

The new AqE gene studied in non-tetrapoda vertebrates is high conserved

Lyudmila V. Puzakova (✉ kvluda@yandex.ru)

Institut biologii pivdennih moriv imeni O O Kovalevskogo Nacional'na akademii nauk Ukraini
<https://orcid.org/0000-0001-6747-4313>

Mikhail V. Puzakov

A.O. Kovalevsky institute of biology of the southern seas of RAS

Olga L. Gostyukhina

A.O. Kovalevsky institute of biology of the southern seas of RAS

Research

Keywords: gene evolution, gene loss, vertebrates, LDH2/MDH2 oxidoreductase, L-sulfolactate dehydrogenase

Posted Date: February 19th, 2020

DOI: <https://doi.org/10.21203/rs.2.23929/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The AqE gene encoding the NAD(P)H-dependent LDH2/MDH2 oxidoreductase has been described in living organisms predominantly in the aquatic environment. It was shown that the gene is present in bacteria and archaea. In the plant kingdom this gene is present only in algae. Animals have the AqE gene in all groups, from protozoa to fish. In the ancestor of tetrapods, the gene disappears and, accordingly, is absent in amphibians, reptiles, birds and mammals. It was suggested that the gene is involved in anaerobic respiration. The loss of the AqE gene in animals and plants is likely to be associated with living on land and the corresponding restructuring of their metabolic pathways, due to oxygen saturation and the absence of natural hypoxia as in aquatic organisms

Results: A 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, as well as 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.

Conclusions: In most orders of non-tetrapoda vertebrates, the AqE gene remains highly conserved. This suggests that the AqE in aquatic vertebrates is an essential gene and undergoes to rigorous selection. AqE gene has the highest homology with the archaeal ComC that encoding SLDH. Based on the similarity of substrates, it cannot be excluded that the enzyme encoded by the AqE gene is involved in the malate-aspartate shuttle mechanism or biosynthesis of energy coenzyme M equivalent.

Background

Oxygen bioavailability for hydrobionts is the main factor determining the possibility of their existence and distribution [1]. Currently a tendency to increase the formation of hypoxia zones where oxygen concentration less than 0.5 mg l⁻¹ is observed in the World Ocean [2–4].

As a rule, the formation of such zones due to a combination of several anthropogenic factors – excessive intake of biogenic compounds (eutrophication), increased water turbidity and, as a result, decreased water transparency, siltation of the bottom, and the development of phytoplankton [5]. In addition, the vertical stratification as a result of the formation of thermocline and/or halocline, weak water mixing, summer temperature increase, and others can serve as a cause of the formation of hypoxic zones [6–8].

In addition to external causes, aquatic organisms may experience hypoxia of internal origin, or so-called functional hypoxia, which is caused primarily by physiological processes that require high energy consumption (for example, active muscle work) [9, 10], the action of toxicants that affect metabolism, the processes of growth, adaptation, development and reproduction, the intensity of nutrition [11–13], as well as various pathologies [7].

The "hypoxic" type of energy metabolism is the most ancient and, apparently, the most diverse regarding metabolic pathways. Undoubtedly, switching to anaerobic energy generation pathways is of particular importance [6, 10, 14]. Despite the modern knowledge of metabolic processes, many alternative, unknown

metabolic pathways are likely to exist. It is likely that hypoxia-tolerant animals may have undescribed metabolic pathways that “turn on” in a low-oxygen environment [15].

Having passed to a terrestrial life organisms no longer experience oxygen deficiency and many anaerobic processes important for aquatic organisms became insignificant or even redundant. Under these conditions, genes of “hypoxic” metabolic systems could be lost. In particular, such losses could affect individual NAD (P) H-dependent oxidoreductases, one of the most important groups of enzymes involved in carbohydrate metabolism and playing a key role in the adaptation of organisms to hypoxia (anoxia) conditions [6, 10, 14].

The AqE gene encoding the NAD(P)H-dependent LDH2/MDH2 oxidoreductase has been described in living organisms predominantly in the aquatic environment [16]. It was shown that the gene is present in bacteria and archaea. In the plant kingdom this gene is present only in algae. Animals have the AqE gene in all groups, from protozoa to fish. In the ancestor of tetrapods, the gene disappears and, accordingly, is absent in amphibians, reptiles, birds and mammals. It was suggested that the gene is involved in anaerobic respiration. The loss of the AqE gene in animals and plants is likely to be associated with living on land and the corresponding restructuring of their metabolic pathways, due to oxygen saturation and the absence of natural hypoxia as in aquatic organisms [16].

