate isomerase (Ferris 1984).

From the evolutionary processes following duplication, new “advantageous” allele combinations for the genome or completely new alleles could appear (Lynch and Conery 2000; Walsh 2003). According to S. Copley (Copley et al. 2017), on average, an enzyme can have ten different activities, any of which can be a starting point for the evolution of a new enzyme. Gene duplication is supposed to promote the formation of completely different enzymes. Thus, a wide variety of dehydrogenases may have formed (Eventoff and Rossmann 1975).

The fact that *AqE* is present in the overwhelming proportion of the studied genomes as a singleton is also surprising, since most of these organisms share a fairly recent whole-genome duplication (the third vertebrate WGD). Consequently, the loss of one *AqE* copy following the teleost WGD occurred very early. However, as can be seen from Table 1, the gene encoding mitochondrial malate dehydrogenase (*mdh2*) is also represented by one copy in many species (especially in non-Cypriniformes and non-Salmoniformes). The loss of duplicates of the *AqE* and *mdh2* genes in teleost fish can be explained by possible features of metabolic pathways in which an excess of the gene dose leads to fatal consequences. Unfortunately, without knowing the function of a gene, it is difficult to give a more detailed explanation.

Putative Function of the *AqE* Gene

In our previous study, we determined that the *AqE* gene encoded the LDH2/MDH2 oxidoreductase. This enzyme has the greatest sequence and structure similarity with sulfolactate dehydrogenase S-SLDH [EC 1.1.1.337] and MDH/SLDH [EC 1.1.1.310] and to MDH(1V9N) and MDH(3I0P) whose functions have not been determined experimentally (Puzakova et al. 2019). Sulfolactate is a natural compound in many organisms (bacteria, archaea, plants, and animals) (Denger and Cook 2010). However, there is some diversity in the metabolic pathways that sulfolactate is involved in. At least three different sulfolactate dehydrogenase (SLDH) have

Table 1 Representation of copies of genes encoding some enzymes of energy metabolism in representatives of certain orders

Orders	<i>AqE</i>	<i>mdh1</i> -like	<i>mdh2</i> -like	<i>ldh</i> -like	<i>me</i> -like	<i>aat</i> -like
Cypriniformes	0.00(0.00)	3.47(1.46)	2.07(1.10)	6.33(2.50)	5.73(2.79)	5.40(2.20)
Gymnotiformes, Characiformes, Siluriformes, Clupeiformes	1.14(0.38)	2.00(0.00)	1.00(0.00)	3.14(1.21)	3.71(0.76)	2.86(0.69)
Salmoniformes	0.86(0.38)	3.71(0.76)	2.29(0.49)	6.71(1.98)	5.43(0.79)	6.14(1.35)
Esociformes, Osmeriformes, Stomiiformes	1.00(0.00)	2.00(0.00)	1.00(0.00)	3.33(0.58)	2.67(0.58)	2.33(0.58)
Perciformes	1.00(0.00)	2.50(1.00)	1.50(1.00)	3.75(1.71)	3.00(0.00)	4.00(0.00)

AqE aquatic enzyme, *mdh1* malate dehydrogenase, cytoplasmic, *mdh2* malate dehydrogenase, mitochondrial; *ldh*—L-lactate dehydrogenase; *me*—malic enzyme; *aat*—aspartate aminotransferase. The cells indicate the average of copies of each gene. The standard deviation is shown in parentheses. Full details are provided in Additional file 7

