Genomic signature of shifts in selection and alkaline adaptation in highland fish

Chao Tong[1,2,*], Miao Li[3], Yongtao Tang[1,4], Kai Zhao[1,*]

[1] Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of Plateau Biology, Chinese Academy of Sciences, Xining, 810001, China;

[2] Department of Biology, University of Pennsylvania, Philadelphia, PA, 19104, USA;

[3] Center for Advanced Retinal and Ocular Therapeutics, Scheie Eye Institute, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, 19104, USA;

[4] College of Fisheries, Henan Normal University, Xinxiang, 453007, China

* Corresponding author:

Chao Tong, tongchao1990@gmail.com

Kai Zhao, zhaokai@nwipb.cas.cn

Abstract

Understanding how organisms adapt to aquatic life at high altitude is fundamental in evolutionary biology. This objective has been addressed mainly related to hypoxia adaptation by recent comparative studies, whereas highland fish has also long suffered extreme alkaline environment, insight into the genomic basis of alkaline adaptation has rarely been provided. Here, we compared the genomes or transcriptomes of 15 fish species, including two alkaline tolerant highland fish species and their six alkaline intolerant relatives, three alkaline tolerant lowland fish species and four alkaline intolerant relatives. We found putatively consistent patterns of molecular evolution in alkaline tolerant species in a large number of shared orthologs within highland and lowland fish taxa. Remarkably, we identified consistent signatures of accelerated evolution and positive selection in a set of core shared genes associated with ion transport, apoptosis, immune response and energy metabolisms in alkaline tolerant species within both highland and lowland fish taxa. This is one of the first comparative studies that began to elucidate the consistent genomic signature of alkaline adaptation shared by highland and lowland fish. This finding also highlights the adaptive molecular evolution changes that support fish adapting to extreme environments at high altitude.

Keywords

Comparative genomics; Molecular evolution; Alkaline adaptation; Schizothoracine fish

Significance Statement

Little is known about how wild fish responds to extreme alkaline stress besides hypoxia at high altitude. Comparative genomics has begun to elucidate the genomic basis of alkaline adaptation in lowland fish, such as killifish, but insight from highland fish has lagged behind. The common role of adaptive molecular evolution during alkaline adaptation in highland and lowland fish has rarely been discussed. We address this question by comparing 15 fish omics data. We find numbers of shared orthologs exhibited consistent patterns of molecular evolution in alkaline tolerant species relative to intolerant species. We further identify remarkably consistent signatures of rapidly evolving and positive selection in a substantial shared core of genes in both highland and lowland alkaline tolerant species.

Introduction

Environments shape the genetic landscape of the populations that inhabit them (Witt & Huerta-Sánchez 2019). The Tibetan Plateau had experienced continuous uplift during the India-Asia collision since approximately 45 million years ago, that triggered numerous environmental changes (Li & Fang 1999; Favre et al. 2015). As elevation above sea level increases, a decrease in barometric pressure results in fewer oxygen molecules in the air, which causes hypoxia. Besides, other harsh environments highland wildlife have encountered include the long-term low temperature, and intensified ultraviolet radiation (An et al. 2001). Large numbers of endemic Tibetan animals had developed unique morphological, physiological or genetic features to tolerate such harsh conditions (Wen 2014; Tong, Tian, et al. 2017; Tong, Fei, et al. 2017). Basically, understanding how organisms adapt to extreme environment is fundamental to address many evolutionary questions, but it remains a formidable task to fully uncover the mechanism of adaptive process (Scheinfeldt & Tishkoff 2010; Tong, Fei, et al. 2017; Tong, Tian, et al. 2017). Adaptation at molecular level can occur by adaptive mutation in key genes over prolonged evolutionary time scales (Orr 2005). Recent studies employing genome-wide approaches have identified candidate genes associated with hypoxia and long-term cold response in Tibetan terrestrial wildlife adaptation to high altitude (Qu et al. 2013; Wu et al. 2020). Nevertheless, the draft genomes of very few Tibetan aquatic wildlife are sequenced (Xiao et al. 2020; Liu et al. 2019), the genomic basis of highland adaptation in aquatic animals (e.g. fish) still remains largely unknown.

The schizothoracine fishes (Teleostei: Cyprinidae), the predominant fish fauna in the Tibetan Plateau, had evolved specific phenotypic characteristics to adapt to extreme aquatic environments, including hypoxia and long-term low temperature (Wu 1992; Cao et al. 1981). Recent comparative studies have identified key genes showing signals of positive selection during adaptation to such harsh environments (Yang et al. 2014; Wang et al. 2015; Kang et al. 2017), such as Hypoxia-inducible factor (HIF) (Guan et al. 2014) and Erythropoietin (EPO) (Xu et al. 2016) associated with hypoxia response, ATPase Family AAA Domain Containing 2 (ATAD2) (Tong, Fei, et al. 2017) and cAMP-dependent protein kinase catalytic subunit alpha (PRKACA) that involved into energy metabolism (Tong, Fei, et al. 2017). The main focus of previous studies in schizothoracine fishes are still on hypoxia and cold response. Notably, an increasing number of lakes in the Tibetan Plateau have been existing or towards alkaline due to

the global climate changes and human activities (Zheng 1997). Thus, the increasing alkalization of fresh water has been the potential challenge to schizothoracine fishes. Among the schizothoracine fishes, *Gymnocypris przewalskii* and *Gymnocypris kelukehuensis* are the only two species inhabited extremely alkaline environment. Unlike other broadly distributed schizothoracine fishes, such as *Gymnocypris eckloni*, *Schizopygopsis pylzovi* and *Platypharodon extremus* that inhihat in the Yellow river basin (Cao et al. 1981; Wu 1992; Qi et al. 2012), *G. przewalskii* only inhabits in saline and alkaline lake. As the largest salt lake in China, Lake Qinghai (FIG. 1) is highly saline (up to 13‰) and alkaline (up to PH 9.4) water environment, a typical salt lake with unusually high sodium, potassium and magnesium concentration (Zheng 1997; Zhu & Wu 1975). In addition, *G. kelukehuensis* only inhabits in a soda lake located at the Tsaidam Basin in the northestern Tibetan Plateau. Lake Keluke (FIG. 1) is also an soda lake with low salinity of 0.79‰ and high PH value up to 9.3 (Zheng 1997). Both schizothoracine fish species had developed unique physiological or genetic features to tolerate such harsh living conditions. Therefore, this provides an exceptional model to investigate the genetic mechanisms underlying alkaline adaptation, and may provide novel insights to fully understand the mechanism of highland adaptation in fish as complement.

Unlike highland alkaline tolerant fish, a huge amount of studies had explored the mechanisms of high saline and high alkaline tolerance in lowland fish species, such as killifish (e.g. *Fundulus heteroclitus*) (Wood et al. 2010; Johnson et al. 2020; Oziolor et al. 2019), tilapia (e.g. *Oreochromis niloticus*) (Zhao et al. 2020; Wood et al. 1994), and salmonids (e.g. *Salmo salar*) (Levings 2016; Lien et al. 2016). These studies had provided insights in physiology of acid-base balance in fishes response to high salinity or high PH environments, and suggested key genes, such as ion transport associated genes under selection during the adaptation (Oziolor et al. 2019; Lien et al. 2016).

In this study, we generated and assembled the transcriptomes of two alkaline tolerant schizothoracine fish species, *Gymnocypris przewalskii* and *G. kelukehuensis* inhabited in high PH environment in the northeastern Tibetan Plateau (FIG. 1). We performed a comparative genomics study together with recently sequenced schizothoracine fish transcriptomes and other lowland fish genomes (FIG. 2A), and sought to identify consistent genomic signature associated with alkaline adaptation in highland and lowland fishes. Specifically, we focused our

comparisons on testing whether alkaline adaptation in highland and lowland alkaline tolerant fishes is associated with the following signatures of molecular evolution: (1) consistent patterns of molecular evolution in protein-coding genes across the phylogeny; (2) consistent shifts in evolutionary rates for specific genes; and (3) consistent signals of positive selection in particular genes.