Due to the fact that the primary sequence of genomic DNA is determined in more than 30 thousand species at the moment, in a previous study we worked on large taxa in order to cover all groups of organisms when searching for the AqE gene [16]. As a result, we could observe the distribution of AqE only in the “thick” branches of the tree of life. In animals, the AqE gene loss was revealed in four classes: Amphibians, Reptiles, Birds and Mammals, that are of Vertebrata group. In this paper, the presence and evolution of the AqE gene in non-tetrapod vertebrates were studied in detail with taxonomic resolution up to order.

Results

The AqE gene landscape in genomes of non-tetrapod vertebrates

In a study of the AqE gene distribution among vertebrates, 86 orders of vertebrates not belonging to the Tetrapoda group were analyzed. Whole-Genome Shotgun sequences (WGS) in the NCBI database at the time of the study were presented only for members of 58 orders out of 86 analyzed. As a result of the analysis in non-redundant protein sequences and WGS databases using BLASTp and tBLASTn, respectively, 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 to study this enzyme [16]. In addition, channel catfish AqE has a predicted exon-intron structure (GeneID:100528876).

Overall, nucleotide sequences homologous to AqE were revealed in 101 species (56 orders). In 16 organisms belonging to 2 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 fact is confirmed by the introns absence of *H. hucho* gene and the high homology of the *H. hucho* protein to bacteria *Vogesella perlucida* AqE (80,24%) while low identity of the *H. hucho* protein to AqE of *I. punctatus* (26.97%).

In three species *Tachysurus fulvidraco*, *Amphiprion ocellaris* and *Acipenser ruthenus* (orders Siluriformes, Pomacentridae и Acipenseriformes, respectively) AqE gene was found in two copies. Such a duplication of the gene is the exception rather than the trend, since 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 WGS were absent, so there is no any data about AqE presence in organisms of these taxa. Only in representatives of one order (Cypriniformes) no any homologies to AqE gene were found. Since WGS of 15 species of Cypriniformes were analyzed, this result can be considered reliable and, accordingly/respectively, the AqE gene was lost in this group (Additional file 1, Fig. 1).

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

In the studied representatives of chimaeras and sharks AqE gene was detected (Additional file 1, Fig. 1). In *Raja erinacia* (rays) no any homologies to AqE gene was found. Since this is the only species with WGS from superorder Batoidea, the gene absence in *R. erinacia* can be considered both as an 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 (26x). In no rank group Cyclostomata, the AqE gene was detected in representatives of both orders of jawless vertebrates, Petromyzontiformes (lampreys) and Myxiniiformes (hagfishes).

Structure Variations of the AqE Genes

To better understand the evolution of the AqE gene in vertebrates, the exon-intron structure of the coding sequence (coding DNA sequence, CDS) was analyzed. It was first defined for 50 species. AqE gene CDS of the predominant number of non-tetrapod vertebrates was quite conservative and on average consisted of 11 exons. Due to the different length of the introns, the length of the CDS with introns varied from 3 kbp 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 was characterized as hypothetically whole. These species did 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, since its extent is extremely small (apparently 3 aa). For the same reason, we could not detect exon 11 (the last) due to its significant variability in some species pointed above. Nevertheless, we suppose that these organisms 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, as well as exons 5 and 6, was revealed. In *Xiphophorus couchianus* (Cyprinodontiformes), exons 2 and 6 were only partially present, and exons 8 and 9 were fused. Whereas in the other seven representatives of Cyprinodontiformes, the gene was similar to the "classical" one. In representatives of 8 orders (Ateleopodiformes, Aulopiformes, Lophiiformes, Beryciformes, Lampriformes, Polymixiiformes, Stylephoriformes, 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 18322 bp. The assemblies of species from these orders had rather low genome coverage (< 25x) and 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 and pseudogenization, and we believe that the AqE gene in these orders is full-sized.

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

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

Phylogenetic relationships among the AqE genes of 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 taxonomic division. Three orders Beryciformes, Pleuronectiformes и Blenniiformes were exception, whose representatives turned to be distribute on different brunches. Nevertheless, taken into consideration the low bootstrap values within the Euteleostomorpha clade (Additional file 4) and the associated polytomy, the overall picture of phylogenetic relationships remains classical. The dendrogram also demonstrates that the increase in copies in individual species is associated exclusively with intraspecific duplications.