been described. The most studied enzyme [EC 1.1.1.272] is SLDH, which is encoded by the *ComC* gene. This enzyme is involved in coenzyme M biosynthesis in methanogenic archaea and spore-formers. ComC oxidoreductase interconverts R-sulfolactate to sulfopyruvate. In subsequent biochemical transformations, sulfopyruvate becomes coenzyme M (Graupner et al. 2000; Muramatsu et al. 2005; Zhang et al. 2017). Using sulfolactate sulfo-lyase (*SuyAB*), R-sulfolactate can be transformed into pyruvate, which is “a crossing point” of many metabolic pathways. Another studied sulfolactate dehydrogenase [EC 1.1.1.–], encoded by the *SlcC* gene, converts S-sulfolactate to sulfopyruvate, which in turn can be used by *ComC*. Thus, *ComC* and *SlcC* perform sulfolactate racemization in a pair (Denger and Cook 2010). The third, least studied enzyme [EC 1.1.99.–], which is encoded by the *SlcD* gene, is bound to the membrane and is involved in the degradation of sulfolactate in the bacteria, *Roseovarius nubinihibens* (Denger et al. 2009). The phylogenetic analysis, which involved the enzymes of the LDH/MDH and LDH2/MDH2 families, revealed that the LDH2/MDH2 oxidoreductase (which was encoded by the *AqE* gene) had the highest sequence similarity with the archaeal *ComC* clade members (Fig. 2, Additional file 8). The enzymes encoded by bacterial *ComC* form a separate clade. The enzymes encoded by the *SlcC* gene may not even be members of the LDH2/MDH2 oxidoreductase family because it formed its own clade. The archaeal *ComC* clade includes L-sulfolactate dehydrogenases found in methanogenic archaea. Although these enzymes can also utilize malate and α -ketoglutarate as substrates, their classification is based on the preferential use of sulfolactate. In methanogenic archaea and in spore-formers, this enzyme is involved in the biosynthesis of coenzyme M (methanogenic cofactor) (Muramatsu et al. 2005; Denger and Cook et al. 2010). In other organisms, this enzyme has not been described; thus, the function of the *AqE* gene remains unknown, in particular, in eukaryotes. As Irimia et al. (2004) proposed, converting sulfolactate to sulfopyruvate in eukaryotes does not make sense because there are no corresponding metabolic processes. Therefore, the main substrates for SLDH encoded by the *AqE* gene in non-Tetrapoda vertebrates are likely to be malate and/or α -ketoglutarate or even another compound. However, it cannot be excluded that SLDH in eukaryotes is still involved in the conversion of sulfolactate to sulfopyruvate with the formation of certain energy equivalents of coenzyme M.

The most important function of the oxidoreductases group, which *AqE* belongs to, is associated with the ecological and biochemical role in adaptive reactions, which is usually expressed in the regulation of the balance of aerobic and anaerobic processes. SLDH in aquatic vertebrates is likely to be a reserve pathway enzyme, which supplements the main metabolic energy processes under conditions of oxygen deficiency. This assumption has been confirmed by

two facts. First, the *AqE* gene disappears in terrestrial vertebrates (we associate this with the presence of free oxygen in the atmosphere). Second, the *AqE* protein is the most similar to SLDH, which is involved in anaerobic processes (Graupner et al. 2000; Muramatsu et al. 2005; Zhang et al. 2017). The malate–aspartate shuttle mechanism, in which malate and α -ketoglutarate are the key compounds, is considered the most effective process that allows aquatic organisms to survive under hypoxia (anoxia) conditions (Hochachka and Somero 2002). Malate and α -ketoglutarate are transferred to the mitochondria through the antiport, where they are oxidized to oxaloacetate by the mitochondrial enzyme, MDH2. Since the SLDH protein encoded by the *ComC* gene can use malate and α -ketoglutarate as a substrate in addition to sulfolactate (Muramatsu et al. 2005), we also suggest that the product of the *AqE* gene can be included in the malate–aspartate shuttle mechanism (Fig. 3). The key enzyme in this process is the cytoplasmic fraction of malate dehydrogenase (MDH1, 1.1.1.37). However, some other enzymes from the malate dehydrogenase family also take part in the combination of protein and carbohydrate metabolism (Hochachka and Somero 2002). We do not exclude that this could be an enzyme encoded by the *AqE* gene.