Materials and Methods

Sample collection

We collected eight adult *Gymnocypris przewalskii* (FIG. 1) individuals (four males and four females, 172 ± 0.7 g) from Lake Qinghai and eight adult *Gymnocypris kelukehuensis* (FIG. 1) individuals (four males and four females, 139 ± 0.3 g) from Lake Keluke using gill nets. All the fish samples were dissected after anesthesia with MS-222 (Solarbio, Beijing, China). All individuals were classified based on the gender and dissected after anesthesia with MS-222 (Solarbio, Beijing, China). Tissues from gill, kidney, brain, heart and liver from each individual were collected and immediately stored in liquid nitrogen at -80 °C. All the animal experiments were approved by the Animal Care and Use Committees of the Northwest Institute of Plateau Biology, Chinese Academy of Sciences (NWIPB-CY-010).

Transcriptomics

Total RNA of each tissue sample was extracted using TRIzol reagent (Invitrogen, CA, USA) in accordance with manufacturer's instructions, and detected for quality and quantity of RNAs with Nanodrop 1000 (NanoDrop Technologies, DE, USA) and Agilent Bioanalyzer 2100 (Agilent Technologies, CA, USA). Equal amount of RNA from eight individual of five tissue was pooled to construct transcriptome library as previously described (Tong, Tian, et al. 2017; Tong, Fei, et al. 2017), and sequenced with an Illumina NovaSeq 6000 yielding 150-bp paired-end (PE150) reads.

Sequencing reads were checked for quality using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Sequencing adapters and reads with a quality score < 20 were trimmed with Trimmomatic (Bolger et al. 2014). We built a *de*

novo transcriptome assembly based on clean reads using Trinity v2.6.5 (Grabherr et al. 2011) with default parameters. Next, we removed the redundant transcripts using CD-HIT (Fu et al. 2012) with the threshold of 0.90 and extracted the longest transcript as putative genes. We predicted the open reading frame of each putative genes using TransDecoder (https://github.com/TransDecoder/TransDecoder).

Additional data retrieval

We downloaded six alkaline intolerant schizothoracine fish transcriptomes (Zhou et al. 2020) including *Oxygymnocypris stewartii*, *Schizopygopsis younghusbandi*, *Gymnocypris namensis*, *Platypharodon extremus*, *Schizopygopsis pylzovi* and *Gymnocypris eckloni* from NCBI SRA database (https://www.ncbi.nlm.nih.gov/sra) (supplementary table S1), and performed assembly following above pipeline. In addition, we downloaded the genomes of four alkaline intolerant fish species of *Danio rerio*, *Ctenopharyngodon idellus*, *Cyprinus carpio*, and *Carassius auratus*, and three alkaline tolerant fish species of *Fundulus heteroclitus*, *Oreochromis niloticus* and *Salmo salar* (supplementary table S1).

Species phylogeny and gene orthology

We obtained the phylogenetic tree of 15 fish species by pruning the Fish Tree of Life (https://fishtreeoflife.org/) using R package, phangorn (Schliep 2011). To obtain phylogeny-based orthology relationships between different fish taxa, we included all the predicted proteomes of seven lowland fish genomes, and translated nucleotide sequences of protein-coding genes from eight schizothoracine fish transcriptome assemblies into amino acid sequences, and pooled these datasets as input for an orthology inference tool, OMA (Altenhoff et al. 2018). In this way, we identified one-to-one, one-to-many, and many-to-many orthologs among these 15 fish species. For further comparison, we restricted our analysis to 1:1 orthologs, that is the geneset for which only one gene from each species representing the orthology. In addition, we extracted the shared orthologs among highland fish taxa, lowland fish taxa, and all fish taxa. At last, gene ontology (GO) terms were assigned to each ortholog using Trinotate (https://trinotate.github.io/).

Pattern of molecular evolution in shared orthologs

To determine whether highland fish and lowland fish showing consistent patterns of molecular molecular in the set of alkaline tolerant species branches across the phylogeny, we charactererized the rates of non-synonymous to synonymous rate (dN/dS) in each shared ortholog. For this, we performed the protein sequence alignment using MUSCLE v3.8.31 (https://www.ebi.ac.uk/Tools/msa/muscle). We prepared the codon alignments of shared orthologs, that derived from protein alignments and the corresponding DNA sequences using PAL2NAL v.14 (Suyama et al. 2006). Then, we executed the filtration for shared ortholog alignments with length of at least 50 codons.

We took advantage of HyPHY pipeline (Kosakovsky Pond, Poon, et al. 2020) to test the hypotheses by comparing selective pressures (dN/dS) between *a prori* defined alkaline tolerant fish species branches (focal foreground branch) and alkaline intolerant species fish branches (background branch) in the specified fish phylogeny at ortholog-wide scale. Before that, a common approach to test the hypothesis is to perform separate analyses on subsets of sequences, and compare the parameter estimated in a *post hoc* fashion, that is one branch to the rest of branches in a phylogenetic tree like which we previously described (Tong, Tian, et al. 2017; Tong, Fei, et al. 2017; Tong et al. 2020). While, such approach is statistically suboptimal, and not always applicable (Kosakovsky Pond, Wisotsky, et al. 2020). In this way, we estimated two discrete categories of dN/dS for alkaline tolerant and alkaline intolerant species under the MG94 rate matrices combined with HKY85 model using HyPHY (Kosakovsky Pond, Poon, et al. 2020), and compared nested models with constrained relationships among them. We detected the shift of dN/dS ratios between alkaline tolerant and alkaline intolerant (alternative model, H1), relative to the null model that assuming all fish taxa have the same dN/dS ratio (H0). We constructed the log-likelihood ratio score for each ortholog (ΔlnL) as: $\Delta \ln L = 2(\ln LH_1 - \ln LH_0)$ and employed the likelihood ratio test. In this way, we repeated this analysis with shared ortholog datasets of all fish taxa, lowland taxa and highland taxa, respectively.

Analysis of accelerated evolution

To determine if consistent shift in evolutionary rates (e.g. acceleration) in particular genes within alkaline tolerant species across the phylogeny, we defined rapidly evolving genes with significantly higher dN/dS ratio for alkaline tolerant species than alkaline intolerant species ($P < 0.05$, likelihood ratio test). In this way, we identified the rapidly evolving genes in alkaline

tolerant species within highland and lowland based on earlier estimated dataset including two discrete dN/dS ratios for each shared ortholog, respectively. In addition, these sets of rapidly evolving genes were examined for GO enrichment relative to the full set of all shared orthologs using R package, topGO (https://bioconductor.org/packages/release/bioc/html/topGO.html). We finally visualized all significantly enriched GO terms remaining after the REVIGO (http://revigo.irb.hr/) similarity filter.

Analysis of positive selection

To further determine whether consistent signals of positive selection in a set of branches representing alkaline tolerant species across the phylogeny in specific genes, we used a branch-site model, BUSTED (Murrell et al. 2015) to test for positive selection signal in a gene at any site on focal branches (at least one site on at least one branch). In this way, we identified positively selected genes in alkaline tolerant species within highland fish taxa and lowland fish taxa, respectively. We finally performed GO enrichment analysis using topGO and REVIGO as earlier described.