Sequence Modification Of The Aqe Proteins

To identify conserved motifs (CM) and definition 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), the other 5 were rare (3–14 proteins). The number of conserved motifs in each AQE protein ranged from 5 to 10. The most AQE proteins had 10 conserved motifs. A detailed analysis of conserved motifs showed that CM-15 is an alternative (modification) of CM-10 in representatives of Chondrichthyes. CM-12 were found in deleted proteins and isoforms PRODUCED FROM transcripts with an alternative start of translation in exon 2, that is actually a fragment of CM-2. CM-13 as well as CM-12 is a shortened version 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*, *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 – CM-11). Therefore, AqE genes demonstrate rather high conservatism during the evolution of vertebrates.

Discussion

Possible reasons for Variation of AqE distribution in non-tetrapoda vertebrates

Distribution and structure of the AqE gene in vertebrates not belonging to the Tetrapoda group (with taxonomic resolution up to order) were analyzed in this investigation for the first time. It was shown that representatives of the most part (54 from out 56) of studied non-tetrapod vertebrates orders have the AqE gene (Additional file 1, Fig. 1). In three species (*Tachysurus fulvidraco*, *Amphiprion ocellaris*, and *Acipenser ruthenus*) duplication of AqE gene was revealed, what was most likely associated with intraspecific duplications exclusively. An increase of gene copies number in genomes usually occurs with a determine frequency and most often “extra” copies are gradually pseudogenized [17].

No homologies to the AqE gene were found in genomes of two orders (Rajiformes, Cypriniformes) representatives. Genome sequences of only one representative were available for the analysis in the order Rajiformes. Therefore we cannot confirm that the AqE gene absence is a characteristic feature of this order, but not a species-specific phenomenon or the result of insufficiently high-quality sequencing or assembly. In the order Cypriniformes the gene is absent in all 15 species analyzed, therefore, the AqE gene loss in this taxon is out of doubt. In species of the order Salmoniformes the AqE gene underwent substantial deleting. This result is confirmed by the study of 7 species of this taxon. Thus, the fact that organisms of Cypriniformes and Salmoniformes lost the AqE gene (entirely or partly) is reliable.

The result obtained is quite unexpected since we initially considered the hypothesis that all aquatic organisms had the AqE gene [16]. In representatives of the rest orders of non-tetrapoda vertebrates, the gene is not only present, but still retains a rather high conservatism and is transcribed. These data confirm our assumption that the AqE gene is required for aquatic organisms and therefore it is under the influence of stabilizing selection. In that case why did Cypriniformes и Salmoniformes lose the AqE gene which was so necessary for other taxa? Gene loss is known to be a rather common phenomenon in the evolution. Gene loss can have a neutral effect on vital activity [18, 19] or significantly increase the adaptive potential of a species [20–23]. 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 and its transformation to pseudogene and further gradual degradation (fragmentation). Another way of the process is sudden and complete gene loss (deletion) due to unequal crossing over during meiosis or mobile genetic elements transposition. Gene pseudogenization occurs when the gene becomes redundant. When a 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.

It is cannot be excluded 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, but the process resulted in such a CONSIDERABLE degradation that homologies cannot be found. In Salmoniformes only 2 exons were preserved. Such gene degradation is characteristic of pseudogenization.

The loss of the AqE gene in these bony fish orders may have occurred as a result 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 the lack of a specific LDH-like MDH [24] to act as analogous MalDHs to compensate 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 have occurred independently in Cypriniformes и Salmoniformes (the fourth whole-genome duplication) [25, 26]. An excess of oxydoreductases resulted from genome duplication may have become a reason of «painless» AqE gene loss, because more “substance” for non-homologous replacement or the formation of new pathways and the occurrence of new enzymes appeared. As a result of evolutionary processes following after duplication, new «advantageous» allele combinations for the genome or completely new alleles could appear. According to S. Copley [27], on average, an enzyme can have 10 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 [28]. Thus, polyploidy in Salmoniformes resulted in the presence of at least 30 aldehyde dehydrogenase genes, it is

more than in other higher vertebrates [29]. A trend towards an increase in the number of genes is observed for many enzymes, including lactate dehydrogenase, creatine kinase, glucose phosphatase, etc. [cited on 30].

