Transcriptomic signature of rapidly evolving immune genes in a highland fish

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

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

[2] Center for Advanced Retinal and Ocular Therapeutics, Scheie Eye Institute, University of Pennsylvania

School of Medicine, Philadelphia, PA, USA;

[3] Key Laboratory of Adaptation and Evolution of Plateau Biota, Northwest Institute of

Plateau Biology, Chinese Academy of Sciences, Xining, 810001, China

* 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

Kai Zhao: zhaokai@nwipb.cas.cn

Running title: Rapidly evolving immune genes in highland fish

Abstract

Recent genome-wide studies have begun to elucidate the genomic basis of hypoxia, long-term cold and high saline and alkaline adaptation in highland fish, and a number of key genes contributed to its highland adaptation were identified. An increasing number of studies indicated that immune genes of Tibetan endemic fish species underwent positive selection towards functional shift, while the insight into immune gene repertoire of Tibetan highland fishes from genome-wide studies has largely lagged behind. In this study, we performed one of the first comparative genomics study in particular focusing on the signatures of immune genes in a highland fish, *Gymnocypris przewalskii* based on immune-relevant tissue transcriptome assemblies. We identified seven putative rapidly evolving immune genes with elevated molecular evolutionary rate (dN/dS) relative to lowland fish species. Using tissue-transcriptome data, we found most of rapidly evolving immune genes were broadly expressed in head-kidney, spleen, gills and skin tissues, which significantly enriched for complement activation and inflammatory response processes. In addition, we found a set of complement activation related genes underwent accelerated evolution and showed consistently repressed expression patterns in response to parasite *Ichthyophthirius multifiliis infection*. Moreover, we detected a number of immune genes involved in adaptive immune system exhibited distinct signature of upregulated expression patterns after parasite infection. Taken together, this study provided putative transcriptomic signatures of rapidly evolving immune genes, and will gain the insight into Schizothoracine fish adaptation to high-altitude extreme aquatic environments including diversified pathogen challenge.

Keywords: Comparative genomics, immunity, Schizothoracine fish, Tibetan Plateau

1. Introduction

Understanding how fish species adapt to the aquatic life at high altitude is an interest for both ichthyologists and evolutionary biologists [1]. As the largest and highest highland in the world with the average elevation above 4000m [2], the Tibetan Plateau imposes extremely harsh aquatic environments on endemic fish fauna [3], including hypoxia [4], chronic cold [5,6] and high salinity and alkalinity [7,8]. Most endemic Tibetan fish species had developed unique morphological, physiological or genetic features to tolerate such harsh living conditions [6,9]. The Schizothoracinae is the largest and most diverse taxon of ichthyofauna on the Tibetan Plateau, which are distributed throughout the Tibetan Plateau and its peripheral regions [3,10]. Past evidences revealed that Schizothoracine fishes had well adapted to above extremely inhospitable aquatic environments [3,11,12]. In addition, recent genome-wide studies have begun to elucidate the genomic basis of hypoxia, long-term cold and high saline and alkaline adaptation in Schizothoracine fish [5–7,13,14]. Besides, the aquaculture of several Schizothoracine fish species has been developing rapidly due to the high economic values [15], such as *Gymnocypris przewalskii*. However, continuous growth of the aquaculture industry is threatened by infectious disease, mainly known as the "white spot disease", which caused by infection with a highly virulent ciliate parasite *Ichthyophthirius multifiliis* [15–17]. The severe outbreaks of parasite infection could result in even more than 65% mortality and caused great economic losses in farmed *G. przewalskii* [18]. Therefore, these evidences indicated that Schizothoracine fish might have not well established the immune system against diversified pathogens invasion. It may provide novel insight into the highland adaptation by focusing in particular on the mechanism of immune genes in Schizothoracine fishes.

Recent studies employing genome-wide approaches identified massive immune genes in several aquaculture fish species [19,20]. With the recent rapid advances in sequencing technologies for non-model organisms without reference genomes, an increasing number of transcriptome-wide studies mainly

focused on the identification of immune genes in fish species [21]. While, the insight into immune gene repertoire of Schizothoracine fishes from genome-wide studies has largely lagged behind [22,23]. Notably, a growing body of studies provided the evidences of immune genes showing signatures of rapid evolution and positive selection in Schizothoracine fishes, including *Gymnocypris przewalskii* [16,23,24] and *Gymnocypris eckloni* [25–27]. Among Schizothoracine fishes, Tibetan naked carp (*Gymnocypris przewalskii*) is one of the best characterized Tibetan endemic fish species, and a number of key genes contributed to its highland adaptation were identified by previous studies [7,13,14,16,24,28]. Therefore, it is an interesting issue to investigate the evolutionary signature of immune genes that contributed to highland adaptation in Schizothoracinae using *G. przewalskii* as a case study in highland fish species.