Results

Following *de novo* assembly and annotation, each of the eight schizothoracine fish transcriptome assemblies had an average of 39,921 transcripts representing the complete or partial protein-coding regions of genes (supplementary table S2). We further identified a total of 7,309 one-to-one orthologs shared by all 15 fish species, and 6,241 core shared orthologs including all species (supplementary table S3).

Consistent pattern of molecular evolution

We estimated two categories of dN/dS ratios for 5,748 core shared orthologs (after restrict fiteration) in all fish taxa, lowland fish taxa, and highland fish taxa, separately. We found consistent patterns of molecular evolution showing significant ($P < 0.05$, likelihood ratio test) acceleration (increased dN/dS) or deceleration (decreased dN/dS) in a set of terminal branches of alkaline tolerant species relative to the set of terminal branches of alkaline intolerant species

in large numbers of shared orthologs within all fish taxa (n = 1,407), lowland fish taxa (n = 1,630) and highland fish taxa (n = 406) (FIG. 2B, supplementary table S4).

Consistent signature of accelerated evolution

Buiding on earlier dataset of two discrete categories of dN/dS ratios separately representing branches of alkaline tolerant and alkaline intolerant species across the phylogeny, we focused on genes with significantly higher dN/dS ($P < 0.05$, likelihood ratio test) in alkaline tolerant species, that is rapidly evolving gene (REGs) repertoire in alkaline tolerant species within highland and lowland fish taxa. We identified 240 REGs in highland fishes and 781 REGs in lowland fishes (FIG. 3A). Out of 41 overlapping REGs, we found a set of ion transport and transmembrane functions associated genes, such as sodium-dependent phosphate transport protein 2A (SLC34A1), transient receptor potential cation channel subfamily V member 1 (TRPV1), solute carrier family 7, member 3 (SLC7A3), and sodium-dependent phosphate transporter 1-B (SLC20A1b) (FIG, 3B, supplementary table S5). Besides, we identified overlapping REGs related to energy metabolism process, such ADP-dependent glucokinase (ADPGK) and mitochondrial pyruvate carrier 2 (MPC2) (FIG. 3B). Given that different REGs identified in either highland or lowland alkaline tolerant fish species, we found a number of ion transport and ATP synthesis related genes in REGs repertoire of highland fish (FIG. 3B, supplementary table S5), such as solute carrier family 35 member F4 (SLC35F4), calcium-activated potassium channel subunit beta-2 (KCNMB2), ATP-sensitive inward rectifier potassium channel 15 (KCNJ15), sodium-dependent serotonin transporter (SLC6A4), NAD-dependent protein deacetylase sirtuin-2 (SIRT2) and NADP-dependent malic enzyme (ME1). Similarly, in lowland fish, we also identified genes like solute carrier family 45 member 3 (SLC45A3), potassium inwardly-rectifying channel subfamily J member 2a (KCNJ2a), and ATP synthase F0 complex subunit B1 (ATP5PB) showing rapidly evolving in alkaline tolerant species (FIG. 3B, supplementary table S5).

Further, we did GO enrichment analysis for both REGs datasets, showing 134 significantly enriched GO terms (biological process) in highland fish ($P < 0.05$, fisher's exact test) and 137 significantly enriched GO terms in lowland fish. After the filtration by semantic similarity of GO terms, we found a set of enriched dominant GO terms in highland fish were related to ion transport and transmembrane functions (FIG. 3C, supplementary table S6), such as ion

transport (GO:0006811), regulation of calcium ion transport (GO:0051924), anion transport (GO:0006820). In addition, in lowland fish, we observed a set of enriched dominant GO terms associated with ion transport function (FIG. 3D, supplementary table S6), such as phosphate ion transport (GO:0006817) and response to salt stress (GO:0009651). Collectively, this finding suggested the consistent signature of accelerated evolution in alkaline tolerant species within highland taxa and lowland taxa.

Consistent signature of positive selection

Out of 5,748 shared orthologs, we identified 227 orthologs showing signals of positive selection on at least one site on at least one branch of alkaline tolerant species ($P < 0.05$, likelihood ratio test), namely positively selected genes (PSGs) in highland fish (FIG. 4A, supplementary table S7), and 274 PSGs in lowland alkaline tolerant fish species (FIG. 4A, supplementary table S7). Out of 16 overlapping PSGs, we found these genes were mainly related to apoptosis (cell death), ion transport and immune response, such as vitamin K-dependent protein C precursor (PROC), E3 ubiquitin-protein ligase SH3RF1 (SH3RF1), 14-3-3 protein beta/alpha-A (YWHABA) and cadherin-related family member 2 (Cdhr2). Besides, we found PSGs in highland alkaline tolerant species were also mainly involved in similar functional categories as overlapping PSGs (FIG. 4B, supplementary table S7), such as transmembrane protein 268 (TMEM268), transmembrane protein 266 (TMEM266), solute carrier family 35 member F4 (SLC35F4), solute carrier family 7, member 3 (SLC35F4) and solute carrier family 7, member 3 (SLC7A3) involved in ion transport or transmembrane functions, probable phospholipid-transporting ATPase VD (ATP10D) involved in energy metabolism, apoptosis-inducing factor 1 (AIFM1) involved in apoptosis, toll-like receptor 8 (TLR8) and interleukin-2 receptor subunit beta (IL2RB) related to immune response. Similarly, in lowland alkaline tolerant fish species, these PSGs were involved in four main categories (FIG. 4B, supplementary table S7), such as solute carrier family 2 member 15b (SLC2A15b), transmembrane protein 69 (TMEM69) and potassium voltage-gated channel subfamily E member 4 (KCNE4) associated with ion transport function, phosphoinositide 3-kinase regulatory subunit 4 (PIK3R4) and lysosomal-associated membrane protein 1 (LAMP1) associated with apoptosis, CD44 antigen (CD44) and immunoglobulin-like domain containing receptor 1b precursor (ILDR1) involved in immune response, and plasma membrane calcium-transporting ATPase 3 (ATP2B3) associated with energy metabolism.

Further, GO enrichment results showed that a number of significantly enriched GO terms ($P < 0.05$, fisher's exact test) are related to immune response, apoptosis, protein metabolisms for PGSs in highland alkaline tolerant species, such as inflammatory response (GO:0006954), natural killer cell activation (GO:0030101), cell killing (GO:0001906), muscle cell apoptotic process (GO:0010656), phagocytosis (GO:0006909), regulation of protein metabolism (GO:0051246) and negative regulation of macromolecule biosynthesis (GO:0010558) (FIG. 4C, supplementary table S8). Similarly, in lowland alkaline tolerant species, the significantly enriched GO terms were mainly related to four categories as well, such as T cell mediated cytotoxicity (GO:0001913) related to immune response, regulation of cation channel activity (GO:2001257) and regulation of transporter activity (GO:0032409) related to transport function, macroautophagy (GO:0016236) and regulation of apoptotic signaling pathway (GO:1902253) related to apoptosis (cell death), glucose metabolism (GO:0006006) and sulfur amino acid metabolism (GO:0000096) (FIG. 4D, supplementary table S8). Collectively, this finding suggested the consistent signature of positive selection in alkaline tolerant species within highland taxa and lowland taxa.

Discussion

Our results support above three hypotheses that alkaline tolerant species shared the consistent patterns of molecular evolution in protein-coding genes (dN/dS) and consistent signatures of accelerated evolution (rapidly evolving gene) and positive selection (positively selected gene) in highland and lowland fish. Specifically, these signatures include: genes experienced consistent acceleration in evolutionary rates (increased dN/dS) in alkaline tolerant species, which are mainly involved in ion transport, transmembrane and energy metabolism functions; genes showing consistent signals of positive selection in alkaline tolerant species within highland and lowland fish taxa, these are mainly associated with ion transport/transmembrane, apoptosis (cell death), immune response and energy metabolism processes. Altogether, this study provides insights in understanding the common role of adaptive molecular evolution in fish adaptation to alkaline environment, as well as adaptation to extreme environment at high-altitude.