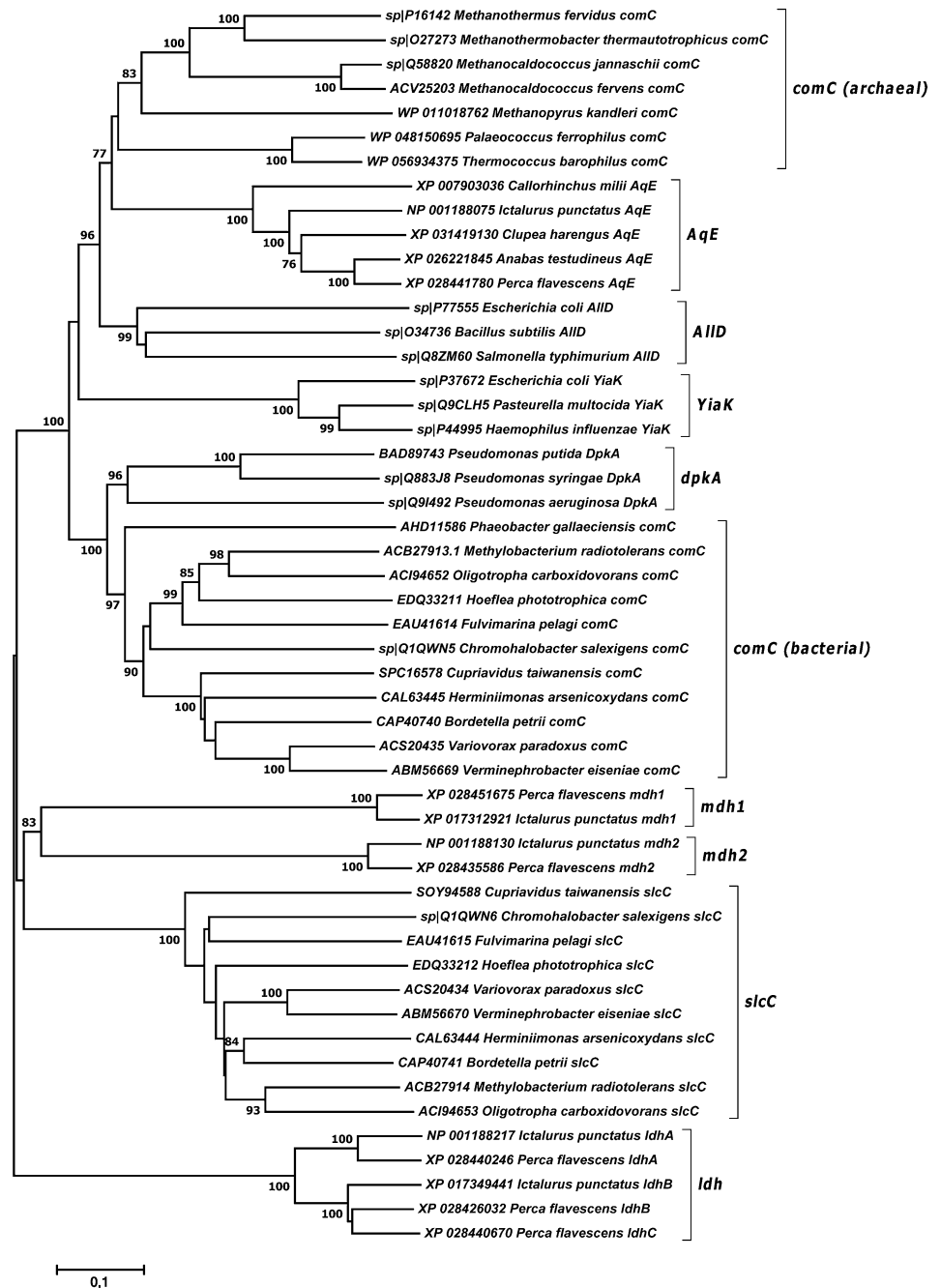
Conclusion

Our study of the distribution of the *AqE* gene among non-Tetrapoda vertebrates showed that it is present in the genomes of bony and cartilaginous fishes and in the genomes of hagfishes and lampreys. In addition, it was reliably shown that, for representatives of Cypriniformes, the *AqE* gene was lost, and for representatives of Salmoniformes, it underwent significant deletions, which most likely led to its pseudogenization. Thus, in most orders of non-Tetrapoda vertebrates, the *AqE* gene remains highly conserved. This suggests that the *AqE* gene in aquatic vertebrates is an essential gene and undergoes rigorous selection. Therefore, the enzyme is actively involved in metabolic pathways that are still unknown.

The *AqE* gene has the highest sequence similarity with the archaeal *ComC* that encodes SLDH. Based on the similarity of substrates, it cannot be excluded that the enzyme encoded by the *AqE* gene is involved in the following metabolic pathways:

- the malate-aspartate shuttle mechanism, which is the most effective process in aquatic organisms living under hypoxia (anoxia) conditions. This mechanism combines protein and carbohydrate metabolism and provides organisms with energy in the form of NADH;
- the pathway of sulfolactate to sulfopyruvate conversion followed by the formation of energy equivalents in the form

Fig. 2 Evolutionary relationships of LDH/MDH and LDH2/MDH2 family enzymes and AqE proteins. The evolutionary history was inferred using the Neighbor-Joining method, bootstrap 1000. The analysis involved 51 amino acid sequences. Only bootstrapping values higher than 70% are written on the branch



of coenzyme M (an analogue of the pathway found in methanogenic archaea).