In this study, we performed the comparative genomics analysis and identified the putative rapidly evolving immune genes in a highland fish, *G. przewalskii*. Specifically, we used recently available *G. przewalskii* transcriptome and other four teleost fish genomes for comparison. In addition, using tissue-transcriptomics, we also characterized the expression pattern of rapidly evolving immune genes in four immune-relevant tissues and response to parasite infection.

2. Materials and methods

2.1. Data collection and transcriptome assembly

We downloaded the available transcriptomic data of *G. przewalskii* from NCBI SRA database (https://www.ncbi.nlm.nih.gov/sra). Specifically, we used four unpublished tissue transcriptome sequencing data including gills, head-kidney, spleen and skin of *G. przewalskii*, plus a spleen transcriptome in response to *Ichthyophthirius multifiliis* infection [17]. Sequencing reads were checked for quality using package FastQC. Sequencing adapters and reads with a quality score < 20 were trimmed

with Trimmomatic [29], resulting in clean reads. Transcriptome was *de novo* assembled with Trinity [30] with default parameters. After assembly, the redundant transcripts were removed using CD-HIT [31] with the threshold of 0.90, and only the longest transcript under each cluster was extracted as unigene (unique gene). Next, the open reading frame (ORF) of each unigene was predicted using MARKER [32] and TransDecoder (https://github.com/TransDecoder/TransDecoder).

2.2. Orthologous gene group identification

We included four well-annotated teleost fish genomes and downloaded from Ensembl database (release 96) (http://useast.ensembl.org/index.html) to build a local protein database, including zebrafish (*Danio rerio*), tilapia (*Oreochromis niloticus*), medaka (*Oryzias latipes*) and spotted gar (*Lepisosteus oculatus*). We then translated the nucleotide sequences of protein-coding genes from the transcriptome assemblies of *G. przewalskii* into amino acid sequences using an in-house-developed perl script, and pooled it into above exiting protein database. Next, we conducted a self-to-self BLAST search for all protein sequences using DIAMOND [33] with an E-value cutoff of $1e^{-5}$, and removed hits with identity $< 30\%$ and coverage $< 30\%$. We identified the orthologous gene groups from the BLAST results using OrthoFinder [34] with default settings. We calculated, mapped and illustrated all the identified orthologous gene groups of each species by venn diagram. Finally, we identified the core orthologous gene groups shared by above five fish species.

Among the genes under above identified orthologous gene groups, we identified one to one, one-to-many, and many-to-many orthologs among five fish species. Notably, for each 1:1 ortholog pair, we only selected the longest transcript associated with the gene for each pair of species as putative single-copy ortholog. In addition, the one-to-many orthologs were treated as multiple copy orthologs, and many-to-many orthologs were labeled as other orthologs. Finally, the number of genes under every category in each fish species were calculated, respectively.

2.3. Phylogenetic inference

We performed the sequence alignment of each putative single-copy ortholog of five fish species using R package, MSA [35] with default parameters and trimmed gaps using trimAl [36] with parameter "-automated1". Moreover, we performed the filtration of single-copy orthologs with stricter constraints, including length (minimum 200 aa), sequence alignment (maximum missing data 50% in CDS alignments). Next, we conducted a genome-scale coalescent-based dataset including thousands of single-copy genes from five fish species. Finally, we detected the best model for tree construction using ModelTest2 [37], and then we built the Maximum Likelihood phylogenetic tree using RAxML 8 [38]. Statistical supports for major nodes were estimated from 1000 bootstrap replicates. We reconstructed the species tree using ASTRAL 4.4.4 [39] for molecular evolution analysis.

2.4. Rapidly evolving gene identification

We sought to estimate the evolutionary rate of each single-copy ortholog in each fish species, and identified a set of genes with an elevated ratio of non-synonymous to synonymous substitutions (dN/dS) in *G. przewalskii* relative to other four fish species. At first, we derived nucleotide alignments from previously prepared protein alignments of each single-copy ortholog using PAL2NAL v14 [40]. We ran two branch models using CodeML package in PAML 4.7a [41] 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 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 [42].

2.5. Gene expression analysis

We mapped all the clean sequencing reads to the transcriptome assemblies of *G. przewalskii* using RSEM [43] to obtain expected counts and transcripts per million (TPM). To avoid the contamination during dissection causes the low levels of expression in the neighboring structures, we removed genes with TPM < 1 in each dataset. We are primarily interested in the expression pattern of rapidly evolving immune genes (REIGs), and calculated the TPM value of each REIG in each tissue. At last, we annotated the differentially expressed REIGs by gene ontology using R software, package TopGO [42].