Acid-base balance and osmoregulation

Alkaline tolerant species requires enhanced physiological abilities including acid-base balance and osmoregulation to respond the elevation in alkalinity or salinity of freshwater (Evans et al. 2005; Marshall 2005). Extremely alkaline environment may accelerate the evolution of genes associated with osmoregulation in these species survived in such harsh environment (Xu et al. 2017; Tong, Fei, et al. 2017; Reid et al. 2016). In this study, we identified a set of genes associated with ion transport and transmembrane that tended to evolve rapidly in alkaline tolerant species than their alkaline intolerant relatives (FIG. 3). This result echoed our previous finding in *G. przewalskii* compared with other teleost fishes (alkaline intolerance) (Tong, Fei, et al. 2017), also was in line with the rapidly evolving gene repertoire of a wild fish, Amur ide that survived in extremely alkaline environment (Xu et al. 2017). Interestingly, osmoregulation related genes including three solute carrier (SLC) genes and one transient receptor potential cation channel (TRPV) gene exhibited consistent signature of accelerated evolution in both highland and lowland alkaline tolerant fishes. *SLC* genes encode transmembrane transporters for inorganic ions, amino acids, neurotransmitters, sugars, purines and fatty acids, and other solute substrates (Dorwart et al. 2008). Recent evidences indicted that adaptive evolution of *SLC* genes contributed to fish adaptation to high salinity and high PH environment (Wang & Guo 2019; Tong, Fei, et al. 2017; Xu et al. 2017). Besides, we also identified numbers of ion transport and transmembrane genes under positive selection in alkaline tolerant fish species, such as potassium voltage-gated channel (*KCN*) genes. In killifish, an excellent model to study extreme environment adaptation, recent genome-wide studies also found a set of *KCN* genes under selection in natural populations that survived in toxic pollution environment (including high PH) (Reid et al. 2016; Osterberg et al. 2018). In addition, given that a set of same rapidly evolving or positively selected genes identified in both highland and lowland alkaline tolerant fish species, we found a large number of different genes under selection but involved in similar functions, such as anion transport (GO:0006820) and phosphate ion transport (GO:0006817), indicating the common role of adaptive molecular evolution during alkaline adaptation in fish.

Apoptosis and cell death

Extremely alkaline stress may cause extensive damage to fish, such as inducing cell apoptosis (Monteiro et al. 2009; Zhao et al. 2016, 2020). Given that, several fish species can survive in this harsh environment, such as schizothoracine (*Gymnocypris przewalskii*) (Tong, Fei, et al. 2017; Tong & Li 2020), tilapia (*Alcolapia grahami*) (Wilkie & Wood 1996), and Amur ide (*Leuciscus waleckii)* (Xu et al. 2017). This may also pose a barrier for alkaline tolerant fish as

compared with alkaline intolerant fish. Significantly, positively selected *AIFM1*, *SH3RF1*, *YWHABA*, *PIK3R4* and *LAMP1* of alkaline tolerant species relative to alkaline intolerant species have enrichment in a set of apoptosis pathways, such regulation of apoptotic signaling pathway (GO:1902253), macroautophagy (GO:0016236) and phagocytosis (GO:0006909) (FIG. 4). Apoptosis is a form of programmed cell death that occurs in multicellular organisms, it plays a significant role in the biochemical events lead to characteristic cell changes (morphology) and death (Green 2011). Few direct evidences suggested the roles of these candidate genes in response to extreme alkaline stress in fish, but numerous studies had defined their functions in tolerance to harsh environments. *AIFM1* is a ubiquitous mitochondrial oxidoreductase involved in apoptosis, involved in sea bream (*Sparus aurata*) response to acute environmental stress (Bermejo-Nogales et al. 2014). *YWHABA* (14-3-3 protein beta/alpha-A) is an important gene showing the ability to bind a multitude of functionally diverse signaling proteins, such as transmembrane receptors (Fu et al. 2000), that involved in spotted sea bass (*Lateolabrax maculatus*) tolerance to saline stress (Zhang et al. 2019). In addition, *LAMP1* plays an important role in lysosome biogenesis and autophagy (Eskelinen 2006), which involved in common carp (*Cyprinus carpio*) response to hydrogen peroxide environment. Collectively, the presence of apoptosis related genes under positive selection may contribute to the alkaline adaptation of fish, how these adaptive molecular changes affect the ability programmed cell death remains unknown.

Immune response

Extreme environments (including high PH) impact on the physiology of animals in a wide variety of ways. Recent advances in the understanding of environmental impacts were identified in relation to specific areas of immune function, such as increase in PH resulted in a general increase in immune function (Bowden 2008; Sridhar et al. 2020). Intriguingly, we identified different immune genes under positive selectin in alkaline tolerant species within highland (IL2RB, TLR8, IF144) and lowland (ILDR1 and CD44), but they all involved similar immune functions, such as inflammatory response (GO:0006954), natural killer cell activation (GO:0030101), and T cell mediated cytotoxicity (GO:0001913). In another word, we found different genes but conserved pathways may underlie fish adaptation to alkaline environment. TLR8, IF144, IL2RB and ILDR1 are important components of toll-like receptor (TLR) signaling pathway that play key roles in the innate immune system (Rebl et al. 2010). For instance, our previous comparative studies in alkaline tolerant fish, *G. przewalskii* identified key genes

involved in TLR pathway under selection (Tong, Fei, et al. 2017; Tong et al. 2015). In Nile tilapia (*Oreochromis niloticus*), another alkaline tolerant species and stands as ideal model to study extreme environment adaptation, a most recent comparative study identified numbers of immune genes involved in natural killer cell mediated cytotoxicity and NF-kappa B signaling pathway in response to alkalinity stress (up to PH = 8.9) (Zhao et al. 2020), and echoes our present results. Thus, it is possible that adaptive evolution changes (e.g. positive selection) acting on innate immune genes in alkaline tolerant species contribute to their adaptation to extremely alkaline environment.

Energy metabolism

It is not surprising that we identified a set of genes under either accelerated evolution or positive selection, that enriched in diverse metabolisms, such as glucose metabolism, phosphate metabolism, sulfur amino acid metabolism and protein metabolism in alkaline tolerant species compared with their alkaline intolerant relatives. In general, metabolism processes were involved in fish response to varies environmental stresses (including alkaline stress) by a huge amount of researches (Wood 1991). A most recent study highlights the genes involved in conserved mitochondrial pathways under selection in adaptation to extreme environment (Greenway et al. 2020). In addition to our previous studies in highland alkaline tolerant fish, we found a number of genes associated with energy metabolism processes under selection as well, such as mitochondrial function and protein metabolism (Tong, Fei, et al. 2017). Moreover, increasing studies on alkaline tolerant fish species (e.g. killifish) pointed out the roles of varied metabolisms in extreme environment (e.g. high PH) adaptation (Scott et al. 2019; Zhao et al. 2020; Wang & Guo 2019; Xu et al. 2017). Altogether, our analysis infer that the adaptive molecular evolution of metabolism associated genes may be indispensable and common features to alkaline adaptation as well as extreme environment adaptation in fish.

Conclusion

In summary, this comparative genomics study of 15 fish species suggests the common role of alkaline adaptation in fish, regardless of highland or lowland background environments. Our results also highlight that the adaptive evolution of protein-coding genes are likely to play a crucial role in fish response to extreme environment, such as extremely high PH. Notably, this

study provides putative genomic signatures of shift in selection and alkaline adaptation in several alkaline tolerant fish species, further study should include large scale of omics data of more alkaline tolerant fish species and their intolerant relatives as multiple pairs to demonstrate the genetic basis of alkaline adaptation in fish at genome-wide scale.