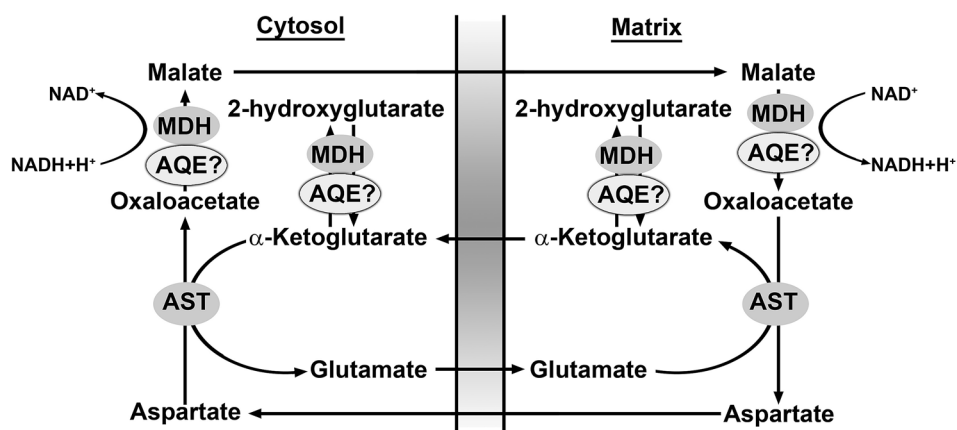
Methods

Mining AqE Genes

The amino acid sequence of the *Ictalurus punctatus* AqE gene (GeneID: 100528876) was used as a query to identify homologous genes in the whole-genome shotgun sequences

(WGS) of non-tetrapod vertebrates (Additional file 3). A search was carried out using the Basic Local Alignment Search Tool (BLAST) (Altschul et al. 1997). First, we searched for homologies among non-redundant protein sequences of non-tetrapod vertebrates to find the species (orders) in which the structure of the AqE gene was determined from automatic annotation. Next, we investigated the genome sequences of representatives of orders that did not have a predicted gene structure. Homologous sequences for the amino acid sequence of the *I. punctatus* AqE were searched using tBLASTn. Exon boundaries were refined

Fig. 3 Possible AQE involvement in the malate–aspartate shuttle. MDH, malate dehydrogenase; AST—aspartate transaminase



visually by the highest sequence similarity between the query and studied sequence and the presence of the 5' and 3' splice site boundary. If homologous sequences to all exons of *AqE* were not found in a representative of a certain order, then, all members of this order were analyzed. The *AqE* coding sequences obtained from the analysis were used to search transcribed RNA sequences in the transcriptome shotgun assembly database (TSA).

TBLASTn was used to study the representation of copies of genes encoding some enzymes of energy metabolism in fish genomes. The following amino acid sequences were used as a query: *Larimichthys crocea* malate dehydrogenase, cytoplasmic isoform X1 (XP_019115288); *L. crocea* malate dehydrogenase, mitochondrial (XP_010738316); *L. crocea* L-lactate dehydrogenase A chain (XP_010731034); *L. crocea* L-lactate dehydrogenase B chain isoform X1 (XP_027128106); *L. crocea* L-lactate dehydrogenase C chain isoform X1 (XP_010755609); *L. crocea* NAD-dependent malic enzyme, mitochondrial (XP_027133890); *L. crocea* aspartate aminotransferase, cytoplasmic isoform X1 (XP_019110932). The maximum number of copies of an identical fragment with E value < 0.001 was counted. In addition, the number of copies with query cover > 50% was separately counted.

Gene Structure and Conserved Motif Analysis of *AqE* Genes

The exon–intron structures of the *AqE* genes were displayed using Gene Structure Display Server 2.0 (Hu et al. 2015) based on the alignment of their coding sequences with their corresponding genomic sequences. The MEME suite server (Bailey et al. 2009) was used to identify the conserved motifs of the proteins encoded by the *AqE* genes, and the parameters used for this study were as follows: maximum number of different motifs, 20; minimum width, 10; and maximum width, 50.