Putative function AqE gene

In the previous study the AqE gene was shown to encode LDH2/MDH2 oxidoreductase. This enzyme is phylogenetically the most similar to sulfolactate dehydrogenase S-SLDH [EC 1.1.1.337] and MDH/SLDH [EC 1.1.1.310] as well to MDH(1V9N) and MDH(3I0P) enzymes whose function was not determined experimentally [16]. Sulfolactate is a natural compound in many organisms (bacteria, archaea, plants and animals) [cited on 31]. But there is some diversity in metabolic pathways which sulfolactate involved in. At least three different sulfolactate dehydrogenase (SLDH) are described. The most studied enzyme [EC 1.1.1.272] is SLDH, encoded by ComC gene. This enzyme involved in the coenzyme M biosynthesis in methanogenic archaea and spore-formers. ComC oxidoreductase interconverts R-sulfolactate to sulfopyruvate. In subsequent biochemical transformations sulfopyruvate becomes coenzyme M [15, 32, 33]. Using sulfolactate sulfo-lyase (SuyAB) R-sulfolactate can be transformed in pyruvate, which is «a crossing point» of many metabolic pathways. Another studied sulfolactate dehydrogenase [EC 1.1.1.-], encoded by SlcC gene, converts S-sulfolactate to sulfopyruvate, which in turn can be used by ComC. Thus, ComC and SlcC perform sulfolactate racemisation acting in pair [31]. The third, least studied enzyme [EC 1.1.99.-], encoded by the SlcD gene, is bound to the membrane and is involved in the degradation of sulfolactate in the bacteria *Roseovarius nubinhibens* [34]. Enzymes of LDH/MDH and LDH2/MDH2 family were involved in phylogenetic analysis which revealed that LDH2/MDH2 oxidoreductase encoded by the AqE gene had the highest homology with the archaeal ComC clade members (Fig. 2). The enzymes encoded by bacterial ComC form a separate clade. The enzymes encoded by SlcC gene may not even be members of LDH2/MDH2 oxidoreductase family, since formed its own clade. 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 preference sulfolactate for using. In methanogenic archaea and in spore-formers, this enzyme is involved in the biosynthesis of coenzyme M (methanogenic cofactor) [15, 31]. In other organisms this enzyme is not described, hence, the function of the AqE gene remains unknown, in particular in eukaryotes. As Irimia et al. [24] suppose converting sulfolactate to sulfopyruvate in eukaryotes does not make sense since 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 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, usually expressed in the regulation of the balance of aerobic and anaerobic processes. SLDH in aquatic vertebrates is likely to be an enzyme of reserve pathways, which supplement the main metabolic energy processes under conditions of oxygen

deficiency. This assumption is confirmed by two facts. Firstly, the AqE gene disappears in terrestrial vertebrates (we associate this with the presence of free oxygen in atmosphere). Secondly, the AqE protein is the most similar to SLDH which involved in anaerobic processes [15, 32, 33]. Malate-aspartate shuttle mechanism, in which malate and α -ketoglutarate are the key compounds, is considered to be the most effective process that allows aquatic organisms to survive under hypoxia (anoxia) conditions [35]. 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 sulpholactate [15] we also suggest that the product of AqE gene can be included in the malate-aspartate shuttle mechanism. The key enzyme of 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 [35]. We do not exclude that this could be an enzyme encoded by the AqE gene.

Conclusion

A 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, as well as 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 to rigorous selection. Therefore, the enzyme is actively involved in metabolic pathways that are still unknown.

AqE gene has the highest homology with the archaeal ComC that encoding 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:

- 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 the organism with energy in the form of NADH;
- a pathway of sulpholactate to sulphopyruvate conversion followed by the formation of energy equivalents in the form 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 query to identify homologous genes in the Whole-Genome Shotgun sequences (WGS) of non-tetrapod vertebrates (Additional file 3). Search was carried out using the Basic Local Alignment Search Tool (BLAST) [36]. First

we searched for homologies among non-redundant protein sequences of non-tetrapod vertebrates in order to find the species (orders) in which the structure of the AqE gene is determined as a result of 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 visually by the highest homology between query and studied sequence and the presence of 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 coding AqE sequences (CDS) obtained from the analysis were used to search transcribed RNA sequences in transcriptome shotgun assembly database (TSA).

Gene Structure and Conserved Motif Analysis of AqE Genes

The exon-intron structure of AqE genes were displayed via Gene Structure Display Server 2.0 [37] based on the alignment of their coding sequences with their corresponding genomic sequences. The MEME suite server [38] was used to identify the conserved motifs of the proteins encoding by AqE genes, and the parameters used in this study were as follows: maximum number of different motifs, 20; minimum width, 10; and maximum width, 50.