Author Contributions

C.T. and K.Z. conceived this project. C.T. designed the project. C.T., Y.T. and K.Z. collected the samples. C.T. and M.L. performed the comparative genomics analyses. C.T. wrote the paper. All authors read and approved the final manuscript.

Acknowledgments

We would like to thank Prof. Sergei L Kosakovsky Pond from Temple University for the suggestion in using HyPhy. We also thank three anonymous reviewers for helpful comments on this manuscript.

Funding

This work was supported by grants from the National Natural Science Foundation of China (31870365). Joint Grant from Chinese Academy of Sciences -People's Government of Qinghai Province on Sanjiangyuan National Park (LHZX-2020-01), China Biodiversity Observation and Research Network, Sino BON-Inland Water Fish Diversity Observation Network.

Conflicts of Interest

The authors declare no conflict of interest.

Data Availability Statement

The Illumina sequencing reads have been deposited at NCBI Sequence Read Archive under the NCBI BioProject accession PRJNA684806.

Literature Cited

Altenhoff AM et al. 2018. The OMA orthology database in 2018: retrieving evolutionary relationships among all domains of life through richer web and programmatic interfaces. Nucleic Acids Res. 46:D477–D485.

An Z, John K, Warren P, Stephen P. 2001. Evolution of Asian monsoons and phased uplift of the Himalaya–Tibetan plateau since Late Miocene times. Nature. 411:62–66.

Bermejo-Nogales A et al. 2014. Metabolic and transcriptional responses of gilthead sea bream (Sparus aurata L.) to environmental stress: new insights in fish mitochondrial phenotyping. Gen. Comp. Endocrinol. 205:305–315.

Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30:2114–2120.

Bowden TJ. 2008. Modulation of the immune system of fish by their environment. Fish Shellfish Immunol. 25:373–383.

Cao W, Chen Y, Wu Y, Zhu S, Others. 1981. Origin and evolution of schizothoracine fishes in relation to the upheaval of the Qinghai-Tibetan Plateau. Science Press: Beijng.

Dorwart MR, Shcheynikov N, Yang D, Muallem S. 2008. The solute carrier 26 family of proteins in epithelial ion transport. Physiology . 23:104–114.

Eskelinen E-L. 2006. Roles of LAMP-1 and LAMP-2 in lysosome biogenesis and autophagy. Mol. Aspects Med. 27:495–502.

Evans DH, Piermarini PM, Choe KP. 2005. The multifunctional fish gill: dominant site of gas exchange, osmoregulation, acid-base regulation, and excretion of nitrogenous waste. Physiol. Rev. 85:97–177.

Favre A et al. 2015. The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas. Biol. Rev. Camb. Philos. Soc. 90:236–253.

Fu H, Subramanian RR, Masters SC. 2000. 14-3-3 proteins: structure, function, and regulation. Annu. Rev. Pharmacol. Toxicol. 40:617–647.

Fu L, Niu B, Zhu Z, Wu S, Li W. 2012. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 28:3150–3152.

Grabherr MG et al. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat. Biotechnol. 29:644–652.

Green DR. 2011. Means to an end: apoptosis and other cell death mechanisms. Cold Spring Harbor Laboratory Press.

Greenway R et al. 2020. Convergent evolution of conserved mitochondrial pathways underlies repeated adaptation to extreme environments. Proc. Natl. Acad. Sci. U. S. A. 117:21822–21822.

Guan L, Chi W, Xiao W, Chen L, He S. 2014. Analysis of hypoxia-inducible factor alpha polyploidization reveals adaptation to Tibetan Plateau in the evolution of schizothoracine fish. BMC Evol. Biol. 14:192.

Johnson LK et al. 2020. Draft genome assemblies using sequencing reads from Oxford Nanopore Technology and Illumina platforms for four species of North American Fundulus killifish. Gigascience. 9. doi: 10.1093/gigascience/giaa067.

Kang J, Ma X, He S. 2017. Evidence of high-altitude adaptation in the glyptosternoid fish, Creteuchiloglanis macropterus from the Nujiang River obtained through transcriptome analysis. BMC Evol. Biol. 17:229.

Kosakovsky Pond SL, Poon AFY, et al. 2020. HyPhy 2.5—A Customizable Platform for Evolutionary Hypothesis Testing Using Phylogenies. Mol. Biol. Evol. 37:295–299.

Kosakovsky Pond SL, Wisotsky SR, Escalante A, Magalis BR, Weaver S. 2020. Contrast-FEL--a test for differences in selective pressures at individual sites among clades and sets of branches. Molecular Biology and Evoluton. https://academic.oup.com/mbe/advance-article-abstract/doi/10.1093/molbev/msaa263/5926108.

Levings CD. 2016. Ecology of Salmonids in Estuaries around the World: Adaptations, Habitats, and Conservation. UBC Press.

Lien S et al. 2016. The Atlantic salmon genome provides insights into rediploidization. Nature. 533:200–205.

Li J, Fang X. 1999. Uplift of the Tibetan Plateau and environmental changes. Chinese science

bulletin. 44:2117–2124.

Liu H-P et al. 2019. The sequence and de novo assembly of Oxygymnocypris stewartii genome. Scientific Data. 6:190009.

Marshall WS. 2005. Ion transport, osmoregulation, and acid-base balance. The physiology of fishes. CRC Press.

Monteiro SM, dos Santos NMS, Calejo M, Fontainhas-Fernandes A, Sousa M. 2009. Copper toxicity in gills of the teleost fish, Oreochromis niloticus: effects in apoptosis induction and cell proliferation. Aquat. Toxicol. 94:219–228.

Murrell B et al. 2015. Gene-wide identification of episodic selection. Mol. Biol. Evol. 32:1365–1371.

Orr HA. 2005. The genetic theory of adaptation: a brief history. Nat. Rev. Genet. 6:119–127.

Osterberg JS, Cammen KM, Schultz TF, Clark BW, Di Giulio RT. 2018. Genome-wide scan reveals signatures of selection related to pollution adaptation in non-model estuarine Atlantic killifish (Fundulus heteroclitus). Aquat. Toxicol. 200:73–82.

Oziolor EM et al. 2019. Adaptive introgression enables evolutionary rescue from extreme environmental pollution. Science. 364:455–457.

Qi D et al. 2012. Convergent, parallel and correlated evolution of trophic morphologies in the subfamily schizothoracinae from the Qinghai-Tibetan plateau. PLoS One. 7:e34070.

Qu Y et al. 2013. Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau. Nat. Commun. 4:2071.

Rebl A, Goldammer T, Seyfert H-M. 2010. Toll-like receptor signaling in bony fish. Vet. Immunol. Immunopathol. 134:139–150.

Reid NM et al. 2016. The genomic landscape of rapid repeated evolutionary adaptation to toxic pollution in wild fish. Science. 354:1305–1308.

Scheinfeldt LB, Tishkoff SA. 2010. Living the high life: high-altitude adaptation. Genome Biol. 11:133.

Schliep KP. 2011. phangorn: phylogenetic analysis in R. Bioinformatics. 27:592–593.

Scott WC et al. 2019. Influence of salinity and pH on bioconcentration of ionizable pharmaceuticals by the gulf killifish, Fundulus grandis. Chemosphere. 229:434–442.

Sridhar A et al. 2020. Activity profile of innate immune-related enzymes and bactericidal of freshwater fish epidermal mucus extract at different pH. Environ. Sci. Pollut. Res. Int. doi: 10.1007/s11356-020-11173-5.

Suyama M, Torrents D, Bork P. 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Res. 34:W609–12.