3. Results

3.1. Sequence analysis and orthologous gene groups

The *de novo* transcriptome assembly of *G. przewalskii* yielded 614,528 transcripts in total, with an N50 of 1,656 bp and an average length of 881 bp. After removing redundant isoforms and extraction of longest isoform among alternative transcripts, a total of 31,982 unigenes were obtained, with an N50 of 3,072 bp and a mean length of 1,985 bp. After prediction of protein-coding genes by MAKER and TransDecoder, we totally obtained 28,917 unigenes with full or partial length of gene coding regions (CDS) in *G. przewalskii* (supplementary table S1).

A total of 113,529 proteins from five fish species protein coding gene sets were binned into 30,218 orthologous gene groups. We identified a total of 10,334 core orthologous gene groups shared by eight

fish species (Figure 1A). After strict one to one ortholog selection, we identified 5,573 putative single-copy orthologs (only one ortholog in each gene family) in each fish species among core shared orthologous gene groups (Figure 1B), making them suitable for phylogenetic inference.

3.3. Phylogeny

After filtration for maximizing the information content of sequences and minimizing the impact of missing data, we eventually obtained 3,653 single-copy orthologs in each fish species. After protein sequence alignment, we constructed the maximum likelihood phylogenetic trees of five species based on coalesced single-copy orthologs using the GTRGAMMA model. The strongly supported (100%) phylogeny suggested that *G. przewalskii* had a closely related phylogenetic relationship with *Danio rerio*, and consistent with the tree on mitochondrial genes or nuclear DNA markers.

3.4. Rapidly evolving gene repertoire

Rapidly evolving gene is a set of genes underwent accelerated evolution with the signature of an increase rate of non-synonymous changes. We identified 468 putative rapidly evolving single-copy orthologs (REGs) in *G. przewalskii*. Among this set of genes, the most interesting finding was REGs included genes functioning in fish innate immune systems, such as Complement C2 (C2), Complement C5 (C5) and Complement factor D (CFD) that involved in complement activation process (Table 1). In addition, we found a number of REGs associated with fish adaptive immune response, such as Tumor necrosis factor alpha-induced protein 8 (TNFAIP8) and Tumor necrosis factor ligand superfamily member 10 (TNFSF10) involved in immune system process (Table 1). That is said, we identified a number of rapidly evolving immune genes in *G. przewalskii*. Besides immune genes, we also found a large number of genes involved energy metabolism functions, such as ATP5c1 and ATP5b associated with ATP binding and oxidative phosphorylation process (supplementary table S2).

3.5. Expression patterns of rapidly evolving immune genes

We focused on the expression patterns of rapidly evolving immune genes (REIGs) among four immune-relevant tissues of *G. przewalskii*, including gills, head-kidney, spleen and skin. Most REIGs were broadly expressed in above four tissues except Complement 2 (Figure 2A, supplementary table S3), and mainly involved in innate immune response, immune system process and complement activation (Figure 2B, supplementary table S4). In addition, we screened the expression patterns of REIGs in response to parasite invasion using a set of spleen transcriptome data. We found that four complement genes were all repressed after *I. multifiliis* infection (Figure 2C, supplementary table S5), which mainly significantly enriched in complement activation and innate immune response process (Figure 2D, supplementary table S6). Moreover, we detected a set of REIGs involved in adaptive immune system including TNFAIP8 and TNFSF10 that were induced by parasite infection (Figure 2C, supplementary table S5), which significantly enriched in defense response to protozoan and type I interferon signaling pathway (Figure 2D, supplementary table S6).

4. Discussion

Recent transcriptome-wide studies had begun to characterize a small scale of immune genes in Schizothoracine fishes [16,17,23,25], while our understanding of the genome-wide evolutionary signature of immune genes in highland fish remains limited. In the present study, for the first time, we used comparative genomics analysis to in particular identify rapidly evolving immune genes in Tibetan endemic fish, *G. przewalskii*. We found evidences of accelerated evolution acting on four complement genes involved innate immune system, two tumor necrosis factor genes associated with fish adaptive immune response process and one interferon regulatory factor gene. Using tissue transcriptome data, we found divergent expression patterns of rapidly evolving immune genes in head-kidney, spleen, gills and

skin tissues of *G. przewalskii*. In addition, we found complement activation related immune genes underwent accelerated evolution exhibited distinct signature of downregulated expression in response to *Ichthyophthirius multifiliis* infection.

