Genomic signature of accelerated evolution in a saline-alkaline lake-dwelling Schizothoracine fish

Chao Tong[1,*], Miao Li[2]

[1] Department of Biology, University of Pennsylvania, Philadelphia, PA, USA

[2] Center for Advanced Retinal and Ocular Therapeutics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA

* Corresponding authors: Chao Tong

Phone: +1-215-452-8712

Email: tongchao1990@gmail.com

ORCID: https://orcid.org/0000-0001-5202-5507

Miao Li: limiao1199@gmail.com

Abstract

Tibetan Plateau imposes extremely inhospitable environment on most wildlife. Besides the harsh aquatic environment including hypoxia and chronic cold, high salinity and alkalinity is an increasing threat to Tibetan endemic fishes. Previous genome-wide studies identified key genes contributed to highland fish adaptation to hypoxia and long-term cold, while our understanding of saline and alkaline adaptation in Tibetan fish remains limited. In this study, we performed a comparative genomics analysis in a saline lake-dwelling highland fish *Gymnocypris przewalskii*, aimed to identify candidate genes that contributed to saline and alkaline adaptation. We found elevated genome-wide rate of molecular evolution in *G. przewalskii* relative to lowland teleost fish species. In addition, we found nine genes encoding biological macromolecules associated with ion transport functions underwent accelerated evolution in *G. przewalskii*, which broadly expressed across kidney, gill, liver, spleen, brain and muscle tissues. Moreover, we found putative evidence of ion transport under selection were interacted by co-expression in *G. przewalskii* adaptation to high salinity and alkalinity environment of Lake Qinghai. Taken together, our comparative genomics study identified a set of rapidly evolving ion transport genes and transcriptomic signatures in Schizothoracine fish adaptation to saline and alkaline environment on the Tibetan Plateau.

Keywords: Comparative genomics, Schizothoracinae, Accelerated evolution, Saline and alkaline adaptation

1. Introduction

Environments may shape the genetic landscape of wildlife that inhabit them [1]. The world's largest and highest highland Tibetan Plateau had undergone continuous uplift during the India-Asia collision since about 45 million years ago, which triggered numerous environmental changes [2,3]. As elevation above sea level increases, a decrease in barometric pressure results in fewer oxygen molecules in the air, which causes hypoxia. Besides, other challenging environments high-altitude dwelling wildlife have encountered are the long-term low temperature and high ultraviolet radiation [4,5]. Understanding how organism adapt to their dwelling environment is central to answering many ecological and evolutionary questions, but it remains a formidable task to fully uncover the mechanisms of adaptive process [6]. Adaptation at the molecular level can occur by adaptive mutation in key genes over prolonged evolutionary time scales [7]. Recent genome-wide studies have identified key genes associated with hypoxia response and energy metabolism in Tibetan terrestrial animals adaptation to high altitude [8–10]. Nevertheless, the adaptive mechanism of Tibetan aquatic organisms to highland water environment is yet well-studied [11].

Unlike Tibetan terrestrial animal, the draft genomes of very few Tibetan aquatic organisms had been sequenced [12,13], the genomic basis of aquatic animals adaptation to water environments at high altitude still remain largely unknown. The Schizothoracine fishes are the predominant aquatic fauna on the Tibetan Plateau, which had evolved specific phenotypic characteristics to adapt to the harsh aquatic environments, such as hypoxia, chronic cold and high salinity and alkalinity. Comparative genomics approaches have the power to facilitate investigation of the genomic basis of evolution and adaptation [14]. Recent comparative genomics studies based on transcriptomic data of several Schizothoracine species have identified a number of candidate genes that underwent positive selection during the long-term adaptive processes to harsh environments on the Tibetan Plateau, such as BYSL and HSF1 associated with hypoxia response [15] and ND1, ATAD2 and ARL3 that involved into cold response [16,17]. Notably, an increasing number of lakes are existing or towards saline and alkaline due to the geological evolution and global climate changes on the Tibetan Plateau [3,18]. For instance, Lake Qinghai, the largest salt lake in China, is highly saline (up to 13‰) and alkaline (up to pH 9.4) water environment. It is also a typical salt lake with unusually high sodium, potassium and magnesium concentration [18,19]. Intriguingly, Lake Qinghai used to be freshwater and connected to the Yellow River, while was late separated with the upper reaches of the Yellow River during the geological event "Gonghe Movement" (approximately 15 mya) [19,20]. Moreover, the increasing of water salinization is a

67 growing threat to freshwater fish species [21,22]. Tibetan freshwater endemic fishes are long suffering
68 these harsh conditions challenges [11,18]. The main focus of the genetic mechanism of highland
69 adaptation in Tibetan fish are still on hypoxia and chronic cold response [15,23–25]. However, the
70 genomic signature of high salinity and alkalinity adaptation in Schizothoracine fish have yet to be
71 comprehensively determined.

72

73 Unlike other broadly distributed Schizothoracinae fish species, *Gymnocypris przewalskii* is only endemic
74 to Lake Qinghai [19,20,26]. Past studies suggested that *G. przewalskii* has gradually evolved from
75 freshwater fish to tolerate high salinity and alkalinity of Lake Qinghai during the early to late Holocene
76 [26]. Because of the unique evolutionary history in Lake Qinghai at high altitude, *G. przewalskii* provides
77 an exceptional model to investigate the genetic mechanisms underlying adaptation to high salinity and
78 alkalinity environment on the Tibetan Plateau. In this study, we performed a comparative genomics
79 analysis and identified a set of ion transport genes that showing strong signals of rapidly evolving in *G.*
80 *przewalskii*. Specifically, we used the *de novo* transcriptome assemblies from multiple tissue RNA-seq
81 data and five well-annotated teleost fish genomes for comparison. In addition, we estimated the genome-
82 wide nucleotide substitution rate of each fish species. Moreover, using the tissue-transcriptomics, we
83 characterized the expression patterns of rapidly evolving ion transport genes in kidney, gill, liver, spleen,
84 brain and muscle of highland fish, *G. przewalskii*.