Tong C et al. 2015. Transcriptome-wide identification, molecular evolution and expression analysis of Toll-like receptor family in a Tibet fish, Gymnocypris przewalskii. Fish Shellfish Immunol. 46:334–345.

Tong C, Fei T, Zhang C, Zhao K. 2017. Comprehensive transcriptomic analysis of Tibetan Schizothoracinae fish Gymnocypris przewalskii reveals how it adapts to a high altitude aquatic life. BMC evolutionary biology. 17:74.

Tong C, Li M. 2020. Genomic signature of accelerated evolution in a saline-alkaline lake-dwelling Schizothoracine fish. Int. J. Biol. Macromol. 149:341–347.

Tong C, Najm GM, Pinter-Wollman N, Pruitt JN, Linksvayer TA. 2020. Comparative Genomics Identifies Putative Signatures of Sociality in Spiders. Genome Biol. Evol. 12:122–133.

Tong C, Tian F, Zhao K. 2017. Genomic signature of highland adaptation in fish: a case study in Tibetan Schizothoracinae species. BMC Genomics. 18:948.

Wang Y et al. 2015. Evidence for Adaptation to the Tibetan Plateau Inferred from Tibetan Loach Transcriptomes. Genome Biol. Evol. 7:2970–2982.

Wang Y, Guo B. 2019. Adaption to extreme environments: a perspective from fish genomics. Rev. Fish Biol. Fish. 29:735–747.

Wen LY. 2014. Uplift of the tibetan plateau influenced the morphological evolution of animals. J. Agric. Sci. 6:244.

Wilkie MP, Wood CM. 1996. The adaptations of fish to extremely alkaline environments. Comp.

Biochem. Physiol. B Biochem. Mol. Biol. 113:665–673.

Witt KE, Huerta-Sánchez E. 2019. Convergent evolution in human and domesticate adaptation to high-altitude environments. Philos. Trans. R. Soc. Lond. B Biol. Sci. 374:20180235.

Wood C et al. 1994. Urea production, acid-base regulation and their interactions in the Lake Magadi tilapia, a unique teleost adapted to a highly alkaline environment. Journal of Experimental Biology. 189:13–36.

Wood CM. 1991. Acid-Base and Ion Balance, Metabolism, and their Interactions, after Exhaustive Exercise in Fish. J. Exp. Biol. 160:285–308.

Wood CM, Bucking C, Grosell M. 2010. Acid--base responses to feeding and intestinal Cl--uptake in freshwater-and seawater-acclimated killifish, Fundulus heteroclitus, an agastric euryhaline teleost. J. Exp. Biol. 213:2681–2692.

Wu DD, Yang CP, Wang MS, Dong KZ. 2020. Convergent genomic signatures of high-altitude adaptation among domestic mammals. National Science Review. 7:952–963.

Wu Y. 1992. The Fishes of the Qinghai-Tibetan Plateau. Sichuan Publishing House of Science & Technology.

Xiao S et al. 2020. Genome of Tetraploid Fish Schizothorax o'connori Provides Insights into Early Re-diploidization and High-Altitude Adaptation. iScience. 23:101497.

Xu J et al. 2017. Genomic Basis of Adaptive Evolution: The Survival of Amur Ide (Leuciscus waleckii) in an Extremely Alkaline Environment. Mol. Biol. Evol. 34:145–159.

Xu Q et al. 2016. Analysis of the erythropoietin of a Tibetan Plateau schizothoracine fish (Gymnocypris dobula) reveals enhanced cytoprotection function in hypoxic environments. BMC Evol. Biol. 16:11.

Yang L, Wang Y, Zhang Z, He S. 2014. Comprehensive transcriptome analysis reveals accelerated genic evolution in a Tibet fish, Gymnodiptychus pachycheilus. Genome Biol. Evol. 7:251–261.

Zhang K-Q et al. 2019. 14-3-3 gene family in spotted sea bass (Lateolabrax maculatus): Genome-wide identification, phylogenetic analysis and expression profiles after salinity stress. Comp. Biochem. Physiol. A Mol. Integr. Physiol. 235:1–11.

Zhao Y et al. 2020. Transcriptome changes for Nile tilapia (Oreochromis niloticus) in response to alkalinity stress. Comp. Biochem. Physiol. Part D Genomics Proteomics. 33:100651.

Zhao Y, Wu J-W, Wang Y, Zhao J-L. 2016. Role of miR-21 in alkalinity stress tolerance in tilapia. Biochem. Biophys. Res. Commun. 471:26–33.

Zheng M. 1997. An Introduction to Saline Lakes on the Qinghai—Tibet Plateau. Springer Science & Business Media.

Zhou C et al. 2020. Comprehensive transcriptome data for endemic Schizothoracinae fish in the Tibetan Plateau. Sci Data. 7:28.

Zhu S, Wu Y. 1975. Study of fish fauna in Qinghai Lake. Science Press: Beijing.

Figure legend

FIG. 1. Overview of geographic distributions of two alkaline tolerant schizothoracine fish species and their habitats in the northeastern Tibetan Plateau. Map depicting the geographic distributions of *G. przewalskii* in Lake Qinghai, and *G. kelukehuensis* in Lake Keluke. Photos showing the representative specimens of *G. przewalskii* and *G. kelukehuensis*. Photo credit: Chao Tong and Kai Zhao.

FIG. 2. Consistent patterns of molecular evolution in alkaline tolerant fish species.
(A) Species tree was pruned from the Fish Tree of Life (https://fishtreeoflife.org/). Alkaline tolerant taxa and alkaline intolerant taxa are depicted in orange and sky blue, respectively.
(B) Number of tested orthologs (N = 5,748), and number of orthologs under consistent shifts in rates of molecular evolution in alkaline tolerant relative to alkaline intolerant species within all fish taxa, highland fish taxa and lowland fish taxa ($P < 0.05$, likelihood ratio test).