An increasing number of transcriptome targeted to identify abundant immune genes in fish by sampling immune-relevant tissues, such as head-kidney and spleen [21–23,25]. To maximize the information content of immune gene sequences, we collected two immune relevant tissues, head-kidney and spleen, plus two less relevant tissues (gills and skin) transcriptome data for assembly, which generated much more unigenes than our previous assembly versions of *G. przewalskii [7,23]*. In addition, compared with our previous comparative genomics study in *G. przewalskii* [7], we identified more immune genes showing signature of rapidly evolving relative to low land fish species. Unlike whole genome data, although transcriptome sequencing is an effective and accessible approach to initiate comparative genomic analyses on non-model organisms [30], it still can not cover all of the protein coding genes in one species. Moreover, most Schizothoracine fish species are polyploidy, the high complexity and large size of polyploidy fish genomes has constrained the development of genomic resources for Schizothoracinae [8]. The whole genome data of tetraploid *G. przewalskii* is long lacking, the limitation of transcriptome is obviously affecting the identification of immune gene repertoire. Therefore, it is not difficult to interpret that the increasing number of rapidly evolving immune gene were found in *G. przewalskii* due to including more immune-relevant tissue transcriptome data.

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 genomics study identified a set of immune genes showing elevated evolutionary rate (dN/dS) in *G. przewalskii* relative to other low land fish species. We found four complement genes underwent accelerated evolution in *G. przewalskii*,

such as C2, C1QC, CFD and C5. Complements are important components in fish innate immune system, and activation of complement can result in the generation of activated protein fragments that contributed to microbial killing, phagocytosis, inflammatory reactions, immune complex clearance, and antibody production [45]. Moreover, we found two rapidly evolving tumor necrosis factor (TNF) genes including TNFAIP8 and TNFSF10, this finding is identical to our previous findings [7,16,23,24]. TNF genes function in many immunological and inflammatory responses under either normal or pathological conditions [46]. Notably, recent study suggested that highland fish, salt-water-dwelling wild *G. przewalskii* was susceptible to infectious disease, which triggered high mortality in freshwater farming industry, such as parasite *Ichthyophthirius multifiliis* infection [15]. In addition, previous surveys on Tibetan lakes showed that hypersaline and alkaline Lake Qinghai had low diversity of pathogens [11,12], which indicated that wild *G. przewalskii* dwelt in a lighter pathogen load aquatic environment. Therefore, we speculated that immune genes of highland fish, *G. przewalskii* have experienced accelerated evolution and functional shifts to well adapted to this specific aquatic environment.

Fish head-kidney and spleen are two key tissues contributed to immune defense against pathogen invasion [47]. In present study, we found most of rapidly evolving immune genes (REIGs) were broadly expressed across all four tissues, indicating that immune genes experiencing accelerated evolution in *G. przewalskii* may have general functions, such as complement activation, inflammatory response and innate immune response. One interesting finding is that we found C2 gene was not expressed in *G. przewalskii* skin tissues, while this was not consistent with previous study in common carp (*Cyprinus carpio*) which indicated C2 was highly expressed in skin tissue [48]. Fish skin mucus is crucial to humoral immunity [49], the lack of complement gene expression in skin may affect the innate immune response of *G. przewalskii* to pathogen invasion. Moreover, it is interpretable that the accelerated evolution may drive complement genes of *G. przewalskii* towards functional shift during long-term adaptation to the low diversity of pathogens environment in Lake Qinghai. Another intriguing finding is that, all REIGs

associated with complement activation function were repressed after *I. multifiliis* infection. Fish complement system is the first line of host defense, which promoted opsonization, lysis and local inflammatory response to pathogens infection [50]. Past evidences suggested that pathogens can avoid complement activation by mimicking host ligands, attracting complement regulators or secreting complement inhibitors [50–52]. In fact, explicit role of complement system was unclear, especially in response to *I. multifiliis* infection. Here, we speculated that *G. przewalskii* complement system may reduce the sensitivity to detect various pathogens during its adaptation to highland aquatic environment, this can result in this parasite *I. multifiliis* can evade or counteract its innate immune system to further trigger high mortality in freshwater farmed *G. przewalskii*. Since the complements were mainly synthesized in fish liver tissue, this hypothesis needs to be further confirmed by large-scale transcriptomics and functional assay in future.