Phylogenetic Analyses

Multiple alignment of the amino acid sequences were performed using MUSCLE [39], and the resulting data were used to construct a phylogenetic tree via the MEGA 7 software [40] with the maximum likelihood (ML) methods. The analysis involved 98 amino acid sequences (Additional file 6). With respect the ML method, the following parameters were used: bootstrap, 100 replicates; Jones–Taylor–Thornton; Gamma distribution. Evolutionary relationships of LDH/MDH and LDH2/MDH2 families enzymes and AqE proteins was inferred using the Neighbor-Joining method, bootstrap 1000. The analysis involved 51 amino acid sequences (Additional file 7).

Declarations

List of abbreviations

bp - base pair, aa - amino acids, ORF - open reading frame, WGS - whole-genome shotgun sequences, TSA – transcriptome shotgun assembly, SLDH – sulfolactate dehydrogenase, CDS – coding DNA sequence

Ethics approval and consent to participate

Not applicable

Consent for publication

Not applicable

Availability of data and materials

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

Competing interests

The authors declare that they have no competing interests

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).

Authors' 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 authors read and approved the final manuscript.

Acknowledgements

Not applicable.

References

1. Mandic M, Lau GY, Nijjar MM, Richards JG. Metabolic recovery in goldfish: A comparison of recovery from severe hypoxia exposure and exhaustive exercise. *Comp Biochem Physiol C Toxicol Pharmacol*. 2008;148(4):332-8.
2. Middelburg JJ, Levin LA. Coastal hypoxia and sediment biogeochemistry. *Biogeosciences*. 2009;6:1273-93.
3. Gewin V. Dead in the water. *Nature*. 2010;466:812-4.
4. Heath A. *Water Pollution and Fish Physiology*. Boca Raton: CRC Press, 1995.
5. Tishchenko P, Tishchenko P, Lobanov V, Sergeev A, Semkin P, Zvalin V. Summertime in situ monitoring of oxygen depletion in Amursky Bay (Japan/East Sea). *Continental Shelf Research*. 2016;118:77–87.
6. Wu SSR. Hypoxia: from molecular responses to ecosystem responses. *Marine Pollution Bulletin*. 2002;45(1):35–45.
7. Farrell AP, Richards JG. Chapter 11. Defining Hypoxia: An Integrative Synthesis of the Responses of Fish to Hypoxia. *Fish Physiology*. 2009;27:487–503.

8. Seibel BA. Metabolic responses and unique environmental adaptations critical oxygen levels and metabolic suppression in oceanic oxygen minimum zones. *The Journal of Experimental Biology*. 2011;214:326-36.
9. Herbert NA, Steffensen JF. The response of Atlantic cod, *Gadus morhua*, to progressive hypoxia: fish swimming speed and physiological stress. *Marine Biology*. 2005;147(6):1403–12.
10. Richards JG. Physiological, behavioral and biochemical adaptations of intertidal fishes to hypoxia. *The Journal of Experimental Biology*. 2011;214:191-9.
11. Meshcheryakova OV, Churova MV, Nemova NN. Mitochondrial lactate oxidation: mechanism and importance at the temperature adaptation. *Comparative Biochemistry and Physiology. Part A: Molecular and Integrative Physiology*. 2012;163:S5-S6.
12. Nemova NN, Meshcheryakova OV, Churova MV. Energy metabolism indicators for Salmonidae growth and development assessment. *Uchenye zapiski Petrozavodskogo gosudarstvennogo universiteta*. 2015;153(8):7-13.
13. Churova MV, Meshcheryakova OV, Veselov AE, Nemova NN. Activity of enzymes involved in the energy and carbohydrate metabolism and the level of some molecular-genetic characteristics in young salmon (*Salmo salar* L.) with different age and weight. *Russian Journal of Developmental Biology*. 2015;46(5):254-62.
14. Hochachka PW. Defense strategies against hypoxia and hypothermia. *Science*. 1986;231(4735):234-41.
15. Muramatsu H, Mihara H, Goto M, Miyahara I, Hirotsu K, Kurihara T, Esaki N. A new family of NAD(P)H-dependent oxidoreductases distinct from conventional Rossmann-fold proteins. *Journal of bioscience and bioengineering*. 2005;99(6):541–7.
16. Puzakova LV, Puzakov MV, Soldatov AA. Gene encoding a novel enzyme of LDH2/MDH2 family is lost in plant and animal genomes during transition to land. *Journal of Molecular Evolution*. 2019;87:52–9.
17. Michael AJ. Evolution of biosynthetic diversity. *Biochemical Journal*. 2017;474:2277–99.
18. Moreau R, Dabrowski K. 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*. 1998;95:10279–82.
19. Drouin G, Godin JR, Pagé B. The genetics of vitamin C loss in vertebrates. *Curr Genom*. 2011;12(5):371–378.
20. Greenberg AJ, Moran JR, Coyne JA, Wu CI. Ecological adaptation during incipient speciation revealed by precise gene replacement. *Science*. 2003;302:1754–57.
21. Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow TA et al. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature*. 2007;450(7167):203–18.
22. McBride CS, Arguello JR, O'Meara BC. Five *Drosophila* genomes reveal nonneutral evolution and the signature of host specialization in the chemoreceptor superfamily. *Genetics*. 2007;177:1395–1416.