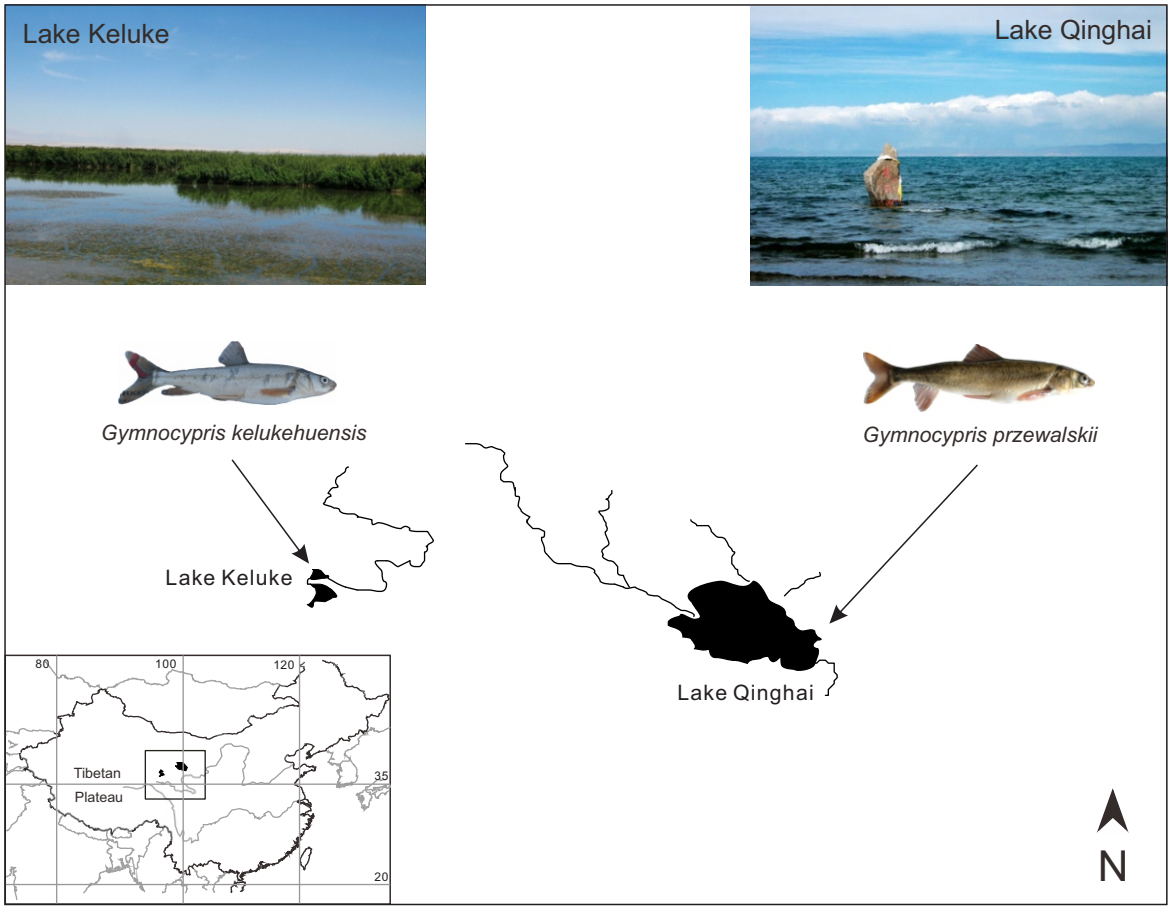
FIG. 3. Consistent signature of accelerated evolution in alkaline tolerant species within highland and lowland fish taxa.
(A) Venn diagram depicting numbers of rapidly evolving genes (REGs) in highland alkaline tolerant species, lowland alkaline tolerant species and both.
(B) Highland fish specific REGs, lowland fish specific REGs and overlapping REGs mainly related to ion transport, transmembrane and energy metabolism functions. The table showing the representative REGs under each category.
(C) REVIGO plot depicting the dominant enriched GO terms for REGs in highland alkaline tolerant fish. The scale of dark dot indicating the number of included enriched GO terms under a dominant GO term, the color scale representing the *P* value transformed by log10.
(D) REVIGO plot depicting the dominant enriched GO terms for REGs in lowland alkaline tolerant fish.

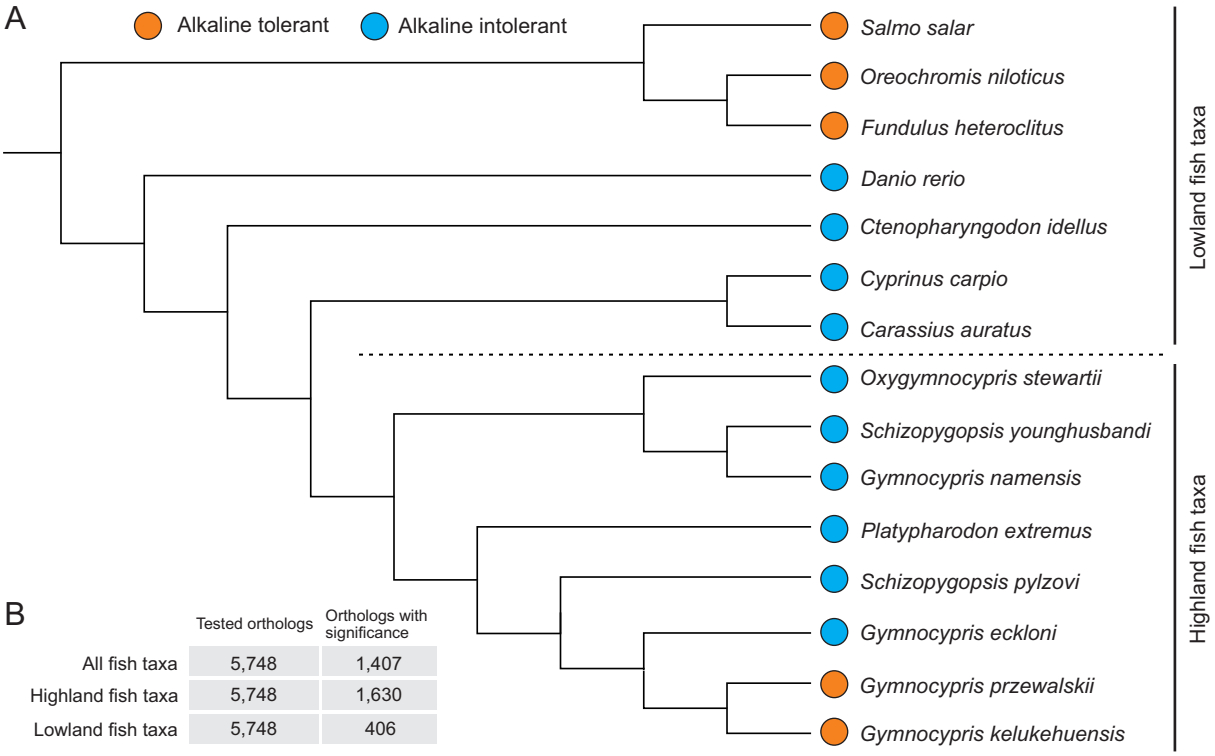
FIG. 4. Consistent signature of positive selection in alkaline tolerant species within highland and lowland fish taxa.
(A) Venn diagram depicting numbers of positively selected genes (PSGs) in highland alkaline tolerant species, lowland alkaline tolerant species and both.
(B) Highland fish specific PSGs, lowland fish specific PSGs and overlapping PSGs mainly

related to four categories, including ion transport/transmembrane, apoptosis/cell death, immune response and energy metabolism. The table showing the representative PSGs under each category.

(C) REVIGO plot depicting the dominant enriched GO terms for PSGs in highland alkaline tolerant fish. The dark dot scale indicating the number of included GO terms under a dominant GO term, the color scale representing the *P* value transformed by log10.

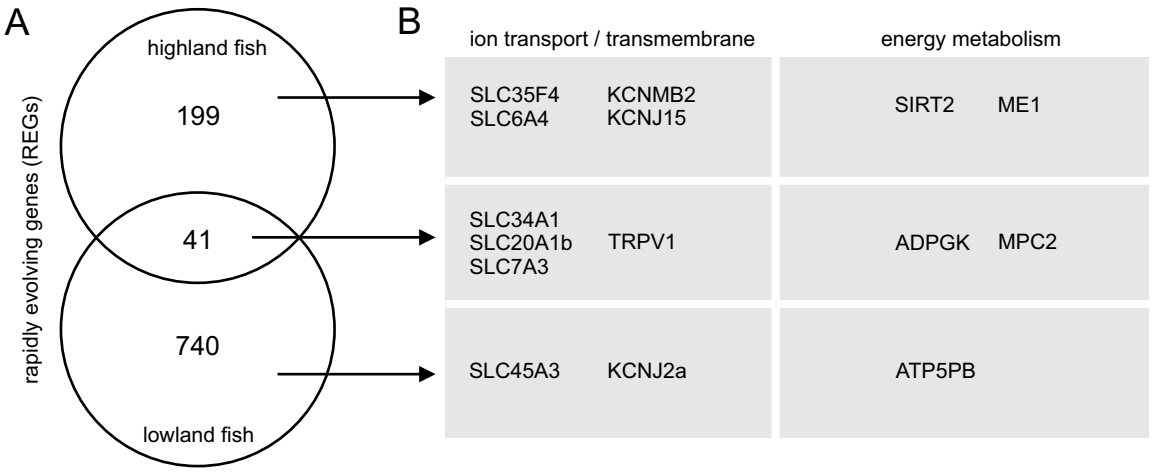
(D) REVIGO plot depicting the dominant enriched GO terms for PSGs in lowland alkaline tolerant fish.