5. Conclusion

We used comparative genomics based on the assemblies from multiple immune-relevant tissue transcriptomes to identify the signatures of immune genes a highland fish, *G. przewalskii*. These putative transcriptomic signatures included: 1) a number of genes involved in innate or adaptive immune system were found in *G. przewalskii* with elevated molecular evolutionary rate (dN/dS) relative to lowland fish species; 2) most rapidly evolving immune genes associated with complement activation and inflammatory response were broadly expressed in head-kidney, spleen, gills and skin of *G. przewalskii*; 3) A set of complement activation related genes underwent accelerated evolution and showed consistently repressed expression patterns in response to parasite *I. multifiliis* infection; 4) A set of genes involved in adaptive immune system exhibited distinct signature of upregulated expression after parasite infection. Taken together, our study will provide the transcriptomic signatures of rapidly evolving immune genes, and gain

the insight into Schizothoracine fish adaptation to high-altitude extreme environments including diversified pathogen challenge.

Supplementary Material

Supplementary table S1. Summary of protein-coding gene prediction.

Supplementary table S2. List of rapidly evolving genes in *G. przewalskii*.

Supplementary table S3. Expression levels of rapidly evolving immune genes in *G. przewalskii* head-kidney, spleen, gills and skin tissues.

Supplementary table S4. Ontology for broadly expressed rapidly evolving immune genes in *G. przewalskii*.

Supplementary table S5. Expression levels of rapidly evolving immune genes in spleen tissue *G. przewalskii* after parasite infection.

Supplementary table S6. Ontology for differentially expressed rapidly evolving immune genes in *G. przewalskii* spleen after parasite infection.

Acknowledgments

References

[1]   P.E. Bickler, L.T. Buck, Hypoxia tolerance in reptiles, amphibians, and fishes: life with variable

oxygen availability, Annu. Rev. Physiol. 69 (2007) 145–170.

[2]   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.

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

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

[5]   C. Zhang, C. Tong, F. Tian, K. Zhao, Integrated mRNA and microRNA transcriptome analyses reveal regulation of thermal acclimation in Gymnocypris przewalskii: A case study in Tibetan Schizothoracine fish, PLoS One. 12 (2017) e0186433.

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

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

[8]   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.

[9]   L.Y. Wen, Uplift of the tibetan plateau influenced the morphological evolution of animals, J. Agric. Sci. 6 (2014) 244.

[10] Y.F. Chen, W.X. Cao, Schizothoracinae. Fauna Sinica, Osteichthyes, Cypriniformes III, (2000).

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

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

[13] C. Tong, F. Tian, C. Zhang, K. Zhao, The microRNA repertoire of Tibetan naked carp Gymnocypris przewalskii: A case study in Schizothoracinae fish on the Tibetan Plateau, PLoS One. 12 (2017) e0174534.

[14] 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.

[15] D. Chen, F. Xiong, J. Shi, H. Qi, The protection research of Gymnocypris przewalskii, (2011).

[16] C. Tong, Y. Lin, C. Zhang, J. Shi, H. Qi, K. Zhao, Transcriptome-wide identification, molecular evolution and expression analysis of Toll-like receptor family in a Tibet fish, Gymnocypris przewalskii, Fish Shellfish Immunol. 46 (2015) 334–345.

[17] F. Tian, C. Tong, C. Feng, K. Wanghe, K. Zhao, Transcriptomic profiling of Tibetan highland fish (Gymnocypris przewalskii) in response to the infection of parasite ciliate Ichthyophthirius multifiliis, Fish Shellfish Immunology. 70 (2017) 524–535.

[18] H. Qi, J. Shi, An overview on the resource enhancement of Gymnocypris przewalskii (Kessler), Chinese Fisheries Economics. (2009). http://en.cnki.com.cn/Article_en/CJFDTotal-ZYJJ200905017.htm.

[19] T. Xu, G. Xu, R. Che, R. Wang, Y. Wang, J. Li, S. Wang, C. Shu, Y. Sun, T. Liu, J. Liu, A. Wang, J. Han, Q. Chu, Q. Yang, The genome of the miiuy croaker reveals well-developed innate immune and sensory systems, Sci. Rep. 6 (2016) 21902.

[20] C. Wu, D. Zhang, M. Kan, Z. Lv, A. Zhu, Y. Su, D. Zhou, J. Zhang, Z. Zhang, M. Xu, L. Jiang, B. Guo, T. Wang, C. Chi, Y. Mao, J. Zhou, X. Yu, H. Wang, X. Weng, J.G. Jin, J. Ye, L. He, Y. Liu, The draft genome of the large yellow croaker reveals well-developed innate immunity, Nat. Commun. 5 (2014) 5227.