85

86 ## 2. Materials and methods

87 ### *2.1. Data collection and transcriptome assembly*

88 We downloaded the transcriptome sequencing data of Schizothoracine fish *G. przewalskii* from NCBI
89 SRA database (https://www.ncbi.nlm.nih.gov/sra). Specifically, we collected six tissues transcriptomics
90 including kidney, gill, liver, spleen, brain and muscle of *G. przewalskii* (supplementary table S1). At first,
91 we checked the quality of the raw sequencing reads using FastQC v 0.11.8
92 (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/). Sequencing adapters and reads with a
93 quality score < 20 were trimmed with Trimmomatic v.0.36 [27], resulting in clean sequencing reads.
94 Then, we performed *de novo* transcriptome assembly using Trinity v2.8.5 [28] with default parameters.
95 After assembly, the redundant transcripts were removed using CD-HIT v4.8.1 [29] with the threshold of
96 0.90, and only the longest transcript under each cluster was extracted as unigene (unique gene). Next, we
97 predicted the open reading frame (ORF) of each unigene using TransDecoder

(https://github.com/TransDecoder/TransDecoder). Finally, we translated the nucleotide sequences of protein-coding genes from the assemblies of *G. przewalskii* into amino acid sequences using an in-house-developed perl script.

2.2. Orthologs identification and sequence alignment

We included five well-annotated teleost fish genomes for comparative genomics analysis and downloaded from Ensembl database (http://useast.ensembl.org/index.html), including zebrafish (*Danio rerio*), tilapia (*Oreochromis niloticus*), medaka (*Oryzias latipes*), fugu (*Takifugu rubripes*) and cod (*Gadus morhua*). Then, we built a local protein database including the sequences from above five fish genomes and *G. przewalskii* transcriptome assemblies. Next, we downloaded the curated orthology map of Actinopterygii (ray-finned fish) from OrthoDB database (release 8) [30] which contains 21,952 orthologous gene groups information. Of these seed orthologous groups in HaMStR v13.2.6 [31], we identified the orthologs in each fish species with E-values of less than 10^{-20}. Moreover, we aligned and trimmed the protein sequences of the orthologous groups using PRANK [32] and MATFF v 7.450 [33], and trimmed using trimAl [34] with the parameter "-automated1". Among the identified orthologs, we identified one-to-one, one-to-many and many-to-many orthologs in each fish species. For each 1:1 orthologous pair (i.e. genes for which only one gene from each species matches the given OrthoDB orthologous gene group), we only selected the longest transcript associated with the gene for each pair of species as putative single-copy ortholog. Finally, we identified the core single-copy orthologs that were shared by above six fish species.

2.3. Genome-scale concatenation and coalescent based dataset construction

We performed the alignment of core shared single-copy orthologs of six fish species using MUSCLE [35] with default parameters and trimmed using trimAl v1.2 [34] with parameter "-automated1". In addition, we filtered the core shared single-copy orthologs with strict constraints, including length (minimum 200 aa) and sequence alignment (maximum missing data 50% in alignments). Next, we prepared two types of datasets after filtration. Firstly, we concatenated all core shared single-copy genes from each species into one-line sequence as a supergene using a custom python script (genome-scale concatenation-based dataset), respectively. Secondly, we conducted a genome-scale coalescent-based dataset including core shared single-copy genes from each species.

2.4. Molecular evolution analysis

We used the clipped species tree (Figure 1A) including above six fish species from a previous study [16]. To estimate lineage-specific evolutionary rates for each fish species, we aligned core shared single-copy orthologs using MUSCLE [35], derived nucleotide alignments from protein alignments using PAL2NAL v14 [36], and estimated pairwise non-synonymous to synonymous substitutions (dN/dS) of nucleotide alignments using the CodeML package in PAML 4.7a [37]. Specifically, we used the free-ratio model ("several ω ratio") to calculate the ratio of dN to dS nucleotide changes separately for each ortholog and a concatenation of all alignments of single-copy orthologs from above six fish species. Parameters, including dN, dS, dN/dS, N*dN, and S*dS values, were estimated for each branch, and genes were discarded if N*dN or S*dS < 1, or dS > 2, following previous studies [11,16,17].

We sought to identify a set of genes with elevated dN/dS in *G. przewalskii* relative to other five fish species. At first, we ran two branch models using CodeML package in PAML 4.7a [37] to identify rapidly evolving genes (REGs) in *G. przewalskii* lineage with corresponding nucleotide alignments, specifically with the null model assuming that all branches have been evolving at the same rate and the alternative model allowing the focal foreground branch (*G. przewalskii*) to evolve under a different evolutionary rate. Next, we used a likelihood ratio test (LRT) in R software, package MASS with df = 1 to discriminate between the alternative model and the null model for each single-copy orthologs in the genesets. We only considered the genes as rapidly evolving with a significantly faster rate in the foreground branch if the adjusted *P* value < 0.05 and higher dN/dS in the focal foreground branch than focal background branches (other four fish species). Finally, we annotated the rapidly evolving genes with gene ontology (GO) function category using R software, package topGO [38].