Phylogenetic Analyses

Multiple alignments of the amino acid sequences were performed using MUSCLE (Edgar 2004), and the resulting data were used to construct a phylogenetic tree using MEGA 7 software (Kumar et al. 2016) with the maximum likelihood (ML) method. Regarding the ML method, the following parameters were used: bootstrap, 100 replicates; Jones–Taylor–Thornton; and Gamma distribution. The analysis involved 97 *AQE* amino acid sequences of non-tetrapod vertebrates (Additional file 9). The evolutionary relationships of LDH/MDH and LDH2/MDH2 family enzymes and *AQE* proteins were inferred using the Neighbor-Joining (NJ) method (bootstrap 1000) and ML method (bootstrap 100; Le_Gascuel_2008 model; and Gamma distribution). The analysis involved 51 amino acid sequences (Additional file 10). Only five proteins encoded by the fish *AqE* gene were used in the analysis, since the gene is highly conserved among vertebrates. In addition, LDH, MDH1 and MDH2 of the same organisms were taken as a possible outgroup. Since proteins encoded by *AqE* are often given the names “*dpkA*-like” and “*yiaK*-like” during annotation, we included in the analysis the well-known LDH2/MDH2 oxidoreductases, which include *dpkA*, *yiaK*, *comC* and *AllD*. However, since these enzymes have been characterized only in archaea and bacteria, proteins from these taxa were used in the study. Next, we added *slcC* encoded sulfolactate dehydrogenases to the analysis, because it was not clear whether this protein is a representative of LDH2/MDH2 oxidoreductases.

Codon-Based Test of Purifying Selection for Analysis Between Sequences

The study of the selection and application of the Z-test of selection was performed using MEGA 7 (Kumar et al. 2016). Analyses were conducted using the Kumar method (Nei and Kumar 2000). The analysis involved 54 transcribed RNA sequences of the *AqE* gene, the accession of which are given

in Additional file 1. The variance of the difference was computed using the analytical method. All positions containing gaps and missing data were eliminated.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00239-021-09997-x>.

Author Contributions PLV made a substantial contributions to the conception and design of the work, interpretation of data; have drafted the work. PMV made substantial contributions to the acquisition, analysis and interpretation of data. GOL made substantial contributions to the acquisition, analysis and interpretation of data. All the authors read and approved the final manuscript.

Funding The reported study was funded by the Russian Foundation for Basic Research and the city of Sevastopol, Project Number 20-44-920006 (investigation of *AqE* gene in the genomes of cartilaginous and bony fishes) and by the Russian Academy of Sciences Research Grant No. AAAA-A18-118021490093-4 (investigation of *AqE* gene in the genomes of lampreys and hagfishes).