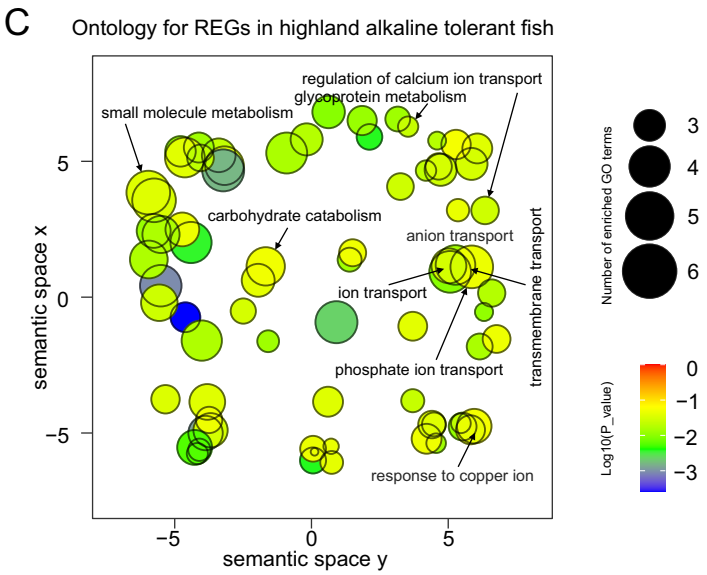

Lake Keluke
Lake Qinghai
Gymnocypris kelukehuensis
Gymnocypris przewalskii
Lake Keluke
Lake Qinghai
80
100
120
35
20
Tibetan
Plateau
N

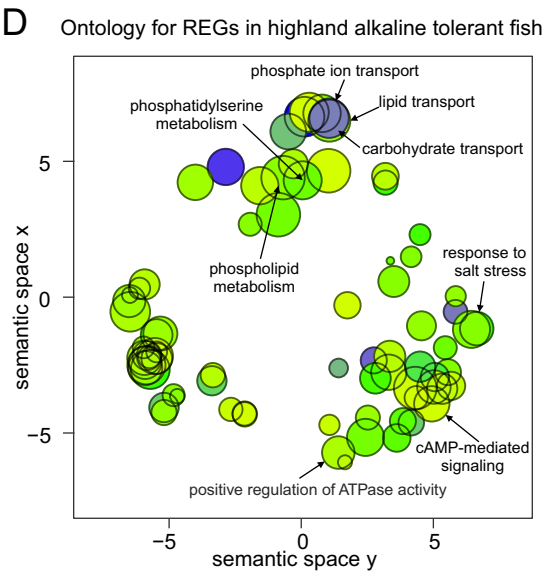


B

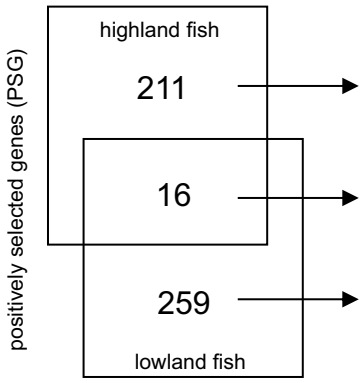
	Tested orthologs	Orthologs with significance
All fish taxa	5,748	1,407
Highland fish taxa	5,748	1,630
Lowland fish taxa	5,748	406


A
highland fish
199
41
740
lowland fish
rapidly evolving genes (REGs)
B
ion transport / transmembrane
energy metabolism
SLC35F4
SLC6A4
KCNMB2
KCNJ15
SIRT2
ME1
SLC34A1
SLC20A1b
SLC7A3
TRPV1
ADPGK
MPC2
SLC45A3
KCNJ2a
ATP5PB


C
Ontology for REGs in highland alkaline tolerant fish
regulation of calcium ion transport
glycoprotein metabolism
small molecule metabolism
carbohydrate catabolism
anion transport
ion transport
phosphate ion transport
transmembrane transport
response to copper ion
semantic space x
semantic space y
Number of enriched GO terms
3
4
5
6
Log10(P_value)
0
−1
−2
−3


D
Ontology for REGs in highland alkaline tolerant fish
phosphate ion transport
phosphatidylserine metabolism
lipid transport
carbohydrate transport
phospholipid metabolism
response to salt stress
cAMP-mediated signaling
positive regulation of ATPase activity
semantic space x
semantic space y

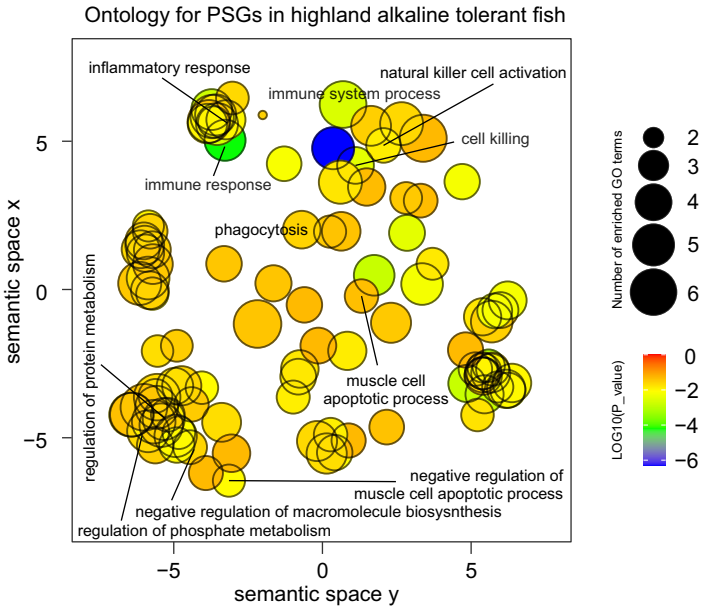
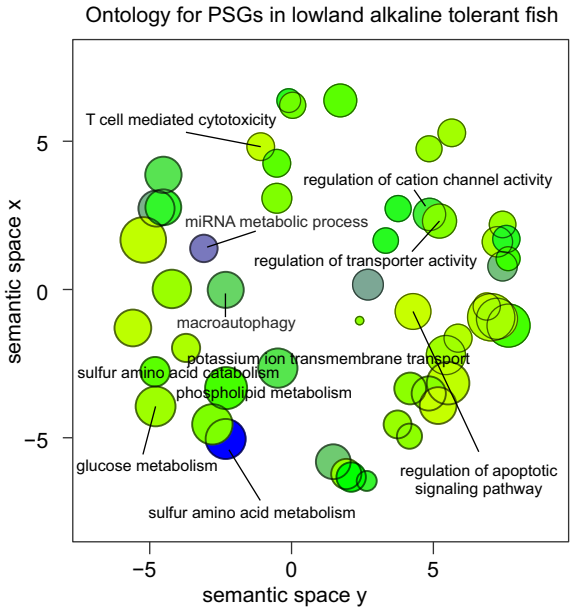
A



B

ion transport/transmenbrane	apoptosis cell death	immune response	energy metabolism
TMEM268 SLC6A4 TMEM266 SLC35F4 TMEM216 SLC7A3 CLDN15 SCN2B TRPV1 SLC13A5	AIFM1	IL2RB TLR8 IF144	ATP10D
PROC Cdhr2	SH3RF1 YWHABA		TPI1b
SLC2A15b SLC12A9 TMEM69 KCNE4	PIK3R4 LAMP1	ILDR1 CD44	ATP2B3

C



D