[21] I. Salinas, S. Magadán, Omics in fish mucosal immunity, Dev. Comp. Immunol. 75 (2017) 99–108.

[22] H. Luo, S. Xiao, H. Ye, Z. Zhang, C. Lv, S. Zheng, Z. Wang, X. Wang, Identification of Immune-Related Genes and Development of SSR/SNP Markers from the Spleen Transcriptome of Schizothorax prenanti, PLoS One. 11 (2016) e0152572.

[23] C. Tong, C. Zhang, R. Zhang, K. Zhao, Transcriptome profiling analysis of naked carp (Gymnocypris przewalskii) provides insights into the immune-related genes in highland fish, Fish Shellfish Immunol. 46 (2015) 366–377.

[24] C. Tong, F. Tian, Y. Tang, C. Feng, L. Guan, Positive Darwinian selection within interferon regulatory factor genes of Gymnocypris przewalskii (Cyprinidae) on the Tibetan Plateau, Fish Shellfish Immunology. (2016).

https://www.sciencedirect.com/science/article/pii/S1050464816300122.

[25] D. Qi, M. Xia, Y. Chao, Y. Zhao, R. Wu, Identification, molecular evolution of toll-like receptors in a Tibetan schizothoracine fish (Gymnocypris eckloni) and their expression profiles in response to acute hypoxia, Fish Shellfish Immunol. 68 (2017) 102–113.

[26] D. Qi, Y. Chao, J. Liang, Q. Gao, R. Wu, I. Mather, Y. Zhao, Q. Chen, Adaptive evolution of interferon regulatory factors is not correlated with body scale reduction or loss in schizothoracine fish, Fish Shellfish Immunol. 73 (2018) 145–151.

[27] D. Qi, Y. Chao, C. Zhang, Z. Wang, W. Wang, Q. Chen, Z. Zheng, Z. Zhang, Duplication of toll-like receptor 22 in teleost fishes, Fish Shellfish Immunol. 94 (2019) 752–760.

[28] C. Tong, C. Zhang, J. Shi, H. Qi, R. Zhang, Y. Tang, G. Li, C. Feng, K. Zhao, Characterization of two paralogous myostatin genes and evidence for positive selection in Tibet fish: Gymnocypris przewalskii, Gene. 565 (2015) 201–210.

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

[30] 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.

[31] 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.

[32] B.L. Cantarel, I. Korf, S.M.C. Robb, G. Parra, E. Ross, B. Moore, C. Holt, A. Sánchez Alvarado, M. Yandell, MAKER: an easy-to-use annotation pipeline designed for emerging model organism genomes, Genome Res. 18 (2008) 188–196.

[33] B. Buchfink, C. Xie, D.H. Huson, Fast and sensitive protein alignment using DIAMOND, Nat. Methods. 12 (2015) 59–60.

[34] D.M. Emms, S. Kelly, OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy, Genome Biol. 16 (2015) 157.

[35] U. Bodenhofer, E. Bonatesta, C. Horejš-Kainrath, S. Hochreiter, msa: an R package for multiple sequence alignment, Bioinformatics. 31 (2015) 3997–3999.

[36] 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.

[37] D. Posada, K.A. Crandall, MODELTEST: testing the model of DNA substitution, Bioinformatics. 14 (1998) 817–818.

[38] A. Stamatakis, RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies, Bioinformatics. 30 (2014) 1312–1313.

[39] S. Mirarab, R. Reaz, M.S. Bayzid, T. Zimmermann, M.S. Swenson, T. Warnow, ASTRAL: genome-scale coalescent-based species tree estimation, Bioinformatics. 30 (2014) i541–8.

[40] 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.

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

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

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

[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] M.C.H. Holland, J.D. Lambris, The complement system in teleosts, Fish Shellfish Immunol. 12 (2002) 399–420.

[46] Q.W. Qin, M. Ototake, K. Noguchi, G.-I. Soma, Y. Yokomizo, T. Nakanishi, Tumor necrosis factor alpha (TNFα)-like factor produced by macrophages in rainbow trout, Oncorhynchus mykiss, Fish Shellfish Immunol. 11 (2001) 245–256.

[47] W.S. Hoar, D.J. Randall, G. Iwama, T. Nakanishi, The Fish Immune System: Organism, Pathogen, and Environment, Academic Press, 1997.

[48] S.F. Gonzalez, K. Buchmann, M.E. Nielsen, Complement expression in common carp (Cyprinus carpio L.) during infection with Ichthyophthirius multifiliis, Dev. Comp. Immunol. 31 (2007) 576–586.

