


RESEARCH ARTICLE

Open Access



# Genomic insights into the seawater adaptation in Cyprinidae

Ying Wang<sup>1,3,6\*†</sup>, Xuejing Zhang<sup>1†</sup>, Jing Wang<sup>2†</sup>, Cheng Wang<sup>2,4†</sup>, Fei Xiong<sup>1†</sup>, Yuting Qian<sup>2,4†</sup>, Minghui Meng<sup>2</sup>, Min Zhou<sup>1</sup>, Wenjun Chen<sup>2,4</sup>, Zufa Ding<sup>2,4</sup>, Dan Yu<sup>2,4</sup>, Yang Liu<sup>2,4</sup>, Yumei Chang<sup>5</sup>, Shunping He<sup>2,3\*</sup> and Liandong Yang<sup>2,3,6\*</sup> 

## Abstract

**Background** Cyprinidae, the largest fish family, encompasses approximately 367 genera and 3006 species. While they exhibit remarkable adaptability to diverse aquatic environments, it is exceptionally rare to find them in seawater, with the Far Eastern daces being of few exceptions. Therefore, the Far Eastern daces serve as a valuable model for studying the genetic mechanisms underlying seawater adaptation in Cyprinidae.

**Results** Here, we sequenced the chromosome-level genomes of two Far Eastern daces (*Pseudaspius brandtii* and *P. hakonensis*), the two known cyprinid fishes found in seawater, and performed comparative genomic analyses to investigate their genetic mechanism of seawater adaptation. Demographic history reconstruction of the two species reveals that their population dynamics are correlated with the glacial-interglacial cycles and sea level changes. Genomic analyses identified *Pseudaspius*-specific genetic innovations related to seawater adaptation, including positively selected genes, rapidly evolving genes, and conserved non-coding elements (CNEs). Functional assays of *Pseudaspius*-specific variants of the prolactin (*prl*) gene showed enhanced cell adaptation to greater osmolarity. Functional assays of *Pseudaspius* specific CNEs near *atg7* and *usp45* genes suggest that they exhibit higher promoter activity and significantly induced at high osmolarity.

**Conclusions** Our results reveal the genome-wide evidence for the evolutionary adaptation of cyprinid fishes to seawater, offering valuable insights into the molecular mechanisms supporting the survival of migratory fish in marine environments. These findings are significant as they contribute to our understanding of how cyprinid fishes navigate and thrive in diverse aquatic habitats, providing useful implications for the conservation and management of marine ecosystems.

**Keywords** Far Eastern dace, Migratory, Osmoregulation, Seawater adaptation

<sup>†</sup>Ying Wang, Xuejing Zhang, Jing Wang, Cheng Wang, Fei Xiong, and Yuting Qian contributed equally to this work.

\*Correspondence:

Ying Wang  
xinyuanwangying@163.com

Shunping He  
clad@ihb.ac.cn

Liandong Yang  
yangliandong1987@163.com

<sup>1</sup> Hubei Engineering Research Center for Protection and Utilization of Special Biological Resources in the Hanjiang River Basin, College of Life Sciences, Jiangnan University, Wuhan 430056, China

<sup>2</sup> State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan 430072, China

<sup>3</sup> Academy of Plateau Science and Sustainability, Qinghai Normal University, Xining 810016, China

<sup>4</sup> University of Chinese Academy of Sciences, Beijing 100049, China

<sup>5</sup> National and Local Joint Engineering Laboratory for Freshwater Fish Breeding, Heilongjiang Province's Key Laboratory of Fish Stress Resistance Breeding and Germplasm Characteristics On Special Habitats, Heilongjiang River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Harbin 150070, Heilongjiang, China

<sup>6</sup> School of Biological Sciences, University of Bristol, Life Sciences Building, 24 Tyndall Avenue, Bristol BS81TQ, UK



## Background

Fish migration is the movement from one habitat or region to another typically for spawning [1, 2]. There are three styles among migratory fishes: oceanodromous fishes migrating exclusively within seawater, potamodromous fishes migrating exclusively within freshwater, and diadromous fishes migrating between seawater and freshwater [3, 4]. During migration, fish are confronted multiple challenges posed by environmental conditions including temperature, water flow, and salinity [5, 6]. Three hundred sixty-eight species of diadromous fishes are known (<http://www.fishbase.cn>), including salmonoids, osmeroids, galaxioids, lampreys, sturgeons, and sicydiine gobies [7–9]. However, the Far Eastern daces (*Pseudaspius* species) are among the only diadromous fishes in Cyprinidae, with limited migratory behavior observed in some other cyprinid species, including idle (*Leuciscus idus*), roach (*Rutilus rutilus*), and bream (*Vimba vimba*) [10–12].

Meanwhile, environmental salinity fluctuations exert significant influences on various aspects of fish biology such as osmotic regulation, hormone control, energy metabolism, and growth [13–15]. The Far Eastern daces, *Pseudaspius* species (Pisces, Cyprinidae), speciated and dispersed around the Sea of Japan [16]. The genus *Pseudaspius* encompasses two freshwater resident species (*P. nakamurai* and *P. sachalinensis*) and two diadromous species (*P. hakonensis* and *P. brandtii*), belonging to the subfamily Leuciscinae within the large family Cyprinidae [17]. As migratory species of cyprinid fishes, *P. hakonensis* and *P. brandtii* comprise a unique group with the freshwater-seawater transition [18], which has attracted much research interest. *P. hakonensis* and *P. brandtii* are commonly known as gold beachhead fish and black beachhead fish, respectively, and return from the Japan sea for spawning in Suifen River (Heilongjiang Province in China) from April to July each year [19]. Specifically, they migrate upstream into the estuary in batches after their gonadal maturation, and the parent fish immediately return to the offshore area after spawning; the juveniles grow in the river then they overwinter in the river depths before migrating to the sea [20, 21]. A previous study reported that *P. brandtii* evolved unique strategies and exhibited higher tolerance and adaptability to salinity and alkalinity than other cyprinids [22, 23]. Therefore, the *P. hakonensis* and *P. brandtii* genomes may represent a genetic resource for studies on seawater adaptation in migratory fishes.

Multiple studies have documented the genetic basis of seawater adaptation of salmonid fishes and disentangled osmoregulatory genes and functional groups [24–26]. Recently, genomic studies in salmon and trout strikingly revealed a small genomic region (including *rock1* and

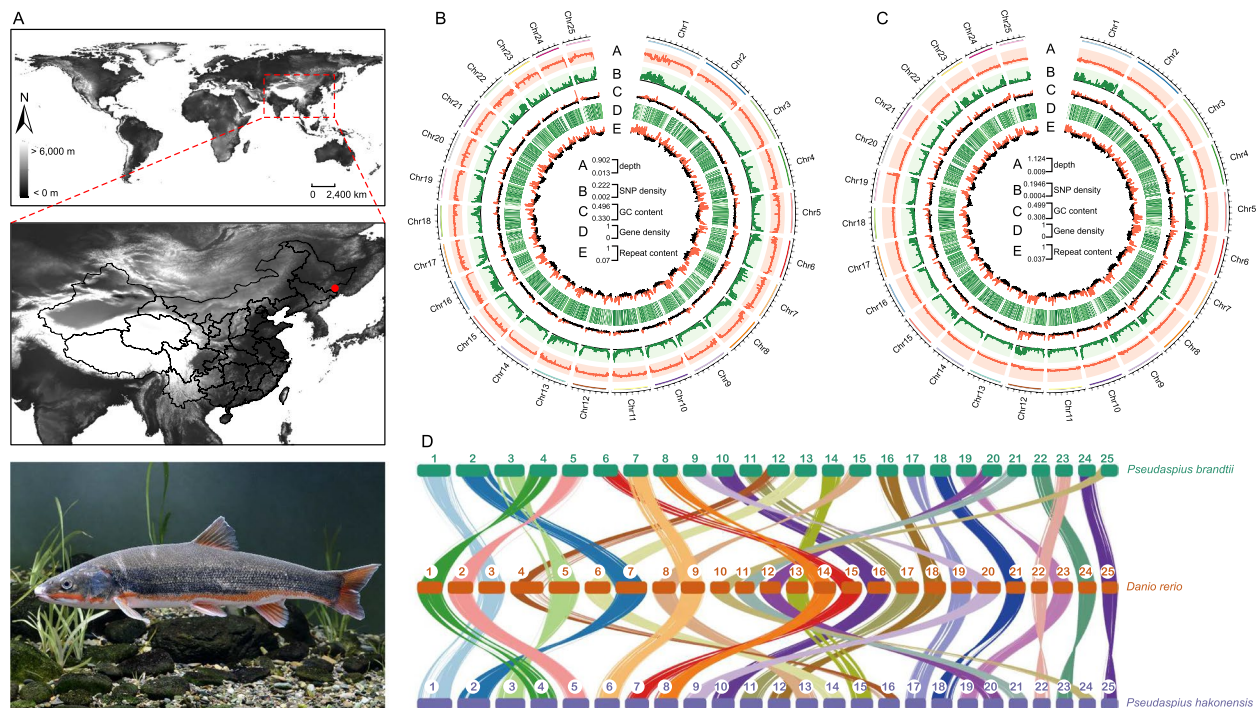
*greb1l* genes) was strongly associated with migration behavior and timing [26–29]. Furthermore, simulated experiments conducted on fishes subjected to salinity treatment have revealed insights into the significance of the osmoregulatory gene family, associated metabolic pathways, and gene expression patterns in determining crucial adaptations to seawater [24, 30–32]. For example, the euryhaline Javafish medaka (*Oryzias javanicus*) genome sequencing provided insights into the molecular basis of osmotic regulation and hatching enzyme activity [32]. Nevertheless, the genetic mechanism of *Pseudaspius* species (*P. brandtii* and *P. hakonensis*) adapting to seawater is unclear.