23. Goldman-Huertas B, Mitchell RF, Lapoint RT, Faucher CP, Hildebrand JG, Whiteman NK. Evolution of herbivory in Drosophilidae linked to loss of behaviors, antennal responses, odorant receptors, and ancestral diet. *Proc Natl Acad Sci USA*. 2015;112:3026–31.
24. Irimia A, Madern D, Zaccari G, Vellieux FMD. Methanoarchaeal sulfolactate dehydrogenase: prototype of a new family of NADH-dependent enzymes. *The EMBO Journal*. 2004;23:1234–44.
25. 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. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nat Commun*. 2014;5:3657.
26. Petit J, David L, Dirks R, Wiegertjes GF. Genomic and transcriptomic approaches to study immunology in cyprinids: What is next? *Developmental & Comparative Immunology*. 2017;75:48-62.
27. Copley SD. Shining a light on enzyme promiscuity. *Curr Opin Struct Biol*. 2017;47:167-75.
28. Eventoff W, Rossmann MG. The evolution of dehydrogenases and kinases. *CRC Crit Rev Biochem*. 1975;3(2):111-40.
29. Holmes RS. Polyploidy among salmonid aldehyde dehydrogenase genes and proteins. *Chemico-Biological Interactions*. 2019;303:22–6.
30. Ferris SD. Tetraploidy and the evolution of the catostomid Fishes. In: Turner B.J. (eds) *Evolutionary Genetics of Fishes*. Monographs in Evolutionary Biology. Springer:Boston, MA. 1984
31. Denger K, Cook AM. Racemase activity effected by two dehydrogenases in sulfolactate degradation by *Chromohalobacter salexigens*: purification of (S)-sulfolactate dehydrogenase. *Microbiology*. 2010;156:967–74.
32. Graupner M, Xu H, White RH. Identification of an archaeal 2-hydroxy acid dehydrogenase catalyzing reactions involved in coenzyme biosynthesis in methanoarchaea. *J Bacteriol*. 2000;182:3688–92.
33. Zhang Y, Schofield LR, Sang C, Dey D, Ronimus RS. Expression, purification, and characterization of (R)-sulfolactate dehydrogenase (ComC) from the rumen methanogen *Methanobrevibacter millerae* SM9. *Archaea*. 2017:5793620. doi:10.1155/2017/5793620.
34. Denger K, Mayer J, Buhmann M, Weinitschke S, Smits THM, Cook AM. Bifurcated degradative pathway of 3-sulfolactate in *Roseovarius nubinhibens* ISM via sulfoacetaldehyde acetyltransferase and (S)-cysteate sulfo-lyase. *J Bacteriol*. 2009;191:5648–56.
35. Hochachka PW, Somero GN. *Biochemical adaptation: mechanisms and process of physiological evolution*. Oxford University Press;New York. 2002.
36. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25:3389–402.
37. Hu B, Jin J, Guo A-Y, Zhang H, Luo J, Gao G. GSDS 2.0: an upgraded gene feature visualization server. *Bioinformatics*. 2015;31(8):1296-7.

38. Bailey TL, Bodén M, Buske FA, Frith M, Grant CE, Clementi L, Ren J, Li WW, Noble WS. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Research*. 2009;37:W202-W208.
39. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;32:1792–7.
40. Kumar S, Stecher G, Tamura K. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol Biol Evol*. 2016;33:1870–4.

Supplementary Files Legend

Additional file 1. List of vertebrates orders and species that are studied in the work. NSSF - no significant similarity found, nd – no data

Additional file 2. The exon–intron structure of non-tetrapoda vertebrates *AqE* genes. Exons are indicated by rectangles.