[49] F.A. Guardiola, A. Cuesta, E. Abellán, J. Meseguer, M.A. Esteban, Comparative analysis of the humoral immunity of skin mucus from several marine teleost fish, Fish Shellfish Immunol. 40 (2014) 24–31.

[50] D. Serruto, R. Rappuoli, M. Scarselli, P. Gros, J.A.G. van Strijp, Molecular mechanisms of complement evasion: learning from staphylococci and meningococci, Nat. Rev. Microbiol. 8 (2010) 393–399.

[51] P.F. Zipfel, R. Würzner, C. Skerka, Complement evasion of pathogens: common strategies are shared by diverse organisms, Mol. Immunol. 44 (2007) 3850–3857.

[52] P. Ronza, D. Robledo, R. Bermúdez, A.P. Losada, B.G. Pardo, A. Sitjà-Bobadilla, M.I. Quiroga, P. Martínez, RNA-seq analysis of early enteromyxosis in turbot (Scophthalmus maximus): new insights into parasite invasion and immune evasion strategies, Int. J. Parasitol. 46 (2016) 507–517.
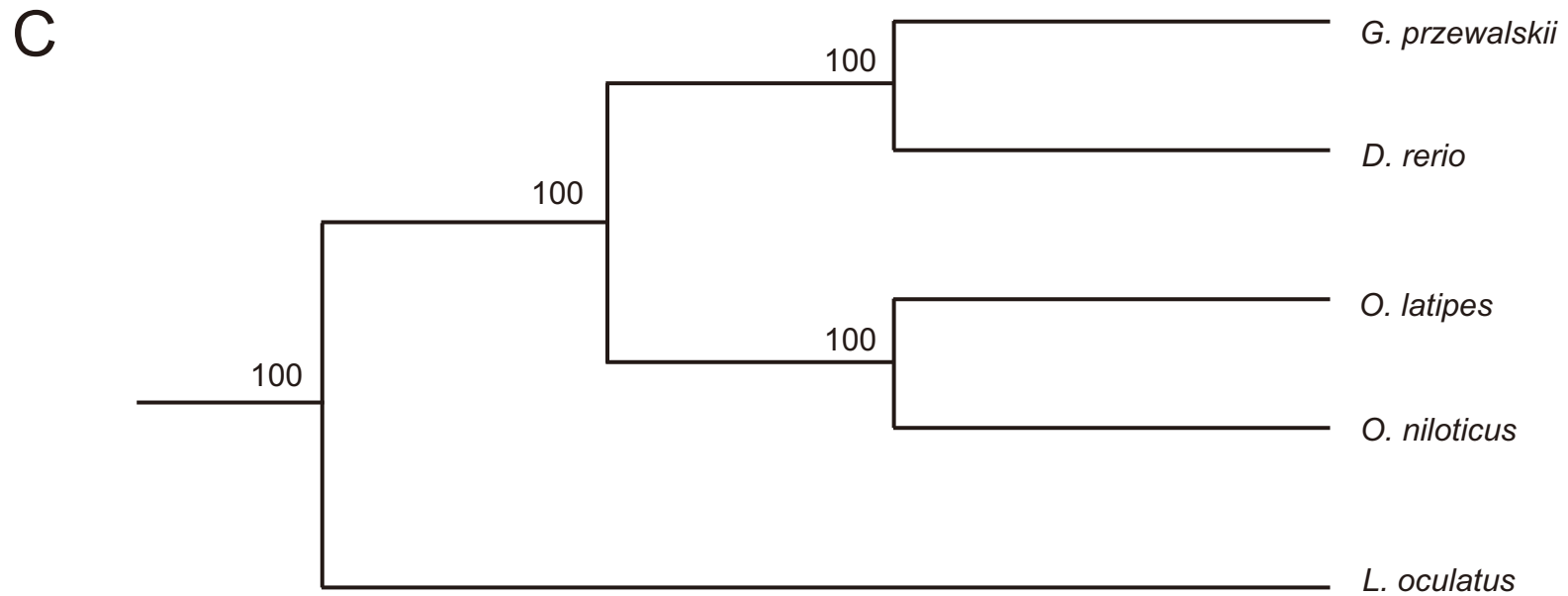
Table 1. List of rapidly evolving immune genes in *G. przewalskii*.