In this study, we focused on two species of cyprinid fishes in seawater, *P. hakonensis* and *P. brandtii*, to unravel the genetic basis of their adaptation to seawater during the migratory process. To achieve this, we first generated high-quality, chromosome-level assemblies of the genomes of *P. hakonensis* and *P. brandtii*. By comparing these genomes with those of other teleosts, we identified genes that underwent positive selection and rapid evolution as well as several conserved non-coding elements (CNEs) that are likely responsible for their seawater adaptation. Importantly, we validated the *Pseudaspius*-specific genetic innovations using functional assays. The findings of this study shed light on the intricate genetic mechanisms underlying seawater adaptation in *Pseudaspius* species. These discoveries serve as a foundation for further investigations into the adaptation of migratory fishes and contribute to our understanding of their remarkable ability to thrive in osmotically contrasting.

## Results

### Genome assembly and genome annotation

Using Illumina, Nanopore, and Hi-C sequencing, we acquired a total of 236.91 Gb data for *P. hakonensis* and 231.44 Gb data for *P. brandtii*, respectively (Fig. 1A, B, and C, Additional file 1: Table S1). Subsequently, we generated chromosome-level genome assemblies for *P. hakonensis* and *P. brandtii*. The assemblies of *P. hakonensis* and *P. brandtii* genomes were 840.77 Mb and 848.49 Mb, with contig N50 length of 3.68 Mb and 3.63 Mb, GC content of 39% and 38%, respectively (Additional file 1: Table S2 and S3). To further refine the assemblies, we anchored the contigs onto 25 chromosomes with mounting rates of 96.54% for *P. hakonensis* and 97.20% for *P. brandtii*. The genome-wide Hi-C heatmaps showing chromosome crosstalk adhered to the interaction rule, with noticeably stronger signal strength around the diagonal compared to other positions (Additional file 2: Figure S1, Additional file 1: Table S4). This suggests the high quality and completeness of the genome assembly. The summary of genome characteristics of *P. hakonensis* and *P. brandtii*



**Fig. 1** Sampling sites and genome assemblies for *P. hakonensis* and *P. brandtii*. **A** A sampling site (red circle) in Suifen River (Helongjiang province, China) and a photo of the *P. brandtii*. **B** Circos plots showed the distributions of genomic components in *P. hakonensis*; the outer circle was the chromosome. **C** Circos plots showed the distributions of genomic components in *P. brandtii*; the outer circle represents the chromosome. **D** Chromosomal syntenic relationship of *P. brandtii*, *P. hakonensis*, and *D. rerio*. The number (1–25) represented the number of the chromosome

are shown in Fig. 1B, C. Evaluation of genome completeness based on BUSCO identified that 94.50% of the complete BUSCO in the *P. hakonensis* genome assembly including 88.00% of the complete and single-copy genes and 6.50% of the complete and duplicated genes, while 94.90% of the complete BUSCO in the *P. brandtii* genome assembly including 87.90% of the complete and single-copy genes and 7.00% of the complete and duplicated genes (Additional file 1: Table S5). These results confirm the high quality of our genome assemblies for both species, providing a solid foundation for our investigations into the genetic basis of seawater adaptation in *P. hakonensis* and *P. brandtii*.

We conducted an analysis of repetitive sequences and protein-coding genes in the genomes of *P. hakonensis* and *P. brandtii*. Approximately 36.93% of the bases in the *P. hakonensis* genome and 37.20% of the bases in the *P. brandtii* genome were classified as repetitive sequences (Additional file 1: Table S6). We also predicted a total of 23,839 and 24,012 protein-coding genes in *P. hakonensis* and *P. brandtii* genomes, respectively (Additional file 1: Table S7). Approximately 97.54% of *P. hakonensis* and 97.69% of *P. brandtii* predicted genes were successfully annotated using the Non-Redundant Protein Sequence Database (NR), InterProScan, Uniport (Universal

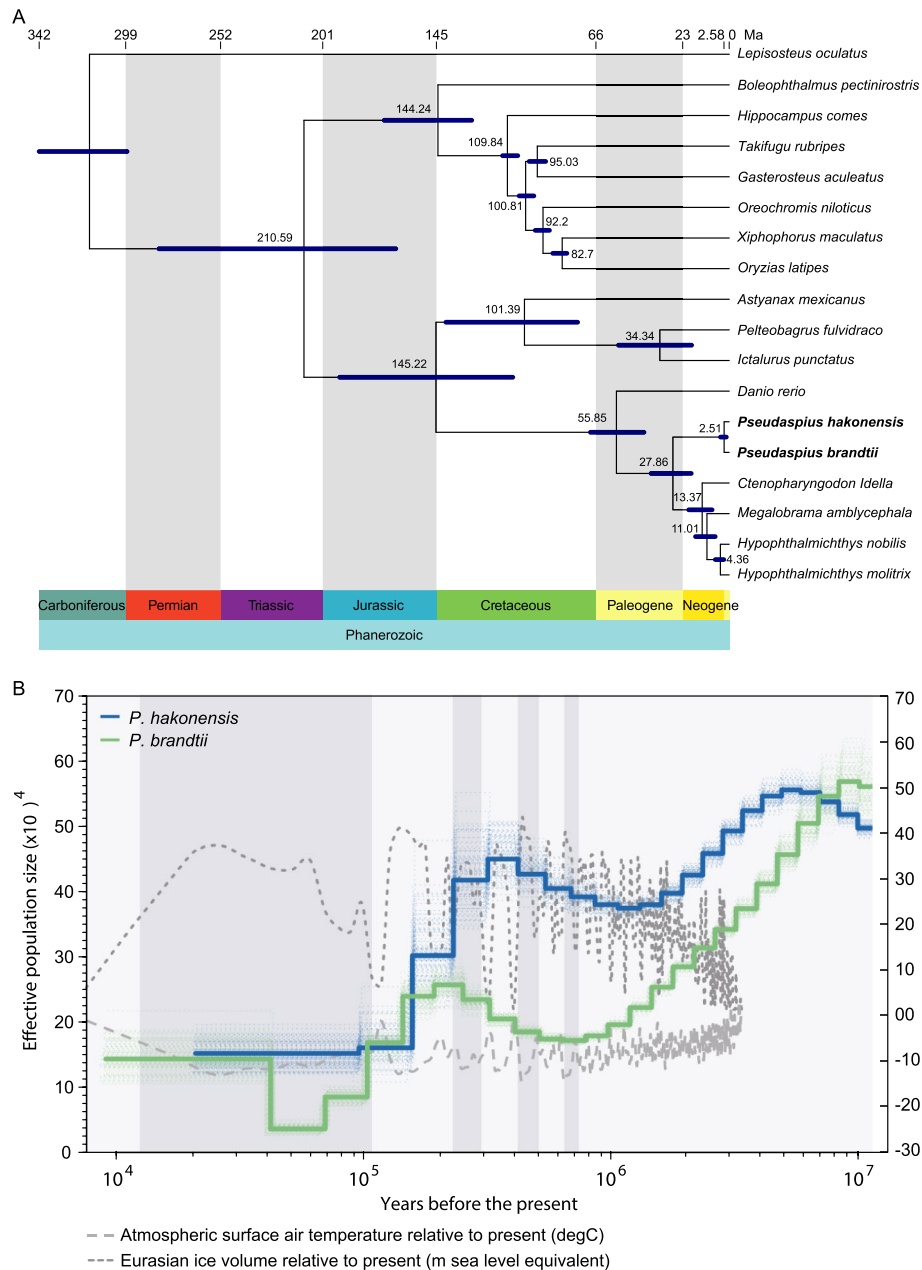
Protein), Pfam, and eggNOG (evolutionary genealogy of genes: Non-supervised Orthologous Groups) databases (Additional file 1: Table S8). To gain further insights into the genomic relationships, we conducted a genome collinearity analysis between *P. hakonensis*, *P. brandtii*, and *Danio rerio* [33]. The results showed that these species exhibited good collinearity without any evidence of chromosomal fission or fusion events (Fig. 1D). This corroborates the high-quality and chromosome-level nature of our assembled genomes for *P. hakonensis* and *P. brandtii*, which provide a solid foundation for our investigations into the genetic basis of seawater adaptation in these two Far Eastern daces.

#### Phylogenetic relationships and population history

We conducted a phylogenetic analysis based on 1597 single-copy orthologous genes to explore the evolutionary relationships and divergence times of *P. hakonensis*, *P. brandtii*, and 16 other teleost fishes (Additional file 2: Figure S2). The time-calibrated phylogeny showed that *P. hakonensis* and *P. brandtii* formed a sister clade to grass carp (*Ctenopharyngodon idella*), bighead carp (*Hypophthalmichthys nobilis*), silver carp (*H. molitrix*), and blunt snout bream (*Megalobrama amblycephala*). The two clades shared a most recent common ancestor

(MRCA) approximately 27.86 million years ago (Ma) during the Palaeogene period (Fig. 2A). The divergence time between *P. hakonensis* and *P. brandtii* was estimated to be 2.51 million years ago (Ma). These results shed light on the evolutionary history and genetic relationships of these two Far Eastern daces and their place within the broader teleost fish lineage.