Additional file 3. Assembly features of the studied genomes. The genomes with genome coverage less 25x and number of sequences more 50,000 are indicated by red.

Additional file 4. The phylogeny based on amino-acid sequences of the non-tetrapoda vertebrates *AqE* gene. This tree was generated in MEGA7 with the Maximum Likelihood method, using the JTT+G model. Only bootstrapping values higher than 70% are written on the branch. The analysis involved 98 amino acid sequences.

Additional file 5. The distribution of conserved motifs in AQE proteins, where the conserved motifs are indicated by colored boxes.

Additional file 6. List of amino acid sequences of AQE proteins used for phylogenetic analysis.

Figures

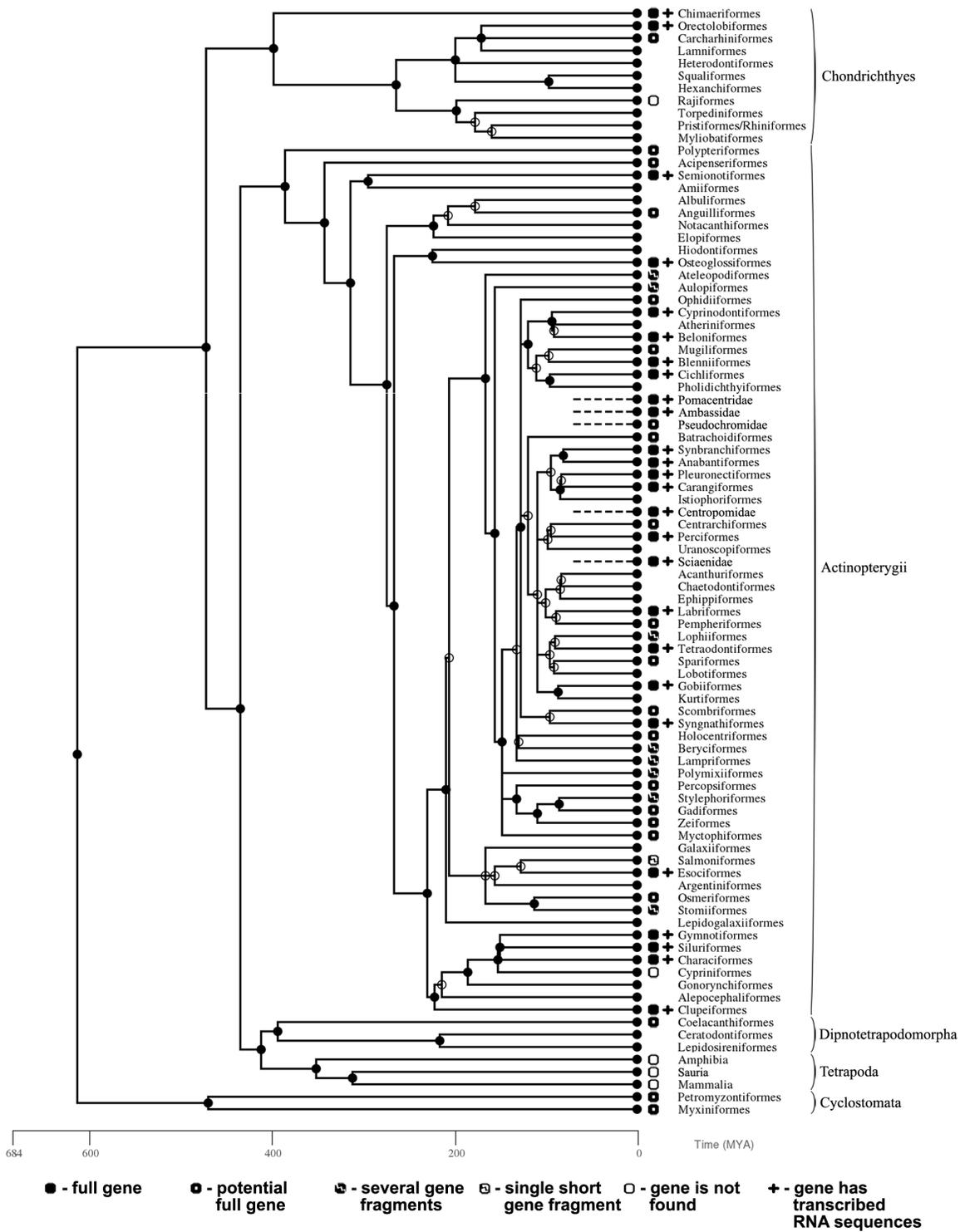


Figure 1

The distribution of AqE gene among vertebrates.