2.5. *Gene expression analysis*

After preparation of clean reads from six tissue-transcriptomics (kidney, gill, liver, spleen, brain and muscle) of *G. przewalskii*, we performed the transcript quantification by mapping all clean reads to the transcriptome assemblies using RSEM v1.3.1 [39] to obtain expected counts and fragments per kilobase million (FPKM). In addition, we primarily focused on the expression pattern of rapidly evolving genes (REGs), and calculated the FPKM value of each REG in each tissue. At last, we annotated the differentially expressed REGs by gene ontology using R software, package TopGO [38].

3. Results

3.1. Transcriptome assemblies and orthologs

By pooling six tissue-transcriptome sequencing data, the de novo transcriptome assembly of *G. przewalskii* yielded 409,685 transcripts, with an N50 of 1,796 bp and an average length of 986 bp. After removing redundant isoforms and extraction of longest isoform among alternative transcripts, a total of 357,601 unigenes were obtained, with an N50 of 3,079 bp and a mean length of 1,992 bp. After protein-coding gene prediction with TransDecoder, we totally obtained 137,539 unigenes with full or partial length of gene coding regions (CDS) in *G. przewalskii*.

After identification of orthologs according to the curated orthologous gene groups of Actinopterygii in each fish species, we obtained a total of 74,107 putative orthologs in 16,379 orthologous gene groups (Table 1). After strict 1:1 ortholog selection, we identificated 16,379 longest orthologs that represent their gene groups as unique ortholog (Table 1). In addition, we eventually obtained core 10,260 orthologs that shared by all six fish species, making them suitable for comparative genomics analysis.

3.2. Genome-wide nucleotide substitution rate

After estimation of the nucleotide substitution rates of each branch that represented each fish species based on 6,742 core shared single-copy orthologs, we found that Schizothoracine fish *G. przewalskii* had elevated terminal genome-wide concatenation-based dN/dS compared to other five fish species (Figure 1B). Furthermore, we also found similar elevated pattern of genome-wide coalescent-based dN/dS in *G. przewalskii* relative to other species (Figure 1C).

3.3. Rapidly evolving genes

A set of genes with the signature of an increase rate of non-synonymous changes and underwent accelerated evolution, namely rapidly evolving genes. We identified 466 putative rapidly evolving single-copy orthologs (REGs) in *G. przewalskii* (supplementary table S2). Among this set of genes, the most interesting finding was REGs included genes associated with ion transport functions. This group included sodium channel subunit beta-3 (SCN3B), solute carrier family 13 member 3 (SLC13A3), transmembrane protein 175 (TMEM175) and $H(^{+})/Cl(^{-})$ exchange transporter 7 (CLCN7) (Table 2). Moreover, we found a number of REGs associated with mitochondrial function and also involved ion transport process, such as sodium/potassium-transporting ATPase subunit beta-2 (ATP1B2), calcium uniporter protein (MCU)

and calcium uptake protein 2 (MICU2) (Table 2). Besides the ion transport genes, we found a large number of genes involved energy metabolism function, such as ATP5c1 and ATP5b associated with ATP binding and oxidative phosphorylation process (supplementary table S2). Although previous comparative genomics studies with highland fish identified several candidate genes with the signals of positive selection [15,23,24], here, we failed to identify any gene that potentially associated with hypoxia response.

3.4. Tissue expression patterns of rapidly evolving ion transport genes

After mapping the clean reads from six tissue-transcriptome sequencing data, we obtained the expression level of each unigenes based on FPKM value (supplementary table S3). We focused on the expression pattern of ion transport genes with the signature of accelerated evolution. Notably, we found eight rapidly evolving ion transport genes were broadly expressed in six tissues, except *transient receptor potential cation channel subfamily V member 6* (*TRPV6*) that only expressed in liver (Figure 2A). In addition, the hierarchical clustering which illustrated by heatmap indicated that four genes (*ATP1B2*, *MCU*, *CLCN7* and *NALCN*) and another five genes (*MICU2*, *SCN3B*, *TMEM175*, *TRPV6* and *SLC13A3*) showed similar tissues expression patterns, respectively (Figure 2B). Moreover, gene ontology (GO) enrichment analysis showed that this set of differentially expressed REGs was significantly enriched multiple functions, such as ion transport (GO:0006811, $P = 0.00031$), sodium ion transport (GO:0006814, $P = 0.00047$), calcium ion transport (GO:0006816, $P = 0.00056$), chloride transport (GO:0006821, $P = 0.00067$), response to pH (GO:0009268, $P = 0.00069$) and response to calcium ion (GO:0051592, $P = 0.00078$) (Figure 2C).

4. Discussion

Over the past few years, transcriptome-based assembly approach enables comparative genomics studies widely employed in many Tibetan endemic organisms to provide insights of highland adaptation [15–17,23,40,41]. Unlike whole genome data, although transcriptome sequencing is an effective and accessible approach to initiate comparative genomic analyses on non-model organisms [28], it still can not cover the whole protein coding gene repertoire of one species. Previous transcriptome studies on Tibetan fishes mainly included one or two tissues [17,40,41], our present study included six tissues (kidney, gill, liver, spleen, brain and muscle) RNA-seq data of *G. przewalskii* and generated much more transcripts than previously assemblies [16,40]. In addition, we used curated orthology mapping approach [42] and identified more than 15,000 pairwise orthologous genes in each fish species and over 10,000

core single-copy orthologs shared by six species, which included much more orthologs than our early studies [16,17]. These putative single-copy orthologs are the important bases for comparative genomic analysis. Notably, most high-altitude dwelling Schizothoracine fish species are polyploidy with high complexity and large size of genomes, the whole genome data is long lacking [11]. Therefore, comparative genomic analysis based on transcriptome assemblies of Schizothoracine fish will still be the tendency in recent years.