The population history was reconstructed by calculating the heterozygosity of the genomes of two species using pairwise sequentially Markovian coalescent (PSMC) model [34]. The population demography analysis of the two species exhibited a similar starting point around  $10^4$  to  $10^7$  years ago, but thereafter, they followed divergent trends. These patterns could be correlated with



**Fig. 2** Phylogenetic analysis and demographic histories of *P. hakonensis* and *P. brandtii*. **A** Phylogenetic tree and divergence times estimated for the *P. brandtii*, *P. hakonensis*, and other 16 teleosts. Error bars indicate 95% confidence levels. *P. hakonensis* and *P. brandtii* were highlighted in bold. **B** Demographic histories were reconstructed using PSMC. Blue lines represent *P. hakonensis*, and green lines represent *P. brandtii*. Dark grey box represents glacial periods, and French grey box represents interglacial periods

significant climatic oscillations, including glacial-interglacial cycles and sea level changes. Such environmental factors might have played a crucial role in shaping the population dynamics and distribution of these Far Eastern daces over time (Fig. 2B). We observed that the first population contraction occurred around 6 million years ago (Ma), followed by a remarkable and massive expansion around 0.9 Ma, reaching its peak at approximately 0.6 Ma during the largest Quaternary glaciation (0.8–0.2 Ma). Subsequently, a second contraction phase began. These fluctuations in population size over geological timeframes shed light on the species' responses to major climatic events and offer valuable insights into its evolutionary history and adaptive strategies. Similarly, the size of the *P. brandtii* population decreased around 9 Ma, followed by a population expansion at approximately 0.7 Ma and peaked around 0.4 Ma during the largest Quaternary glaciation (0.8–0.2 Ma). Our analysis suggests that both *Pseudaspius* species populations experienced multiple bottlenecks that correlated with glacial activities. Specifically, the population declines coincided with the onset of warm interglacial periods when deglaciation dramatically reduced the available habitat. Notably, throughout various periods of evolution, the effective population size of *P. hakonensis* remained relatively higher and more stable compared to *P. brandtii*. This observation may be attributed to *P. hakonensis*' robust adaptability and resistance to changing environments. These population dynamics and differences in effective population size provide valuable insights into the responses and resilience of the two *Pseudaspius* species to historical climatic changes and the complex interplay between environmental conditions and their evolutionary strategies.

### Gene family expansion and contraction

The expansion and contraction of gene families can have a significant impact on speciation and environmental adaptation [35]. The analysis of gene family evolution showed that 148 gene families were expanded, while 837 gene families were contracted in the common ancestor of *P. hakonensis* and *P. brandtii*, respectively (Additional file 2: Figure S3). Based on the results, further enrichment analysis was performed using the GO and KEGG databases. The expanded gene families were significantly enriched in 174 GO terms and 16 KEGG pathways (Additional file 2: Figure S4A and S4B), such as protein-DNA complex subunit organization (GO:0071824), nucleosome (GO:0000786), DNA packaging (GO:0006323), chromatin assembly (GO:0031497), DNA conformation change (GO:0071103), natural killer cell mediated cytotoxicity (ko04650), glycosylphosphatidylinositol (GPI)-anchored proteins (ko00537), and CD molecules

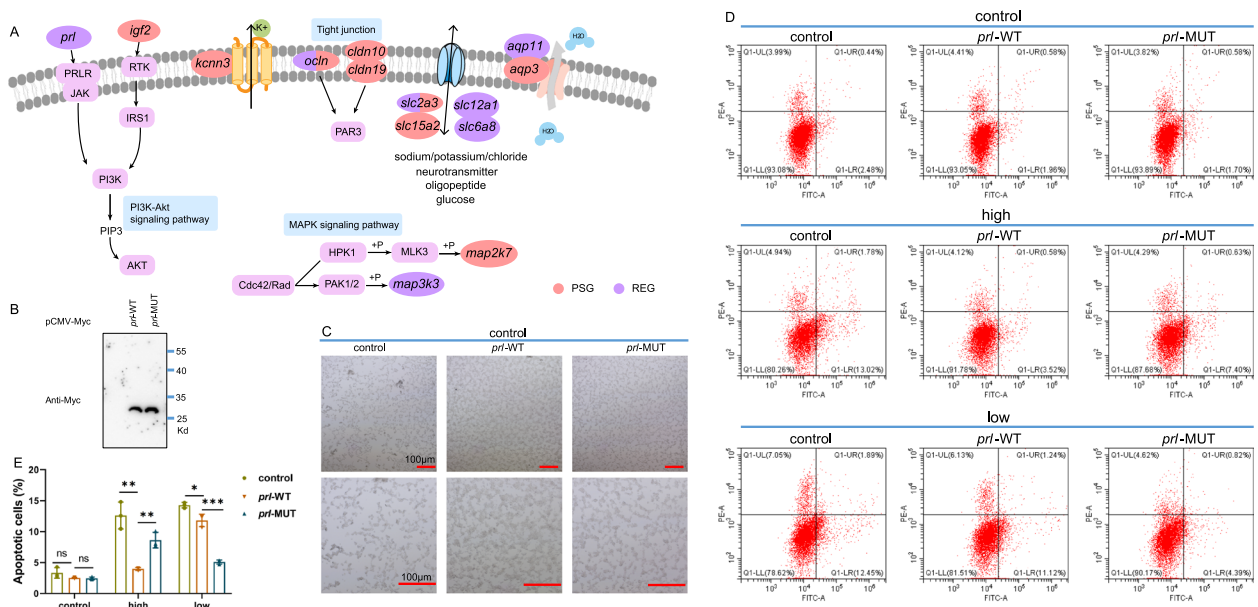
(ko04090) that associated with cell growth and metabolism and cell damage repair.

Conversely, the contraction gene families were significantly enriched in 645 GO terms and 61 KEGG pathways (Additional file 2: Figure S4C and S4D), such as negative regulation of proteolysis (GO:0045861), activation of immune response (GO:0002253), cell adhesion molecules (ko04514), and phagosome (ko04145) that associated with immune response. These results suggested that these biological processes of gene family expansion and contraction in the two *Pseudaspius* species may be associated with their adaptation to the seawater.

### Adaptation to osmoregulation

Migratory fish exhibit the capacity to persist in both freshwater and high salinity seawater environments by employing osmotic regulation mechanisms. These regulatory processes are essential for maintaining cellular homeostasis and achieving internal environmental balance [36]. The branch-site model was employed to identify the genes evolving under positive selection for seawater adaptation. Among the 7194 single-copy orthologous genes across seven species, a total of 355 positively selected genes (PSGs) were identified in the ancestor of *P. hakonensis* and *P. brandtii* genomes (Additional file 2: Figure S5). We found nine PSGs were strongly associated with osmoregulation, including *cldn10*, *igf2*, *slc2a3*, *kcnn3*, *ocln*, *map2k7*, *aqp3*, *slc15a2*, and *cldn19* (Fig. 3A, Table 1, Additional file 1: Table S9), presumably reflecting the seawater adaptation of *P. hakonensis* and *P. brandtii*. Additionally, we detected 704 rapidly evolving genes (REGs) in the ancestor of *P. hakonensis* and *P. brandtii* genomes. Among them, seven genes related to osmoregulation were found, including *ocln*, *slc2a3*, *prl*, *aqp11*, *slc6a8*, *slc12a1*, and *map3k3* (Fig. 3A, Table 2, Additional file 1: Table S9).

Among these important genes, *ocln* and *slc2a3* genes had both rapid evolution and positive selection (Fig. 3A, Table 1 and Table 2). The gene *slc2a3*, namely the facilitative Na<sup>+</sup>-independent sugar transporters, which contributes to osmoregulatory activities in the fish kidney [37], showed positive selection and accelerated evolution. It was found with A42E and V147L specific amino acid site mutations in both *P. hakonensis* and *P. brandtii*, these mutations occurred in the MFS\_1 domain and were likely to have an impact on the three-dimensional structure of the protein (Additional file 2: Figure S6A and S6B). Another gene, *cldn10*, which participates in the creation and maintenance for cation-selective pore junctions in salt-secreting tissues of teleost fishes [38], was found with Y79H and L92T specific amino acid site mutations which occurred in the Claudin\_2 domain in both *P. hakonensis* and *P. brandtii*, and this was also likely to affect the



**Fig. 3** Genomic evidence of osmoregulation related genes in *P. hakonensis* and *P. brandtii*. **A** Genes and pathways involved in osmoregulation are shown in the colored circles and light blue box, respectively. PSGs are marked in pink circles, REGs are marked in purple circles. **B** Immunoblotting assay of *prl* in HEK293T cells. HEK293T cells were cultured in 6-well plates and transfected with indicated plasmids (3 μg each) for 24 h, and then the cells were harvested to perform immunoblotting. **C** *prl*-WT reduces cell crumbling and abscission after high osmotic pressure treatment. HEK293T cells were cultured in 6-well plates and transfected with indicated plasmids (3 μg each) for 8 h and then cultured under high osmolarity condition (600 mOsmol/kg) for 16 h and imaged under a microscope Nikon TE2000-U. **D–E** *prl*-WT reduces cell apoptosis in high osmolarity condition (600 mOsmol/kg) or low osmolarity condition (200 mOsmol/kg) detected by flow cytometry analysis (**D**), and statistical analyses were performed using the GraphPad Prism software (7.0) (**E**). HEK293T cells were cultured in 6-well plates and transfected with the indicated plasmids (3 μg each) for 8 h and then cultured under high osmolarity condition (600 mOsmol/kg) for 16 h (high group) or cultured under high osmolarity condition (600 mOsmol/kg) for 12 h followed by low osmolarity condition (200 mOsmol/kg) for 4 h (low group). Data show mean ± SD; Student’s two-tailed *t* test. \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001, data from three independent experiments