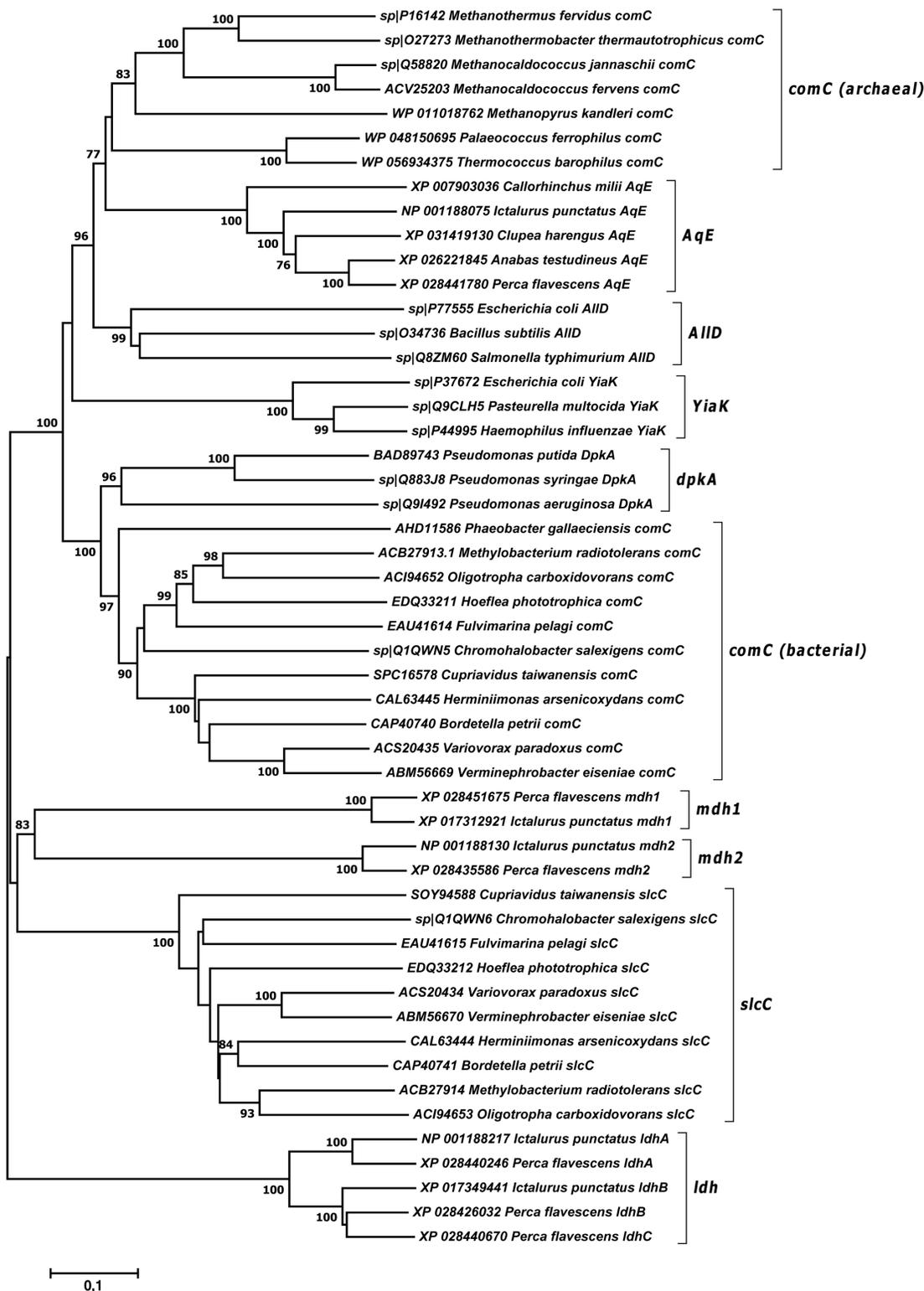


Figure 2

Evolutionary relationships of LDH/MDH and LDH2/MDH2 families 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.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Additionalfile7.pdf](#)
- [Additionalfile6.pdf](#)
- [Additionalfile2.jpg](#)
- [Additionalfile5.jpg](#)
- [Additionalfile1.pdf](#)
- [Additionalfile3.pdf](#)
- [Additionalfile4.jpg](#)