Our present study pinpointed that highland fish, *G. przewalskii* has elevated rate of molecular evolution (dN/dS) on both concatenation and coalescent genomic-scales compared with lowland fish species, indicating that *G. przewalskii* may be under rapidly evolving. Not surprisingly, this result was consistent with previous studies in other Tibetan endemic fish species [15–17,23,41]. In addition, this finding highlighted animals endemic to the Tibetan Plateau underwent accelerated evolution (high dN/dS) relative to low-altitude dwelling organisms [9,10]. Furthermore, species inhabit similar ecological niches may be shaped by convergent evolution to form physiological or morphological similarities [43]. Like other Tibetan terrestrial wildlife, our finding implied that the elevation of genome-wide nucleotide substitution rate is one of adaptive process of *G. przewalskii* to harsh environment in Lake Qinghai, including the increasing of water salinization.

Accelerated evolution at molecular level may be reflected by an increased rate of non-synonymous changes within genes involved in adaptation [44]. Our present comparative study identifies a set of rapidly evolving genes associated with ion transport function in *G. przewalskii*. These genes encoded biological macromolecules which mainly functioning in sodium ion transport, calcuim ion transport, chloride transport and response to pH processes. This result is consistent with findings in an extremely alkaline environment dwelling fish, *Leuciscus waleckii* [45], indicating that the alkaline environment of both Lake Qinghai and Lake Dali Nur spurred accelerated evolution of ion transport genes in both fish species. Notably, the rapidly evolving gene repertoire of *G. przewalskii* included SLC13A3, TMEM175 and CLCN7. Solute carrier (SLC) is a family that encoded transmembrane transporters for inorganic ions, amino acids, neurotransmitters, sugars, purines and fatty acids, and other solute substrates [46]. Past evidence suggested that the adaptive evolution of solute carrier genes contribute to high salinity and alkalinity adaptation in fishes [45–47]. Specifically, SLC13 is a subfamily of sodium sulphate/carboxylate cotransporters [48]. Moreover, CLC gene is a family of H^+/Cl^- exchange transporter that mediate transmembrane Cl− transport [49]. In addition, previous study suggested that TMEM175 is involved in

253 potassium channel activity [50]. Therefore, we suggested that ion transport genes underwent rapidly
254 evolving is another adaptation strategy for *G. przewalskii* to cope with the severe saline and alkaline
255 stress in Lake Qinghai.

256

257 Previous studies identified a number of genes under accelerated evolution that mainly involved energy
258 metabolism pathways [11,15-17,23-25]. Compared with a few rapidly evolving ion transport genes that
259 were found in *G. przewalskii*, this present study identified a number of candidate genes that related to
260 energy metabolism and contributed to long-term cold adaptation. Gene associated with energy
261 metabolism showing signs of adaptive evolution is one of the genomic signatures that had been identified
262 in Tibetan Plateau dwelling animals [16,17]. Our finding is consistent with previous comparative
263 genomics studies in highland fishes as well [15-17]. A set of genes functioning in energy supply and
264 metabolism were under accelerated evolution in *G. przewalskii*, such as ATP5b and ATP5c, ATP
265 synthase subunit beta. In addition, although hypoxia adaptation is one of the significant adaptive
266 processes contributed to highland adaptation in endemic animals that dwelt at high altitude [10,16,17], we
267 still were not able to identify any rapidly evolving genes associated with hypoxia response function in
268 present study. Indeed, there is a long controversial issue about hypoxic environment and hypoxia response
269 for Tibetan fish species. Obviously, more physiological, ecological and genomic analysis were required to
270 reveal the mechanism of highland fish adaptation to hypoxia.

271

272 A set of previous studies indicated that fish gills, kidney, liver and spleen are key tissues that contributed
273 to saline and alkaline tolerance [51,52]. Using tissue-transcriptomic data, we characterized the expression
274 profiles of six tissue types. Most of rapidly evolving ion transport genes have broad expression patterns
275 across all tissues. In addition, these broadly expressed ion transport genes were mainly associated with
276 sodium ion transport, chloride transport and response to pH function by gene ontology annotation. This
277 finding indicates that ion transport genes in *G. przewalskii* experiencing accelerated evolution may have
278 general functions and involve into multiple biological processes. Furthermore, we found a set of rapidly
279 evolving ion transport genes that involved distinct pathways showed the similar tissue expression
280 patterns. That is said, these ion transport genes under selection were putatively interacted by cooperation
281 in *G. przewalskii* adaptation to high salinity and alkalinity environment of Lake Qinghai. Therefore, this
282 finding indicated that future Schizothoracine fish comparative genomics study, including increasing
283 sequencing and function assay, can further clarity the molecular basis of saline and alkaline adaptation of
284 high-altitude dwelling fishes.

5. Conclusion

We used comparative genomics based on the *de novo* assemblies from pooled six tissues transcriptomes to identify the genomic signature of saline and alkaline adaptation in a highland fish, *G. przewalskii*. These putative genomic signatures included: (1) Schizothoracine fish, *G. przewalskii* had an elevated genome-wide nucleotide substitution rate than lowland teleost fish species; (2) a number of genes associated with ion transport and energy metabolism functions were found in *G. przewalskii* with elevated molecular evolutionary rate (dN/dS) showing the signature of rapidly evolving; (3) most of rapidly evolving ion transport genes associated with sodium ion transport, calcium ion transport and chloride transport were broadly expressed in kidney, gill, liver, spleen, brain and muscle of *G. przewalskii*; (4) A set of rapidly evolving ion transport genes exhibited similar tissue expression patterns and were interacted by co-expression in *G. przewalskii*. Altogether, our present study will provide the genomic signatures of rapidly evolving ion transport genes, and gain the insights into the saline and alkaline adaptation of high-altitude dwelling fishes.