**Table 1** Positively selected genes with osmoregulation of both *P. hakonensis* and *P. brandtii*

Gene id in <i>P. hakonensis</i>	Gene id in <i>P. brandtii</i>	Gene name	Description	<i>P</i> -value	Adjusted <i>P</i> -value
evm.model.Contig196.29	evm.model.Contig8.162	<i>ocln</i>	Occludin	0.002161284	0.027519075
evm.model.Contig117.16	evm.model.Contig11.383	<i>igf2</i>	Insulin-like growth factor 2	0.000854549	0.016889081
evm.model.Contig27.200	evm.model.Contig55.130	<i>slc2a3</i>	Solute carrier family 2, facilitated glucose transporter member 3	0.000650832	0.013650395
evm.model.Contig56.10	evm.model.Contig252.6	<i>cldn10</i>	Claudin-10	0.000285494	0.006418262
evm.model.Contig101.4	evm.model.Contig3.11	<i>aqp3</i>	Aquaporin-3	0	0
evm.model.Contig1.321	evm.model.Contig4.99	<i>slc15a2</i>	Solute carrier family 15 member 2-like	3.99E−07	1.45E−05

**Table 2** Rapid evolution genes with osmoregulation and opsin of both *P. hakonensis* and *P. brandtii*

Gene id in <i>P. hakonensis</i>	Gene id in <i>P. brandtii</i>	Gene name	Description	<i>P</i> -value	Adjusted <i>P</i> -value
evm.model.Contig196.29	evm.model.Contig8.162	<i>ocln</i>	Occludin	0.002161284	0.027519075
evm.model.Contig27.200	evm.model.Contig55.130	<i>slc2a3</i>	Solute carrier family 2, facilitated glucose transporter member 3	4.60E−06	0.000387153
evm.model.Contig5.58	evm.model.Contig13a.236	<i>prl</i>	Prolactin	1.10E−05	0.000741856
evm.model.Contig154.2	evm.model.Contig6c.48	<i>aqp11</i>	Aquaporin-11-like	0.000135301	0.004177491
evm.model.Contig2.156	evm.model.Contig45.24	<i>rh1</i>	Rhodopsin	0.000710964	0.013148265

three-dimensional structure (Additional file 2: Figure S6A and S6C).

The *prl* gene directly contributes to the survival and functionalization of branchial ion cells involved in osmoregulation (Additional file 2: Figure S6D) [39]. To further investigate whether rapid evolution of *prl* (*prolactin*) in *Pseudaspius* is involved in their salinity adaptation, we first cloned and tested the expression of *prl* wild-type (*prl*-WT) and mutant (*prl*-MUT) (Additional file 2: Figure S6D). Immunoblotting showed that both *prl*-WT and *prl*-MUT could be detected in transfected HEK293 cells (Fig. 3B). Given that hypertonicity induced by high bay salt, the main component of which is NaCl, decreases cell volume, increases cytosolic osmolality, and alters mitochondrial osmotic equilibrium (Michea et al., 2002), we analyzed whether *prl* gene prevents the corresponding hypertonic threat. As expected, cells treated with high osmolality (600 mOsmol/kg) showed significant shrinkage and shedding, and then overexpression of *prl*, especially overexpression of *prl*-WT, significantly improved the cell status (Fig. 3C). Previous studies have shown that changes in osmolality induce cells to enter the apoptotic program (Michea et al., 2002; Dmitrieva et al., 2000). To further elucidate the effect of *prl* on osmolality-induced apoptosis of HEK293 cells, flow cytometry analysis was performed. As shown in Fig. 3D, E, apoptosis was significantly induced by high osmolality (600 mOsmol/kg), in contrast to overexpression of *prl*-MUT, overexpression of *prl*-WT significantly inhibited osmolality-induced cell apoptosis. Taken together, the results suggest that the *Pseudaspius prl* gene enhanced cell adaptation to altered osmolality.

### Vision adaptation

In addition to osmotic regulation, freshwater and marine transition also require a suite of adaptations in other physiological and life history traits, including diet, metabolism, defense from predators, reproductive strategies, and even vision (Van Nynatten et al., 2021). The spectra of available light differ significantly between marine and freshwater habitats. Light that penetrates the open ocean is known to shift towards the blue end of the spectrum with increasing depth (Jerlov, 1968). In contrast, the light spectrum in riverine environments attenuates more rapidly, is more spectrally diverse, but has been less explored concerning the visual systems of riverine animals (Levine and MacNichol, 1979; Costa et al., 2013). These distinctions in the aquatic light environment are believed to impact the visual systems of fishes, leading to freshwater fishes generally exhibiting more redshifted visual pigments (Rennison et al., 2016; Musilova et al., 2019).

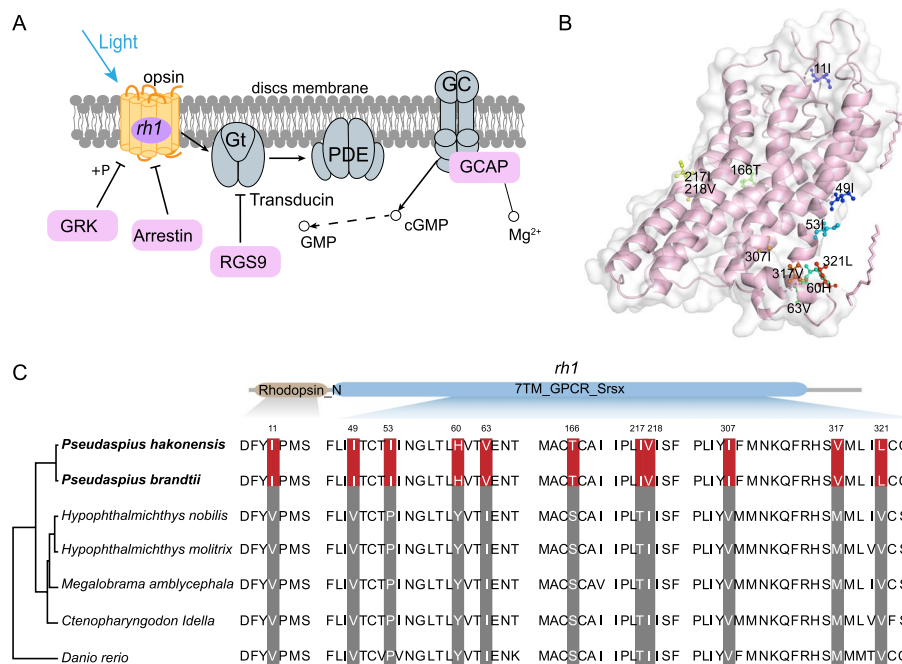
Rhodopsin (*rh1*), the light-sensitive visual pigment expressed in rod photoreceptors, is specialized for vision

in dim-light environments and has been confirmed to be associated with marine to freshwater croaker invasion (Van Nynatten et al., 2021) and was also reported to play an important role in migratory eels [40, 41]. Our results showed that the *rh1* gene involved in phototransduction cascade exhibited a rapid evolutionary rate that may have implications in adapting to distinct light environments (Fig. 4A, Table 2). Subsequently, *rh1* gene underwent 11 specific amino acid replacements which occurred in the Rhodopsin\_N and 7TM\_GPCR\_Srsx domains in both *P. hakonensis* and *P. brandtii*, and these amino acid mutation sites also located in the key position of the functional domain (Fig. 4B, C). Therefore, the accelerated evolution of the *rh1* gene is likely to be closely associated with the diverse light conditions encountered during the migration of *P. hakonensis* and *P. brandtii*.

### Conserved non-coding elements (CNEs) linked to osmoregulation

CNEs are usually distributed in the upstream and downstream regions of genes and served as the core components of gene regulatory networks to regulate gene expression [42]. To determine the extent of CNEs loss or gain in *P. hakonensis* and *P. brandtii*, we predicted genome wide CNEs with reference to the zebrafish genome. Consequently, we successfully identified 18,581 CNEs across 11 teleosts through syntenic alignments. We further analyzed the CNEs that were specifically deleted or inserted in *P. hakonensis* and *P. brandtii* and found that they were present in the neighborhood of 175 genes with 196 CNE sequences, including *atg7* and *usp45* genes related to osmoregulation. In particular, there was one CNE (~160 bp) near *atg7* with an 8-bp specific deletion in the *P. hakonensis* and *P. brandtii* (Fig. 5A, C) and another CNE (~180 bp) near *usp45* with a 7-bp specific insertion in the *P. hakonensis* and *P. brandtii* (Fig. 5B) using zebrafish annotations as reference, which may be associated with adaptation to seawater.

To further explore whether the *Pseudaspius*-specific CNEs near *atg7* and *usp45* genes were involved in osmoregulation, we inserted these CNEs into the pGL3-Basic vector and co-transfected them into HEK293T cells, which were cultured under control (315 mOsmol/kg) or high osmolality conditions (600 mOsmol/kg). Then, their promoter activity was determined using the dual-luciferase reporter assay system. The results of our reporter assay suggest that the upstream fragment of *atg7* and *usp45* from *P. brandtii* and *P. hakonensis* exhibit higher promoter activity compared to corresponding fragments from zebrafish (Fig. 5D, E). Notably, the promoter activity of *P. brandtii* and *P. hakonensis atg7* was significantly induced at high osmolality (Fig. 5D). Overall, these results demonstrated that the *Pseudaspius*-specific



**Fig. 4** Genomic evidence of visual opsin genes in *P. hakonensis* and *P. brandtii*. **A** Genes involved in phototransduction cascade are shown in the colored circles; a REG was marked in purple circle. **B** The three-dimensional structure of RH1 protein. **C** The *rh1* gene had specific AA replacement in *P. hakonensis* and *P. brandtii* based on sequence alignments

CNEs near *atg7* and *usp45* genes played a significant role in osmoregulation.