Data Availability The datasets analyzed during the current study are available in the GenBank repository <https://www.ncbi.nlm.nih.gov/genbank/>. Some data generated or analyzed during this study are included in supplementary information files of this article.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
- Bailey TL, Bodén M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res* 37:W202–W208
- Berthelot C, Brunet F, Chalopin D, Juanchich A, Bernard M, Noël B, Bento P, Da Silva C, Labadie K, Alberti A, Aury JM, Louis A, Dehais P, Bardou P, Montfort J, Klopp C, Cabau C, Gaspin C, Thorgaard GH, Boussaha M, Quillet E, Guyomard R, Galiana D, Bobe J, Volff JN, Genêt C, Wincker P, Jaillon O, Roest Crollius H, Guiguen Y (2014) The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat Commun* 5:3657
- Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA et al (2007) Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* 450(7167):203–218
- Copley SD (2017) Shining a light on enzyme promiscuity. *Curr Opin Struct Biol* 47:167–175
- Denger K, Cook AM (2010) Racemase activity effected by two dehydrogenases in sulfolactate degradation by *Chromohalobacter salexigens*: purification of (S)-sulfolactate dehydrogenase. *Microbiology* 156:967–974
- Denger K, Mayer J, Buhmann M, Weinitschke S, Smits THM, Cook AM (2009) Bifurcated degradative pathway of 3-sulfolactate in *Roseovarius nubinhibens* ISM via sulfoacetaldehyde acetyltransferase and (S)-cysteate sulfo-lyase. *J Bacteriol* 191:5648–5656
- Drouin G, Godin JR, Pagé B (2011) The genetics of vitamin C loss in vertebrates. *Curr Genom* 12(5):371–378
- Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res* 32:1792–1797
- Eventoff W, Rossmann MG (1975) The evolution of dehydrogenases and kinases. *CRC Crit Rev Biochem* 3(2):111–140
- Ferris SD (1984) Tetraploidy and the evolution of the catostomid Fishes. In: Turner B.J. (eds) *Evolutionary Genetics of Fishes. Monographs in Evolutionary Biology*. Springer:Boston, MA.
- Goldman-Huertas B, Mitchell RF, Lapoint RT, Faucher CP, Hildebrand JG, Whiteman NK (2015) Evolution of herbivory in *Drosophilidae* linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. *Proc Natl Acad Sci USA* 112:3026–3031
- Graupner M, Xu H, White RH (2000) Identification of an archaeal 2-hydroxy acid dehydrogenase catalyzing reactions involved in coenzyme biosynthesis in methanoarchaea. *J Bacteriol* 182:3688–3692
- Greenberg AJ, Moran JR, Coyne JA, Wu CI (2003) Ecological adaptation during incipient speciation revealed by precise gene replacement. *Science* 302:1754–1757
- Hochachka PW, Somero GN. *Biochemical adaptation: mechanisms and process of physiological evolution*. Oxford University Press;New York. 2002.
- Holmes RS (2019) Polyploidy among salmonid aldehyde dehydrogenase genes and proteins. *Chem Biol Interact* 303:22–26
- Honka E, Fabry S, Niermann T, Palm P, Hensel R (1990) Properties and primary structure of the L-malate dehydrogenase from the extremely thermophilic archaeobacterium *Methanothermobacter feravidus*. *Eur J Biochem* 188:623–632
- Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G (2015) GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics* 31(8):1296–1297
- Irimia A, Madern D, Zaccari G, Vellieux FMD (2004) Methanoarchaeal sulfolactate dehydrogenase: prototype of a new family of NADH-dependent enzymes. *The EMBO Journal* 23:1234–1244
- Jendrossek D, Kratzin HD, Steinbuechel A (1993) The *Alcaligenes eutrophus ldh* structural gene encodes a novel type of lactate dehydrogenase. *FEMS Microbiol Lett* 112:229–235
- Kumar S, Stecher G, Tamura K (2016) MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol* 33:1870–1874
- Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290(5494):1151–1155. <https://doi.org/10.1126/science.290.5494.1151>
- Madern D (2002) Molecular evolution within the L-malate and L-lactate dehydrogenase super-family. *J Mol Evol* 54:825–840
- McBride CS, Arguello JR, O'Meara BC (2007) Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics* 177:1395–1416
- Moreau R, Dabrowski K (1998) Body pool and synthesis of ascorbic acid in adult sea lamprey (*Petromyzon marinus*): an agnathan fish with gulonolactone oxidase activity. *Proc Natl Acad Sci USA* 95:10279–10282
- Muramatsu H, Mihara H, Goto M, Miyahara I, Hirotsu K, Kurihara T, Esaki N (2005) A new family of NAD(P)H-dependent oxidoreductases distinct from conventional Rossmann-fold proteins. *J Biosci Bioeng* 99(6):541–547
- Nei M, Kumar S (2000) *Molecular Evolution and Phylogenetics*. Oxford University Press;New York.
- Petit J, David L, Dirks R, Wiegertjes GF (2017) Genomic and transcriptomic approaches to study immunology in cyprinids: What is next? *Dev Comp Immunol* 75:48–62

- Puzakova LV, Puzakov MV, Soldatov AA (2019) Gene encoding a novel enzyme of LDH2/MDH2 family is lost in plant and animal genomes during transition to land. *J Mol Evol* 87:52–59
- Walsh B (2003) Population-genetic models of the fates of duplicate genes. *Genetica* 118(2–3):279–294
- Zhang Y, Schofield LR, Sang C, Dey D, Ronimus RS (2017) Expression, purification, and characterization of (R)-sulfolactate dehydrogenase (ComC) from the rumen methanogen *Methanobrevibacter millerae* SM9. *Archaea*. <https://doi.org/10.1155/2017/5793620>