Acknowledgments

This work was funded by the Field Research Fund of University of Pennsylvania Biology Department.

References

[1] K.E. Witt, E. Huerta-Sánchez, Convergent evolution in human and domesticate adaptation to high-altitude environments, Philos. Trans. R. Soc. Lond. B Biol. Sci. 374 (2019) 20180235.

[2] J. Li, X. Fang, Uplift of the Tibetan Plateau and environmental changes, Chinese Science Bulletin. 44 (1999) 2117–2124.

[3] A. Favre, M. Päckert, S.U. Pauls, S.C. Jähnig, D. Uhl, I. Michalak, A.N. Muellner-Riehl, The role of the uplift of the Qinghai-Tibetan Plateau for the evolution of Tibetan biotas, Biol. Rev. Camb. Philos. Soc. 90 (2015) 236–253.

[4] L.B. Scheinfeldt, S.A. Tishkoff, Living the high life: high-altitude adaptation, Genome Biol. 11 (2010) 133.

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

[6] N.G.C. Smith, A. Eyre-Walker, Adaptive protein evolution in Drosophila, Nature. 415 (2002) 1022–1024.

[7] H.A. Orr, The genetic theory of adaptation: a brief history, Nat. Rev. Genet. 6 (2005) 119–127.

[8] J.-T. Li, Y.-D. Gao, L. Xie, C. Deng, P. Shi, M.-L. Guan, S. Huang, J.-L. Ren, D.-D. Wu, L. Ding, Z.-Y. Huang, H. Nie, D.P. Humphreys, D.M. Hillis, W.-Z. Wang, Y.-P. Zhang, Comparative genomic investigation of high-elevation adaptation in ectothermic snakes, Proc. Natl. Acad. Sci. U. S. A. 115 (2018) 8406–8411.

[9] Y. Qu, H. Zhao, N. Han, G. Zhou, G. Song, B. Gao, S. Tian, J. Zhang, R. Zhang, X. Meng, Y. Zhang, Y. Zhang, X. Zhu, W. Wang, D. Lambert, P.G.P. Ericson, S. Subramanian, C. Yeung, H. Zhu, Z. Jiang, R. Li, F. Lei, Ground tit genome reveals avian adaptation to living at high altitudes in the Tibetan plateau, Nat. Commun. 4 (2013) 2071.

[10] Q. Qiu, G. Zhang, T. Ma, W. Qian, J. Wang, Z. Ye, C. Cao, Q. Hu, J. Kim, D.M. Larkin, L. Auvil, B. Capitanu, J. Ma, H.A. Lewin, X. Qian, Y. Lang, R. Zhou, L. Wang, K. Wang, J. Xia, S. Liao, S. Pan, X. Lu, H. Hou, Y. Wang, X. Zang, Y. Yin, H. Ma, J. Zhang, Z. Wang, Y. Zhang, D. Zhang, T. Yonezawa, M. Hasegawa, Y. Zhong, W. Liu, Y. Zhang, Z. Huang, S. Zhang, R. Long, H. Yang, J. Wang, J.A. Lenstra, D.N. Cooper, Y. Wu, J. Wang, P. Shi, J. Wang, J. Liu, The yak genome and adaptation to life at high altitude, Nat. Genet. 44 (2012) 946–949.

[11] C. Tong, M. Li, Y. Tang, K. Zhao, Genomic signature of ongoing alkaline adaptation in a Schizothoracine fish (Cyprinidae) inhabiting soda lake on the Tibetan Plateau, bioRxiv. (2019) 813501. doi:10.1101/813501.

[12] H.-P. Liu, S.-J. Xiao, N. Wu, D. Wang, Y.-C. Liu, C.-W. Zhou, Q.-Y. Liu, R.-B. Yang, W.-K. Jiang, Q.-Q. Liang, Wangjiu, C. Zhang, J.-H. Gong, X.-H. Yuan, Z.-B. Mou, The sequence and de novo assembly of *Oxygymnocypris stewartii* genome, Sci Data. 6 (2019) 190009.

[13] X. Yang, H. Liu, Z. Ma, Y. Zou, M. Zou, Y. Mao, X. Li, H. Wang, T. Chen, W. Wang, R. Yang, The chromosome□level genome assembly of Triplophysa tibetana , a fish adapted to the harsh high□altitude environment of the Tibetan plateau, Molecular Ecology Resources. 19 (2019) 1027–1036.

342 [14] J. Stapley, J. Reger, P.G.D. Feulner, C. Smadja, J. Galindo, R. Ekblom, C. Bennison, A.D. Ball, A.P.
343 Beckerman, J. Slate, Adaptation genomics: the next generation, Trends Ecol. Evol. 25 (2010) 705–
344 712.

345 [15] L. Yang, Y. Wang, Z. Zhang, S. He, Comprehensive transcriptome analysis reveals accelerated genic
346 evolution in a Tibet fish, *Gymnodiptychus pachycheilus*, Genome Biol. Evol. 7 (2014) 251–261.

347 [16] C. Tong, T. Fei, C. Zhang, K. Zhao, Comprehensive transcriptomic analysis of Tibetan
348 Schizothoracinae fish *Gymnocypris przewalskii* reveals how it adapts to a high altitude aquatic life,
349 BMC Evolutionary Biology. 17 (2017) 74.