## Discussion

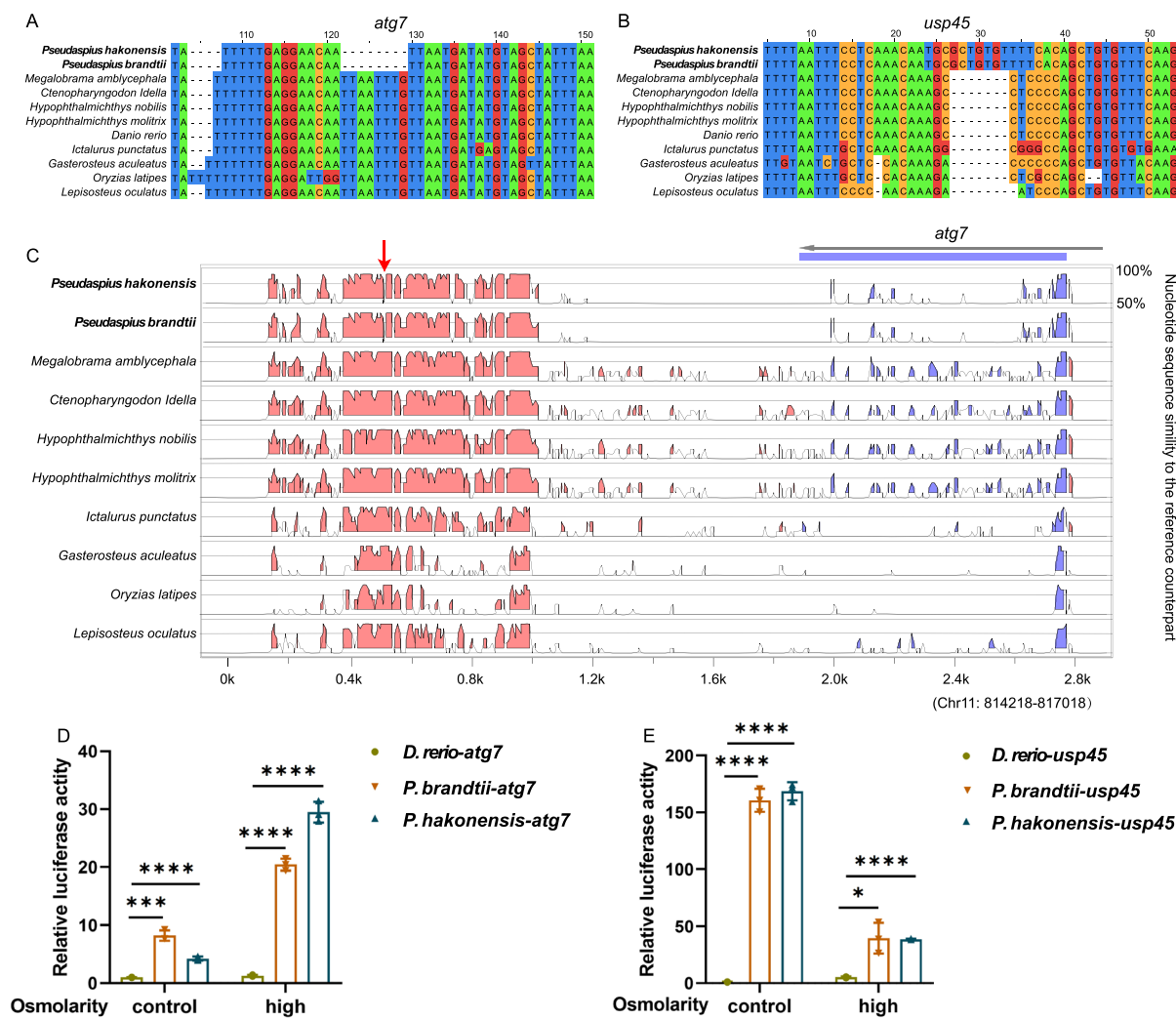
Two of the few migratory species cyprinids, *P. hakonensis* and *P. brandtii*, are ideal fish species to investigate the genetic basis of adaptation to seawater [21, 43]. With advancements in sequencing technology, comparative genomics studies leverage available data to reveal the evolutionary patterns of adaptive traits in fishes in special environmental conditions [44–46]. Here, we generated chromosomal level genome assemblies of *P. hakonensis* and *P. brandtii* with 840.77 Mb and 848.49 Mb, respectively. Both of *P. hakonensis* and *P. brandtii* were anchored with 25 chromosomes and had good collinearity with zebrafish. Analysis of population history showed that two species formed sister groups with *C. idella*, *H. nobilis*, *H. molitrix*, and *M. amblycephala* in Cyprinidae. The genomic information of two species will contribute to a deeper understanding of the genetic mechanism of seawater adaptation.

Surviving in diverse aquatic environments requires ability to regulate their osmotic pressure balance, since salinity plays a crucial role in the ecological environment [47]. Previous studies investigated the biological process genes including metabolism, immune system, cell activity, and growth potentially associated with Nile tilapia

adaptation to different salinities as well as phenotypic differentiation of morphology, physiology, and behavior between the migratory population and the freshwater resident population of *Coilia nasus* in the process of adaptive evolution [48, 49]. We hypothesized that these biological processes of gene family expansion and contraction in the two *Pseudaspius* species might be linked to the physiological and behavioral changes observed in the two *Pseudaspius* species as they adapted to the evolution of seawater.

In previous studies, hatching enzymes associated with the evolution of salinity adaptation has been identified in a whole genome of *Oryzias javanicus*, revealing the genetic mechanism of adaptation to different salinities [32]. Similarly, the Atlantic sturgeon has the physiological capability to move between salinity habitats with varying salinity. Tests on its growth and osmotic regulation have indicated its adaptability to these changing salinity conditions. [50]. In the current study, our analyses yielded a list of adaptively evolving genes, which comprised 355 PSGs and 704 REGs in the ancestor node of *P. hakonensis* and *P. brandtii*. Notably, several genes related to osmotic regulation were identified as PSGs and REGs. The candidate genes *ocln*, *cldn10* and *cldn19* encoded the components of a tight connection (Fig. 3A). Previous studies have provided evidence identifying the *ocln* gene as an osmoregulation-related gene in *Acipenser baerii* [51], while the





**Fig. 5** CNE sequence analysis that may be related to osmoregulation. **A** An 8-bp specific deletion of *P. hakonensis* and *P. brandtii* was located near target gene *atg7*. **B** A 7-bp specific insertion of *P. hakonensis* and *P. brandtii* was located near target gene *usp45*. **C** VISTA sequence conservation plot of the *P. hakonensis* and *P. brandtii* specific deletion CNE around *atg7*, using zebrafish as reference. Red arrows indicate missing bp. **D, E** Luciferase activity assays of the fragments of *atg7* (**D**) and *usp45* (**E**) promoter from *Danio rerio*, *P. brandtii*, and *P. hakonensis* under control (315 mOsmol/kg) or high osmolarity conditions (600 mOsmol/kg) in HEK293 cells. \* $P < 0.05$ , \*\* $P < 0.01$ , \*\*\* $P < 0.001$ , \*\*\*\* $P < 0.0001$

*cldn19* gene played significant roles in osmoregulatory physiology of *Petromyzon marinus* [52]. Moreover, the PSG *cldn10*, which predominantly expressed in osmoregulatory organs of teleost fishes [38, 53], was found have two specific AA replacements in the two *Pseudaspius* species (Additional file 2: Figure S6A). In addition, osmotic regulation affected the expression of many SLCs (solute carriers), such as *slc15a2*, *slc6a8*, *slc12a1*, and *slc2a3* [24, 54–56]. For instance, *slc2a3* gene was important in the fish kidney for osmoregulatory activities [56]. Here, *slc15a2*, *slc6a8*, *slc12a1*, and *slc2a3* genes were involved in material (such as sodium/potassium/chloride, neurotransmitter, oligopeptide, and glucose) exchange

(Fig. 3A). Importantly, *slc2a3* is identified as a both positively selected and rapidly evolving gene that had two specific amino acid replacements in the MFS\_1 domain (Additional file 2: Figure S6A and 6B). Meanwhile, several ion transporters and genes were also important in osmoregulation, such as *kcnn3* that has been found to be upregulated in an euryhaline pupfish adaptation to high salinity conditions [57] (Fig. 3A).