Gene name	Description	Adjusted P-value
C2	Complement C2	0.00267084
IRF8	Interferon regulatory factor 8	0.00181540
C1QC	Complement C1q subcomponent subunit C	0.00981918
CFD	Complement factor D	0.00001774
TNFAIP8	Tumor necrosis factor alpha-induced protein 8	0.00169457
C5	Complement C5	0.00123476
TNFSF10	Tumor necrosis factor ligand superfamily member 10	0.01450549

Figure legends

Figure 1. Genome content composition and genome-wide phylogeny. (A) Venn diagram showed the core shared orthologous gene groups (gene family) by five fish spe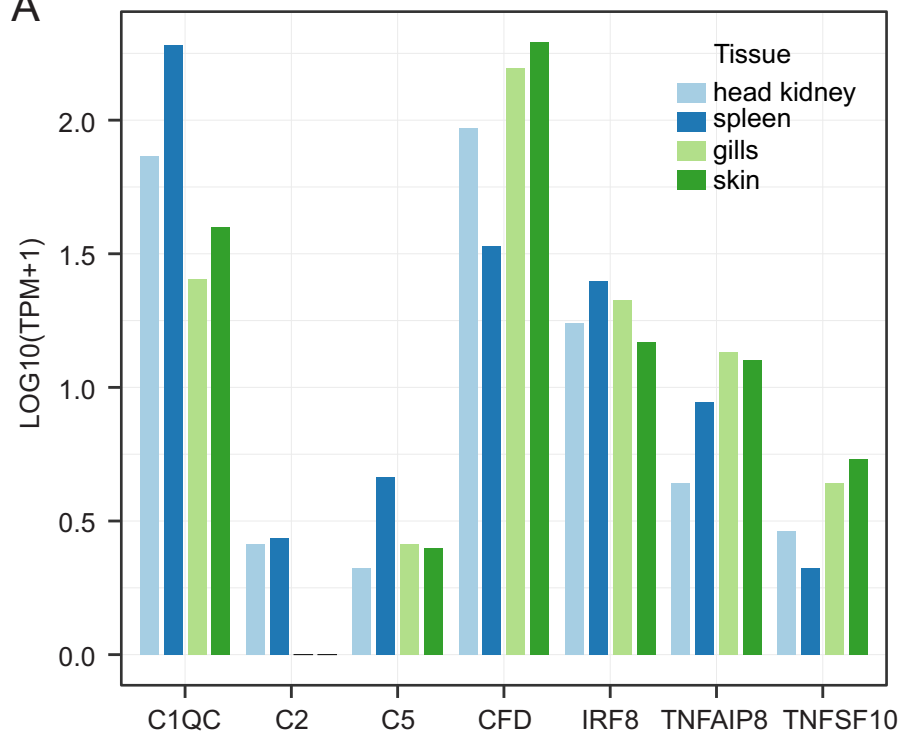cies. (B) Spine plot depicting the composition of different categories of gene families labeled by colors. (C) Genome-wide phylogeny of five fish species.

Figure 2. Expression patterns of rapidly evolving immune genes (REIGs) in four tissues, and spleen after parasite infection. (A) Bar plot depicting the expression level ($Log_{10}(TPM+1)$ value) of seven REIGs in head-kidney, spleen, gills and skin. (B) Bar plot depicting the top five gene ontology for broadly-expressed REIGs. (C) Bar plot depicting the expression level of seven REIGs in spleen between control and infected group. (D) Bar plot depicting the top five gene ontology for up- and down-regulated REIGs in *G. przewalskii*.

A

G. przewalskii
375

L. oculatus
176

D. rerio
287

core shared
orthologous
gene group

N = 10334

787

164

O. niloticus

O. latipes

B

L. oculatus

O. niloticus

O. latipes

D. rerio

G. przewalskii

Single-copy orthologs
Multiple copy orthologs
Other orthologs
Unclustered genes

0   40   80   100

Relative abundance (%)

C

100 — *G. przewalskii*

100 — *D. rerio*

100

100 — *O. latipes*

100 — *O. niloticus*

L. oculatus

A (Bar chart, x-axis: C1QC, C2, C5, CFD, IRF8, TNFAIP8, TNFSF10; y-axis: LOG10(TPM+1))

Tissue:
- head kidney
- spleen
- gills
- skin

B Ontology for broadly-expressed REIGs (x-axis: −LOG10(P−value))
- complement activation
- innate immune response
- immune response
- immune system process
- apoptotic process

C (Bar chart, x-axis: C1QC, C2, C5, CFD, IRF8, TNFAIP8, TNFSF10; y-axis: LOG10(TPM+1))

Group:
- Control group
- Infected group

D Ontology for differentially expressed REIGs (x-axis: −LOG10(P−value))

down regulated:
- complement activation
- innate immune response
- regulation of complement activation
- inflammatory response
- immune system process

up regulated:
- immune response
- defense response to protozoan
- response to bacterium
- defense response to bacterium
- type I interferon signaling pathway