350 [17] C. Tong, F. Tian, K. Zhao, Genomic signature of highland adaptation in fish: a case study in Tibetan
351 Schizothoracinae species, BMC Genomics. 18 (2017) 948.

352 [18] M. Zheng, An Introduction to Saline Lakes on the Qinghai—Tibet Plateau, Springer Science &
353 Business Media, 1997.

354 [19] S. Zhu, Y. Wu, Study of fish fauna in Qinghai Lake, Fish Fauna and Biology of Naked Carp in
355 Qinghai Lake. (1975) 9–26.

356 [20] Y. Wu, The Fishes of the Qinghai-Tibetan Plateau, Sichuan Publishing House of Science &
357 Technology, 1992.

358 [21] E.R. Herbert, P. Boon, A.J. Burgin, S.C. Neubauer, R.B. Franklin, M. Ardón, K.N. Hopfensperger,
359 L.P.M. Lamers, P. Gell, A global perspective on wetland salinization: ecological consequences of a
360 growing threat to freshwater wetlands, Ecosphere. 6 (2015) art206.

361 [22] W.D. Hintz, R.A. Relyea, A salty landscape of fear: responses of fish and zooplankton to freshwater
362 salinization and predatory stress, Oecologia. 185 (2017) 147–156.

363 [23] Y. Wang, L. Yang, K. Zhou, Y. Zhang, Z. Song, S. He, Evidence for Adaptation to the Tibetan
364 Plateau Inferred from Tibetan Loach Transcriptomes, Genome Biol. Evol. 7 (2015) 2970–2982.

365 [24] L. Guan, W. Chi, W. Xiao, L. Chen, S. He, Analysis of hypoxia-inducible factor alpha
366 polyploidization reveals adaptation to Tibetan Plateau in the evolution of schizothoracine fish, BMC
367 Evol. Biol. 14 (2014) 192.

368 [25] Q. Xu, C. Zhang, D. Zhang, H. Jiang, S. Peng, Y. Liu, K. Zhao, C. Wang, L. Chen, Analysis of the
369 erythropoietin of a Tibetan Plateau schizothoracine fish (Gymnocypris dobula) reveals enhanced

cytoprotection function in hypoxic environments, BMC Evol. Biol. 16 (2016) 11.

[26] K. Zhao, Z. Duan, G. Yang, Z. Peng, S. He, Y. Chen, Origin of Gymnocypris przewalskii and phylogenetic history of *Gymnocypris eckloni* (Teleostei: Cyprinidae), Prog. Nat. Sci. 17 (2007) 520–528.

[27] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics. 30 (2014) 2114–2120.

[28] M.G. Grabherr, B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, X. Adiconis, L. Fan, R. Raychowdhury, Q. Zeng, Z. Chen, E. Mauceli, N. Hacohen, A. Gnirke, N. Rhind, F. di Palma, B.W. Birren, C. Nusbaum, K. Lindblad-Toh, N. Friedman, A. Regev, Full-length transcriptome assembly from RNA-Seq data without a reference genome, Nat. Biotechnol. 29 (2011) 644–652.

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

[30] E.V. Kriventseva, F. Tegenfeldt, T.J. Petty, R.M. Waterhouse, F.A. Simão, I.A. Pozdnyakov, P. Ioannidis, E.M. Zdobnov, OrthoDB v8: update of the hierarchical catalog of orthologs and the underlying free software, Nucleic Acids Res. 43 (2015) D250–6.

[31] I. Ebersberger, S. Strauss, A. von Haeseler, HaMStR: profile hidden markov model based search for orthologs in ESTs, BMC Evol. Biol. 9 (2009) 157.

[32] A. Löytynoja, N. Goldman, An algorithm for progressive multiple alignment of sequences with insertions, Proc. Natl. Acad. Sci. U. S. A. 102 (2005) 10557–10562.

[33] K. Katoh, D.M. Standley, MAFFT multiple sequence alignment software version 7: improvements in performance and usability, Mol. Biol. Evol. 30 (2013) 772–780.

[34] S. Capella-Gutiérrez, J.M. Silla-Martínez, T. Gabaldón, trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses, Bioinformatics. 25 (2009) 1972–1973.

[35] R.C. Edgar, MUSCLE: a multiple sequence alignment method with reduced time and space complexity, BMC Bioinformatics. 5 (2004) 113.

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

[37] Z. Yang, PAML 4: phylogenetic analysis by maximum likelihood, Mol. Biol. Evol. 24 (2007) 1586–1591.

[38] A. Alexa, J. Rahnenfuhrer, topGO: enrichment analysis for gene ontology, R Package Version. 2 (2010) 2010.

[39] B. Li, C.N. Dewey, RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome, BMC Bioinformatics. 12 (2011) 323.

[40] R. Zhang, A. Ludwig, C. Zhang, C. Tong, G. Li, Y. Tang, Z. Peng, K. Zhao, Local adaptation of Gymnocypris przewalskii (Cyprinidae) on the Tibetan Plateau, Scientific Reports. 5 (2015) 9780.

[41] X. Ma, W. Dai, J. Kang, L. Yang, S. He, Comprehensive Transcriptome Analysis of Six Catfish Species from an Altitude Gradient Reveals Adaptive Evolution in Tibetan Fishes, G3 . 6 (2015) 141–148.

[42] C. Tong, G.M. Najm, N. Pinter-Wollman, J.N. Pruitt, T.A. Linksvayer, Comparative genomics identifies putative signatures of sociality in spiders, bioRxiv. (2019) 797241. doi:10.1101/797241.