Aquaporins (AQPs) are crucial for rapid osmoregulation by facilitating the diffusion of water and osmolytes across cellular membranes [58]. Recent studies suggested that the *aqp3* gene potentially contributes to water metabolism and osmotic pressure regulation in

*Lateolabrax maculatus* [59]. Moreover, hypertonic conditions have resulted in increased expression of *aqp11* in *Clarias magur* hepatocytes [60, 61]. In particular, our genomic analyses showed that *aqp3* and *aqp11* genes may modulate the exchange of water molecules between the inside and outside of the cell membrane (Fig. 3A). We also found *igf2* to be under selection, which influences osmoregulation by stimulating Na<sup>+</sup>/K<sup>+</sup>-ATPase activity and promoting chloride cell differentiation (Chandhini et al., 2021) (Fig. 3A). A *prl* gene has been showed to have a reduced expression level following seawater adaptation in *Oncorhynchus rhodurus* (Yada et al., 2010) (Fig. 3A). The *prl* and *igf2* genes bind to receptors and transmit signals to PI3K, participating in the PI3K-Akt signaling pathway (Fig. 3A). It has been demonstrated that Mitogen-activated protein kinase (MAPK) cascades are major participants in osmotic stress response of fish, and are involved in regulating cell proliferation, growth, death, and other physiological processes [62–64]. The expression of the *map2k7* gene has been showed to be upregulated in the muscle of *Penaeus monodon* during acute hyposaline stress, and *map3k3* also contributes to the activation of the osmoprotective transcription factor *nfat5* in response to high NaCl [65, 66]. The genes, in particular *map2k7* and *map3k3*, part of MAPK signaling pathway, were identified as PSG and REG in the two *Pseudaspius* species, respectively (Fig. 3A). The role of PRL in osmoregulation has been discovered as early as 1950s by researching on killifish (*Fundulus heteroclitus*), with *F. heteroclitus* unable to survive in freshwater following hypophysectomy while PRL treatment allowed hypophysectomized killifish to survive [67]. Subsequently, several studies found a direct contribution of PRL to osmoregulation by knocking out this gene in zebrafish (Shu et al., 2016). However, it is still unclear whether the rapid evolving of *prl* gene in two *Pseudaspius* species contribute to their adaptation to seawater. In our study, by integrating comparative genomics and functional assays, we found that the rapid evolving of *prl* gene is indeed associated with their osmoregulation in the two *Pseudaspius* species. Osmoregulatory candidate genes are potentially important that contribute to the adaptation of *Pseudaspius* species to seawater.

Previous studies demonstrated that visual systems have undergone diversification in migratory fish [68]; freshwater eels changed their spectral sensitivity by switching chromophore types or expressing different opsin genes in response to changing light environments during migration from fresh water to deep sea environments [68, 69]. Moreover, cloning and expression analysis of freshwater and deep-sea opsin genes in the eel *A. japonica* have revealed that the rod-shaped visual pigment (*rh1*) changes from “freshwater” form to “deep-sea” form

during maturation [70, 71]. The *rh1* underwent positive selection during salmonid evolution, linked to divergent light environments [72]. Interestingly, our genomic analyses found that *rh1* is a rapidly evolving gene with several amino acid replacements (Fig. 4). Therefore, the genetic innovations may be associated with the seawater light environment of *P. hakonensis* and *P. brandtii*.

Many CNEs function as cis-regulatory elements, such as enhancers, repressors, and insulators, and the presence or absence of CNEs plays a significant role in phenotypic differences and morphological innovation [73–75]. Notably, our genomic analyses found that there were CNEs loss near the *atg7* and obtain close to *usp45* (Fig. 5). Thus, we propose CNEs play pivotal roles in osmoregulation during seawater adaptation of *P. hakonensis* and *P. brandtii*. Previous reports have indicated adverse effects of hypotonic stress on various developmental stages of *Macrobrachium rosenbergii* larvae; the expression of the autophagy-related gene *atg7* showed significant changes under hypotonic stress [76]. In salinity-induced transcriptome studies of marine and freshwater threespine stickleback, a decrease in *usp45* expression was observed 6 h after the sudden exposure to seawater [77]. Notably, our genomic analyses found that there were CNEs loss near the *atg7* and obtain close to *usp45* (Fig. 5). Thus, we propose that this may play pivotal roles in osmoregulation and will yield insights into seawater adaptation of *P. hakonensis* and *P. brandtii*.

## Conclusions

We acquired high-quality chromosome-level genome assemblies of *P. hakonensis* and *P. brandtii*. Genome evolution analysis provided insights into genetic changes related to osmoregulation and vision that may be critical to the seawater adaptation in *P. hakonensis* and *P. brandtii*. Our results provide opportunities to elucidate the molecular mechanisms underlying the evolutionary adaptations of migratory fish to seawater and present a valuable resource for future research on adaptive evolution in *Pseudaspius* species.

## Methods

### Sample DNA and RNA extraction, sequencing

*P. hakonensis* individual and *P. brandtii* individual were collected from Suifen River in Dongning City, Heilongjiang Province, China. Genomic DNA was extracted from the muscle using Puregene Tissue Core 81 Kit A (Qiagen, Maryland, USA). DNA integrity was checked using 1% agarose gel electrophoresis and Pultton DNA/Protein Analyzer (Plextech, Berkeley, USA). Total RNA was extracted from tissues of *P. hakonensis* and *P. brandtii*, including muscle, eye, heart, liver, kidney, gill, spleen, spermary, swim bladder, and brain, by using TRIzol

reagent (Invitrogen, USA). RNA quality was checked with a NanoDrop ND-1000 spectrophotometer (Labtech, Palaiseau, France) and a 2100 Bioanalyzer (Agilent Technologies, Waldbronn, Germany).

A paired-end library with inset sizes of 150 bp was constructed according to the Illumina protocol and sequenced on the NovaSeq 6000 platform (Illumina) to assess the complexity of the genome, correction of the genome assembly, and assembly evaluation. Hi-C libraries were constructed and sequenced to obtain a chromosomal-level genome assembly. After reads with low-quality bases (reads with more than 10% N bases or low-quality bases  $\leq 5$ ), adapter sequences, and duplicated sequences were discarded, the clean reads were used for subsequent analysis. Nanopore libraries were generated according to the manufacturer's instructions and sequenced on 13 flow cells using the GridION X5 DNA sequencer (Oxford Nanopore). For the Nanopore reads, the mean quality of each read was calculated, and any reads with mean quality  $> 7$  were retained for subsequent analyses. An SMRTbell library with a fragment size of 20 kb was constructed using a SMRTBell template preparation kit 1.0 (PacBio) by following the manufacturer's protocol, and the library was sequenced with the PacBio Sequel II system. Then, full-length refined consensus transcripts were generated using the PacBio Iso-Seq pipeline (<https://github.com/PacificBiosciences/IsoSeq>).

#### Genome size estimation and genome assembly

All the filtered Illumina pair-end reads were used for k-mer frequency analysis [78]. The genomic size was estimated on the basis of the following formula:  $G = k\text{-mer number} / k\text{-mer depth}$ , where  $G$  represents the genome size, where k-mer number are the total number, and k-mer depth represents the average depth of 17-mers, respectively. After the Nanopore low-quality reads were filtered, the remaining high-quality reads were first calibrated using the CANU (version 1.5) [79] and then assembled using the genome assembler FLYE (version 2.6) [80] with default parameters. The original Nanopore reads were polished for three rounds using racon (version 1.2.1) [81] and further corrected for three rounds using Illumina reads with pilon (version 1.21) [82] for a consistent assembly. The filtered Hi-C reads were first aligned to the assembled genome using the BOWTIE (version 2.2.5) [83]; only the read pairs with both ends uniquely aligned to the genome were selected. Then, the corrected contigs and valid Hi-C reads were used to perform the chromosomal-level genome assembly by LACHESIS [84] with default parameters. In addition, a heatmap based on the interaction signals was constructed to evaluate the quality of the chromosomal-level genome assembly. Finally,

the completeness of our final genome assembly was evaluated using BUSCO (version 3) [85].

#### Genome annotation

In order to predict the repeat elements in *P. hakonensis* and *P. brandtii*, REPEATMODELER (version 1.0.5) [86] was used to construct a de novo transposable element library, which was then used to predict repeats with REPEATMASKER (version 4.0.6) [86]. In the *P. hakonensis* and *P. brandtii* genomes, ab initio, homologues, and RNA-sequencing methods were used for predicting the protein-coding genes. The above three gene sets were integrated to yield a comprehensive and nonredundant gene set using EVIDENCEMODELER (EVM, version 1.1.1) [87]. Then, the integrated gene set was translated into amino acid sequences and performed gene functional annotations by searching against known databases, including NR, InterPro, Uniport, Pfam, and eggNOG, with a cutoff  $E$  value of  $1e-5$ .

#### Phylogenetic analysis

Protein sequences from a total of 18 species including *P. hakonensis* and *P. brandtii* (assembled in this study), *C. Idella* [88], *H. nobilis*, *H. molitrix* [89], *M. amblycephala* [90], *D. rerio*, *Astyanax mexicanus* [91], *Ictalurus punctatus* [92], *Pelteobagrus fulvidraco* [93], *Gasterosteus aculeatus* [94], *Takifugu rubripes* [95], *Oreochromis niloticus* [96], *O. latipes* [97], *Xiphophorus maculatus* [98], *Hippocampus comes* [99], *Boleophthalmus pectinirostris* [100], and *Lepisosteus oculatus* [101] were constructed by ORTHOMCL (version 2.0.9) with default settings, and 1597 single-copy orthologues were identified. Then, the protein sequences of single-copy orthologous genes were further aligned using MUSCLE [102] with default parameter settings. The maximum likelihood method was used to construct a phylogenetic tree using RAXML (version 8.1.19) [103] with 1000 bootstraps under the GTR+I model. MCMCTREE of the PAML package [104] was used to calculate the divergence times among *P. hakonensis* and *P. brandtii* and 16 other teleost fishes, and the fossil records were obtained from TIMETREE database (<http://www.timetree.org/>) for calibration; we set four calibration time points (*C. Idella*–*L. oculatus*: ~298.8–342.5 MYA; *C. Idella*–*D. rerio*: ~41.7–68.9 MYA; *H. comes*–*G. aculeatus*: ~99.5–116.7 MYA; *O. niloticus*–*X. maculatus*: ~81.2–91.6 MYA).