[43] D.L. Stern, The genetic causes of convergent evolution, Nat. Rev. Genet. 14 (2013) 751–764.

[44] G. Zhang, C. Li, Q. Li, B. Li, D.M. Larkin, C. Lee, J.F. Storz, A. Antunes, M.J. Greenwold, R.W. Meredith, A. Ödeen, J. Cui, Q. Zhou, L. Xu, H. Pan, Z. Wang, L. Jin, P. Zhang, H. Hu, W. Yang, J. Hu, J. Xiao, Z. Yang, Y. Liu, Q. Xie, H. Yu, J. Lian, P. Wen, F. Zhang, H. Li, Y. Zeng, Z. Xiong, S. Liu, L. Zhou, Z. Huang, N. An, J. Wang, Q. Zheng, Y. Xiong, G. Wang, B. Wang, J. Wang, Y. Fan, R.R. da Fonseca, A. Alfaro-Núñez, M. Schubert, L. Orlando, T. Mourier, J.T. Howard, G. Ganapathy, A. Pfenning, O. Whitney, M.V. Rivas, E. Hara, J. Smith, M. Farré, J. Narayan, G. Slavov, M.N. Romanov, R. Borges, J.P. Machado, I. Khan, M.S. Springer, J. Gatesy, F.G. Hoffmann, J.C. Opazo, O. Håstad, R.H. Sawyer, H. Kim, K.-W. Kim, H.J. Kim, S. Cho, N. Li, Y. Huang, M.W. Bruford, X. Zhan, A. Dixon, M.F. Bertelsen, E. Derryberry, W. Warren, R.K. Wilson, S. Li, D.A. Ray, R.E. Green, S.J. O'Brien, D. Griffin, W.E. Johnson, D. Haussler, O.A. Ryder, E. Willerslev, G.R. Graves, P. Alström, J. Fjeldså, D.P. Mindell, S.V. Edwards, E.L. Braun, C. Rahbek, D.W. Burt, P. Houde, Y. Zhang, H. Yang, J. Wang, Avian Genome Consortium, E.D. Jarvis, M.T.P. Gilbert, J. Wang, Comparative genomics reveals insights into avian genome evolution and adaptation, Science. 346 (2014) 1311–1320.

[45] J. Xu, J.-T. Li, Y. Jiang, W. Peng, Z. Yao, B. Chen, L. Jiang, J. Feng, P. Ji, G. Liu, Z. Liu, R. Tai, C.

426 Dong, X. Sun, Z.-X. Zhao, Y. Zhang, J. Wang, S. Li, Y. Zhao, J. Yang, X. Sun, P. Xu, Genomic
427 Basis of Adaptive Evolution: The Survival of Amur Ide (Leuciscus waleckii) in an Extremely
428 Alkaline Environment, Mol. Biol. Evol. 34 (2017) 145–159.

429 [46] M.R. Dorwart, N. Shcheynikov, D. Yang, S. Muallem, The solute carrier 26 family of proteins in
430 epithelial ion transport, Physiology . 23 (2008) 104–114.

431 [47] G.D. Kavembe, P. Franchini, I. Irisarri, G. Machado-Schiaffino, A. Meyer, Genomics of adaptation
432 to multiple concurrent stresses: insights from comparative transcriptomics of a Cichlid fish from one
433 of earth's most extreme environments, the hypersaline soda Lake Magadi in Kenya, east Africa, J.
434 Mol. Evol. 81 (2015) 90–109.

435 [48] D. Markovich, H. Murer, The SLC13 gene family of sodium sulphate/carboxylate cotransporters,
436 Pflugers Arch. 447 (2004) 594–602.

437 [49] D. Basilio, K. Noack, A. Picollo, A. Accardi, Conformational changes required for H(+)/Cl(-)
438 exchange mediated by a CLC transporter, Nat. Struct. Mol. Biol. 21 (2014) 456–463.

439 [50] C. Cang, K. Aranda, Y.-J. Seo, B. Gasnier, D. Ren, TMEM175 Is an Organelle K(+) Channel
440 Regulating Lysosomal Function, Cell. 162 (2015) 1101–1112.

441 [51] M.P. Wilkie, C.M. Wood, The adaptations of fish to extremely alkaline environments, Comp.
442 Biochem. Physiol. B Biochem. Mol. Biol. 113 (1996) 665–673.

443 [52] A.G. Heath, Water pollution and fish physiology, CRC press, 2018.
444

Table

Table 1. Summary of orthologous genes in five fish genomes and *G. przewalskii* transcriptomic assemblies.

Species	Genes	Genes in orthologous groups	Unique orthologs
Danio rerio	52,089	29,232	17,001
Gadus morhua	22,100	16,884	16,390
Takifugu rubripes	47,841	25,137	16,071
Oryzias latipes	24,674	17857	15,877
Oreochromis niloticus	26,763	19,432	17,433
Gymnocypris przewalskii	137,539	74,107	16,379

451 **Table 2**. List of rapidly evolving ion transport genes in *Gymnocypris przewalskii*.

Gene name	Description	Adjusted P-value
SCN3B	Sodium channel subunit beta-3	0.000076
ATP1B2	Sodium/potassium-transporting ATPase subunit beta-2	0.020563
NALCN	Sodium leak channel non-selective protein	0.029246
SLC13A3	Solute carrier family 13 member 3	0.002234
TMEM175	Transmembrane protein 175	0.002100
CLCN7	H(+)/Cl(-) exchange transporter 7	0.000019
TRPV6	Transient receptor potential cation channel subfamily V member 6	0.022243
MCU	Calcium uniporter protein, mitochondrial	0.003946
MICU2	Calcium uptake protein 2, mitochondrial	0.002823

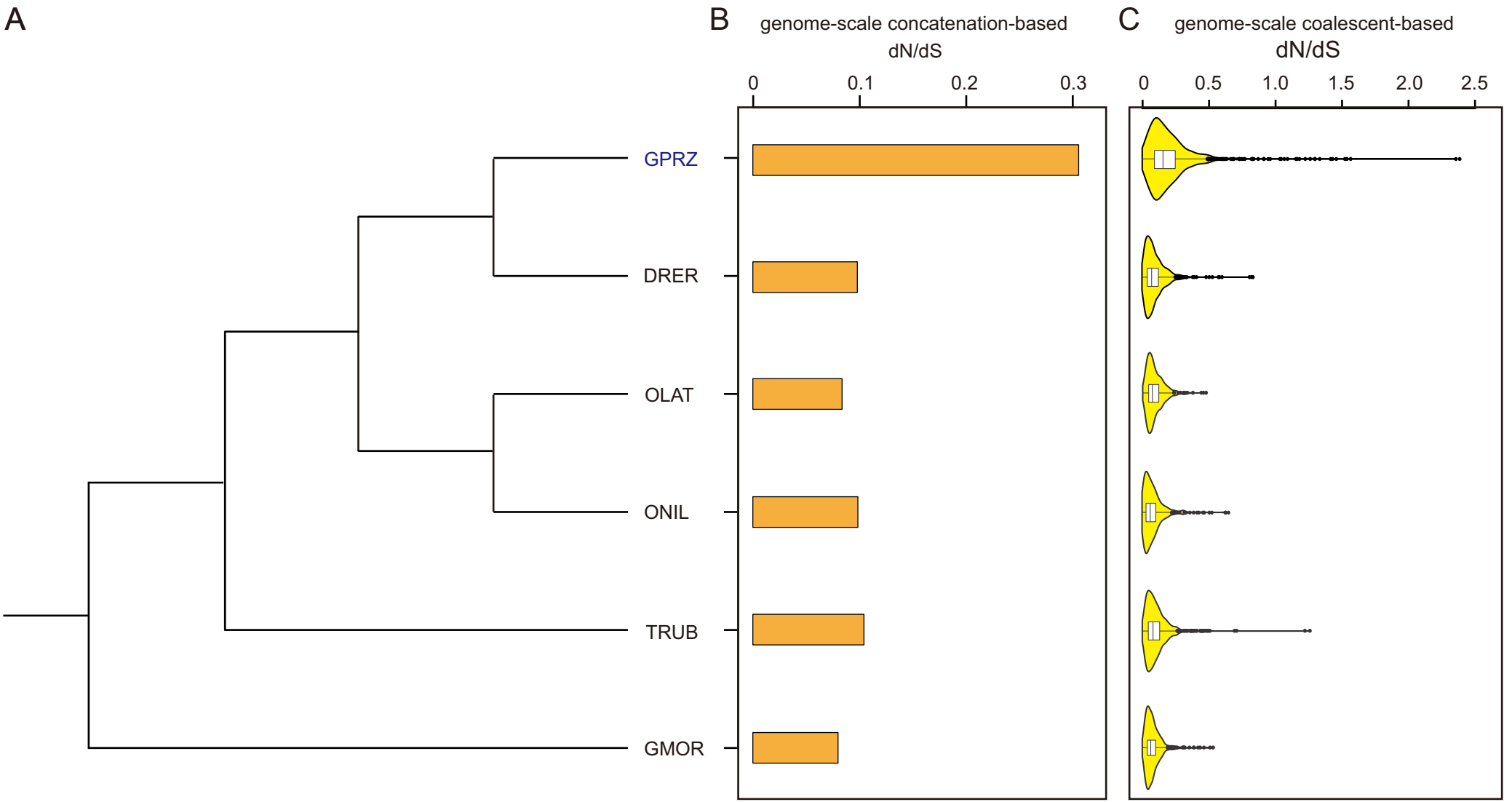
452

453

Figure legends

Figure 1. Genome-wide molecular evolutionary feature of six fish species. (A) Clipped species tree used in this study. GPRZ, *Gymnocypris przewalskii*, DRER, *Danio rerio*, OLAT, Oryzias latipes, ONIL, Oreochromis niloticus, TRUB, Takifugu rubripes, GMOR, *Gadus morhua*. (B) Barplot depicting the dN/dS of concatenated supergenes in each fish species. (C) Violin plot depicting the dN/dS of each coalescent orthologs in each species estimated by free-ratio model.

Figure 2. Expression feature of rapidly evolving ion transport genes (REITGs) in six tissues of *G. przewalskii*. (A) Barplot depicting the expression level of nine REITGs in kidney, gill, liver, spleen, brain and muscle tissues based on Log_{10}(FPKM + 1) value which estimated from RNA-seq data. (B) Heatmap depicting the expression level comparison of each REITG based on Log10(FPKM + 1) values. Tissue type and gene name are shown on the y-axis and x-axis, respectively. Plot colors reflect the expression level, ranging from low (blue) to high (red). (C) Barplot depicting the significantly enriched gene ontology for differentially expressed REITGs.


A
B
genome-scale concatenation-based
dN/dS
0
0.1
0.2
0.3
C
genome-scale coalescent-based
dN/dS
0
0.5
1.0
1.5
2.0
2.5
GPRZ
DRER
OLAT
ONIL
TRUB
GMOR