#### Demographic history

We concluded the demographic history for the *P. hakonensis* and *P. brandtii* by pairwise sequentially Markovian coalescent (PSMC) analysis [34]. Heterozygote sites across the genome were used to estimate changes in effective population size. To generate consensus diploid

sequences for the two individuals separately by SAMtools (version 1.3.1) [105], the “fq2psmcfa” tool was used to create the input file for PSMC modeling, with the option -q20. Moreover, the effective population history was inferred using PSMC with 100 bootstraps and plotted using the `psmc_plot.pl` pipeline. The generation time for two species was set as 4 years, and the mutation rate was  $1.3387e-09$  per year per nucleotide for *P. hakonensis* and  $1.2125e-09$  per year per nucleotide for *P. brandtii*. In addition, we obtained the atmospheric surface air temperature (°C) and Eurasian ice volume (m sea level equivalent) data for the past 3 million years from the NCEI (<http://www.ncdc.noaa.gov/>).

### Gene family expansion and contraction

To determine gene family expansion and contraction in the *P. hakonensis* and *P. brandtii* genomes, CAFÉ (version 3.1) [106] was used with default parameter settings. If the copy number of the *P. hakonensis* and *P. brandtii* were higher or lower than those found in their close ancestral clades, then we identified the gene family as a substantially expanded or contracted gene family. A conditional *P*-value for each gene family was calculated, and gene modules of the significantly expanded and contracted gene family were extracted and were subjected to GO and KEGG functional enrichment analyses with a false discovery rate (FDR) adjusted *P*-value < 0.05, respectively. During functional enrichment, the zebrafish genes were selected as background. GO terms or KEGG pathways with a *P*-value < 0.05 were considered significantly enriched.

### Identification of positively selected genes and rapidly evolving genes

Single-copy orthologous genes were used in the detection of PSGs and REGs. The branch-site model incorporated in the PAML (version 4.8) package [107] was employed to detect PSGs in *P. hakonensis* and *P. brandtii* genomes, with the non-migrated fishes as the background. The null model used in the branch-site test assumed that the comparison of the substitution rates at nonsynonymous and synonymous sites (Ka/Ks ratio) for all codons in all branches must be  $\leq 1$ , whereas the alternative model assumed that the foreground branch included codons evolving at  $Ka/Ks > 1$ . The *P*-values were computed based on the chi-square statistic adjusted by the false discovery rate (FDR) method, and genes with adjusted *P*-value < 0.05 were considered as PSGs. We identified REGs in the *P. hakonensis* and *P. brandtii* genomes using the branch model in the CODEML program in PAML. The likelihood ratio test (LRT) was employed to discriminate between alternative models for each orthologue with *df* (degree of freedom) = 1. We treated the orthologue as

evolving with a significantly faster rate in foreground branch if the FDR adjusted *P*-value < 0.05 and a higher  $\omega$  (Ka/Ks) ratio in the foreground branch than in the background branches.

### Identification of specific amino acid mutations and protein structure simulation

The homologous protein sequences of *P. hakonensis* and *P. brandtii* and other fishes were aligned and compared. A particular amino acid mutation was defined as the *P. hakonensis* and *P. brandtii* amino acid were different from acids of all other species including *C. Idella* [88], *H. nobilis*, *H. molitrix* [89], *M. amblycephala* [90], and *D. rerio*. Missing sites may be present in the protein sequences of species with poorly assembled genomes because of assembly quality among different genomes. Therefore, the specific amino acid mutation was considered only if at least 80% of species had valid amino acid information. Then, it was determined whether the mutated amino acids located in the protein domain of the *P. hakonensis* and *P. brandtii* using Pfam [108].

First, we preliminarily predicted the *slc2a3*, *cldn10*, and *rh1* protein structures based on the amino acid sequence on the SWISS-MODEL online website and downloaded the PDB file (<https://swissmodel.expasy.org/>). Next, the predicted result with the maximal score was selected as the best structure and visualized by Pymol (<http://www.PyMOL.org/>).

### Identification of conserved noncoding elements

To identify CNEs, using the *D. rerio* genome as a reference, whole genome alignments for 11 teleosts were generated, and the genome data were downloaded from ENSEMBL; they were *P. hakonensis* and *P. brandtii* (assembled in this study), *C. Idella*, *H. nobilis*, *H. molitrix*, *M. amblycephala*, *D. rerio*, *I. punctatus*, *G. aculeatus*, *O. latipes*, and *L. oculatus*. First, the results of the LAST whole genome alignment were submitted to the LAST subroutine “Maf-swap” [109], and the sequence in the maf format alignment was changed; then, the MULTIZ v3 subroutine “roast” [110] was used to extract sequence alignments at the same location on the chromosomes of different species. Next, further analyses were run with subprograms of PHAST [111]. “msa\_view” was used to identify aligned 4D sites of alignments with parameters “-informat MAF -4d -features.” We estimated a neutral phylogenetic model using “phyloFit” [112]. We estimated rho (expected substitution rate of conserved elements relative to neutrality) using “phastCons” with “-estimate-rho -no-postprobs -most-conserved -score.” Conserved models for each window were combined by phyloBoot [112] and then used for initial conserved element prediction. The coding regions and non-coding regions of each

gene based on the GFF3 file of the *D. rerio* genome were separated by using ANNOVAR [113] after conserved elements had been identified. Finally, a series of python scripts were used to detect the divergent CNEs and divergent coding genes. A final set of 18,581 CNEs with a minimum of 30-bp in length were identified. We mainly focused on CNEs which are located in the intergenic regions, in introns, and within the 10-kb upstream or downstream flanking regions of genes. Finally, the VISTA [114] was used to plot the alignments after determining the required CNE sequence.

### Detection of apoptotic cells

HEK293T cells were cultured under different osmolarity conditions as indicated. HEK293T cells were cultured in 6-well plates and transfected with indicated plasmids (3 µg each) for 8 h and then cultured under high osmolarity condition (600 mOsmol/kg) for 16 h and imaged under a microscope Nikon TE2000-U. For flow cytometry analysis, cells were harvested and stained with FITC Annexin V and PI using the FITC Annexin V Apoptosis Detection Kit I (556,547, BD Pharmingen) according to the manufacturer's instructions. The detection of apoptotic cells was performed using the Beckman CytoFLEXS, and the data were analyzed using the CytExpert software.

### Dual-luciferase assay

The chemically synthesized fragments of specific CNEs near *Danio rerio atg7* (NC: 007122.7\_814119-815,755), *P. brandtii atg7* (Chromosome 21: 1,036,539–1038266) and *P. hakonensis atg7* (Chromosome 21: 42,224–43,950), *D. rerio usp45* (NC: 007127.7\_32717090-32718855), *P. brandtii usp45* (Chromosome 10: 21,314,205–21315977), and *P. hakonensis usp45* (Chromosome 10: 20,933,460–20935232) were inserted into the pGL3-Basic vector (Promega, USA) and were sequenced (Sangon Biotech, China). The HEK293T cells were seeded in 24-well plates for 24 h and then co-transfected with 200 ng constructed pGL3-Basic vectors and 3 ng pRL-CMV as an internal control using VigoFect (Vigorous Biotech). Sixteen hours after transfection, cells were cultured under control (315 mOsmol/kg) or high osmolarity conditions (600 mOsmol/kg) for 8 h and then harvested using passive lysis buffer. Luciferase activity in cell extracts was determined using a dual-luciferase reporter assay system (Promega) according to the manufacturer's protocol. The light output of the transcription activity was divided by the output of the Renilla luciferase activity for sample normalization.

### Statistical analysis

All statistical analyses were performed using the GraphPad Prism software (7.0). Results with error bars represent mean ± SD. Statistical analysis was performed

using Student's two-tailed *t*-test. A *P*-value less than 0.05 was considered significant. Statistical significance is indicated as follows: \**P* < 0.05, \*\**P* < 0.01, \*\*\**P* < 0.001, \*\*\*\**P* < 0.0001.

### Abbreviations

CNEs	Conserved non-coding elements
Gb	Gigabase
Mb	Megabase
BUSCO	Benchmarking Universal Single-Copy Orthologs
NR	Non-Redundant Protein Sequence Database
MRCA	Most recent common ancestor
Ma	Million years ago
PSMC	Pairwise sequentially Markovian coalescent
GO	Gene Ontology
KEGG	Kyoto Encyclopedia of Genes and Genomes
PSGs	Positively selected genes
REGs	Rapidly evolving genes
WT	Wild-type
AQPs	Aquaporins

### Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12915-024-01885-2>.

**Additional file 1: Table S1.** The statistics of sequencing of *P. hakonensis* and *P. brandtii* genome. **Table S2.** The statistics of the assembly of *P. hakonensis* and *P. brandtii* genome. **Table S3.** Comparison of the genome assembly quality from fishes in Leuciscidae. **Table S4.** Statistics of the pseudochromosome assemblies using Hi-C data. **Table S5.** BUSCO analysis result of *P. hakonensis* and *P. brandtii* genome. **Table S6.** Repeat elements in the *P. hakonensis* and *P. brandtii* genome with Repeatmasker. **Table S7.** Genetic structure characteristics of the *P. hakonensis* and *P. brandtii* compared to other fish species. **Table S8.** Gene function annotation of *P. hakonensis* and *P. brandtii*. **Table S9.** Positively selected genes and rapid evolution genes with osmoregulation of both *P. hakonensis* and *P. brandtii*.

**Additional file 2: Figure S1.** Genome-wide Hi-C heatmap of *P. hakonensis* (A) and *P. brandtii* (B). The blocks represent the 25 pseudochromosomes. The color bar illuminates the contact density from high (red) to low (white) in the plot. **Figure S2.** Comparison of gene family clusters. The horizontal axis display species and the vertical axis display genes number. **Figure S3.** Expanded/contracted gene families for *P. hakonensis* and *P. brandtii* and another 16 teleosts with the divergence time. **Figure S4.** GO and KEGG enrichment analysis of expanded/contracted gene families for *P. hakonensis* and *P. brandtii*. (A) Top 20 go\_term of expanded gene families for *P. hakonensis* and *P. brandtii*. (B) Top 16 KEGG pathway of expanded gene families for *P. hakonensis* and *P. brandtii*. (C) Top 20 go\_term of contracted gene families for *P. hakonensis* and *P. brandtii*. (D) Top 20 KEGG pathway of contracted gene families for *P. hakonensis* and *P. brandtii*. **Figure S5.** Phylogenetic tree for positive selective genes analysis. **Figure S6.** Genomic evidence of osmoregulation related genes in *P. hakonensis* and *P. brandtii*. (A) The *slc2a3* and *cldn10* genes had specific AA replacement in *P. hakonensis* and *P. brandtii* based on sequence alignments. (B) The three-dimensional structure of SLC2A3 protein. (C) The three-dimensional structure of CLDN10 protein. (D) Alignment of PRL amino acid sequences showing specific amino acid mutations in the *Pseudaspius prl* gene.

**Additional file 3.** Uncropped blots.

### Acknowledgements

This research was supported by the Wuhan Branch, Supercomputing Center, Chinese Academy of Sciences, China. We are grateful to Junhao Huang, Zhuocheng Zhou, and Dr. Fan Li for their help with sampling and fish image. We thank Prof. Martin J. Genner from University of Bristol with the English editing.

**Authors' contributions**

S.H. lead the study. L.Y., and Y.W. designed research. X.Z., C.W., Y.Q., M.Z., and M.M. performed data analyses. F.X., W.C., Z.D., D.Y., Y.L., Y.C., and L.Y. collected the sample. J. W. performed the experiments. X.Z., Y.W., and L.Y. wrote the paper with input from all authors. All authors read and approved the final manuscript.

**Funding**

This work was supported by the National Natural Science Foundation of China (32170480 and 31972866), Youth Innovation Promotion Association, Chinese Academy of Sciences (<http://www.yicas.cn>), and the Young Top-notch Talent Cultivation Program of Hubei Province to L.Y. This work was also supported by the National Natural Science Foundation of China (Grant number 31702016), Jiangnan University Research Project Funding Plan (Grant number 2023KJZX42), and CSC scholarship to Y.W.

**Availability of data and materials**

Raw sequencing reads have been deposited in NCBI with the BioProject accession PRJNA980574 [115]. Uncropped blots are available in Additional file 3.

**Declarations****Ethics approval and consent to participate**

All experimental protocols were approved by the ethics committee of Institute of Hydrobiology, Chinese Academy of Sciences (Approval No.: Y21304506).

**Consent for publication**

Not applicable.

**Competing interests**

The authors declare that they have no competing interests.

Received: 11 December 2023 Accepted: 11 April 2024

Published online: 19 April 2024

**References**

- Chapman BB, Skov C, Hulthén K, Brodersen J, Nilsson PA, Hansson L, Brnmark C. Partial migration in fishes: definitions, methodologies and taxonomic distribution. *J Fish Biol.* 2012;81(2):479–99.
- Dodson JJ, Aubin-Horth N, Thériault V, Páez DJ. The evolutionary ecology of alternative migratory tactics in salmonid fishes. *Biol Rev Camb Philos Soc.* 2013;88(3):602–25.
- Gross MR. Evolution of diadromy in fishes. *Common Strat Anadrom Catadrom Fishes.* 1987;1:14–25.
- Mcdowall RM. The evolution of diadromy in fishes (revisited) and its place in phylogenetic analysis. *Rev Fish Biol Fish.* 1997;7(4):443–62.
- Luo YP. Research advances in energetic changes during fish migration: a review. *Fish Sci.* 2012;31(6):375–81.
- Rachmawati F. Energy content and viscerasomatic index of eel, *Anguilla bicolor McClelland*, exposed in different salinity. *Proceeding of International Conference of Aquaculture Indonesia (ICAI) 2012:76–79.*
- Alò D, Lacy SN, Castillo A, Samaniego HA, Marquet PA, Tittensor D. The macroecology of fish migration. *Glob Ecol Biogeogr.* 2020;30(1):99–116.
- Chadwick E, Mcdowall RM. Diadromy in fishes: migrations between freshwater and marine environments. *Copeia.* 1990;1990(1):271.
- Mcdowall RM. A recent marine ancestry for diadromous fishes? Sometimes yes, but mostly no! *Environ Biol Fishes.* 1993;37(4):329–35.
- Härmä M, Lappalainen A, Urho L. Reproduction areas of roach (*Rutilus rutilus*) in the northern Baltic sea: potential effects of climate change. *Can J Fish Aquat Sci.* 2008;65(12):2678–88.
- Vetemaa M, Eschbaum R, Verliin A, Albert A, Eero M, Lillemägi R, Pihlak M, Saat T. Annual and seasonal dynamics of fish in the brackish-water Matsalu Bay. *Estonia Ecol Freshw Fish.* 2006;15(2):211–20.
- Skovrind M, Olsen MT, Vieira FG, Pacheco G, Carl H, Gilbert MTP, Møller PR. Genomic population structure of freshwater-resident and anadromous ide (*Leuciscus idus*) in north-western Europe. *Ecol Evol.* 2016;6(4):1064–74.
- Gu J, Dai S, Liu H, Cao Q, Yin S, Lai KP, Tse WKF, Wong CKC, Shi H. Identification of immune-related genes in gill cells of Japanese eels (*Anguilla japonica*) in adaptation to water salinity changes. *Fish Shellfish Immunol.* 2018;73:288–96.
- Shin HS, Choi YJ, Kim NN, Lee J, Ueda H, Choi CY. Effects of exogenous cortisol and seawater adaptation on thyroid hormone receptors in the smolt stage of the sockeye salmon, *Oncorhynchus nerka*. *Ichthyol Res.* 2014;61:9–16.
- Zhou S, Zhang N, Fu Z, Yu G, Ma Z, Zhao L. Impact of salinity changes on the antioxidation of juvenile yellowfin tuna (*Thunnus albacares*). *J Mar Sci Eng.* 2023;11(1):132.
- Sakai H, Goto A, Jeon SR. Speciation and dispersal of *Tribolodon* species (Pisces, Cyprinidae) around the Sea of Japan. *Zool Sci.* 2002;19(11):1291–303.
- Sakai H. Life-histories and genetic divergence in three species of *Tribolodon* (Cyprinidae). *Memoirs Facul Fisheries Hokkaido Univ.* 1995;42(1–2):1–98.
- Nakamura M. Cyprinid fishes of Japan. Studies on the life history of cyprinid fishes of Japan. *Spec Publ Res Inst Nat Resources.* 1969;4:189–90.
- Chang Y, Cheng L, Sun B, Su B, Liang L, Wang W. Species classification and origin of *Tribolodon* in the Suifen River, China. *J Fish Sci China.* 2018;25(4):811–8.
- Dong CZ, Zhao CG, Jin ZL. An approach on the breeding population structure of *Leuciscus brandti* during migration in Suifen River. *J Fish China.* 1993;17(4):304–11.
- Zhao CG, Wang WK, Ma GJ, Wang QS, Zhang LM. The structure of spawning population of anadromous Japanese dace in Suifen River. *Chinese Journal of Fisheries.* 2011;24(3):25–8.
- Chi B, Liang L, Liu C, Chang Y, Wang S, Han Q, Gao G. Adaptability of *Tribolodon brandti* (Dybowski) to NaCl concentration and alkalinity. *J Fish Sci China.* 2011;18(3):689–94.
- Kaneko T, Hasegawa S, Uchida K, Ogasawara T, Oyagi A, Hirano T. Acid tolerance of Japanese dace (a cyprinid teleost) in Lake Osorezan, a remarkable acid lake. *Zool Sci.* 1999;16(6):871–7.
- Lee SY, Lee HJ, Kim YK. Comparative transcriptome profiling of selected osmotic regulatory proteins in the gill during seawater acclimation of chum salmon (*Oncorhynchus keta*) fry. *Sci Rep.* 2020;10(1):1987.
- Bernat M, Pablo B, Amaya A, Cindy P, Simon M, Shu N, Munetaka S, Tom ON, Harald S, Lars OE, et al. Effects of different photoperiod regimes on the smoltification and seawater adaptation of seawater-farmed rainbow trout (*Oncorhynchus mykiss*): Insights from Na<sup>+</sup>, K<sup>+</sup>-ATPase activity and transcription of osmoregulation and growth regulation genes. *Aquaculture.* 2019;507:82–292.
- Thompson NF, Anderson EC, Clemente AJ, Campbell MA, Garza JC. A complex phenotype in salmon controlled by a simple change in migratory timing. *Science.* 2020;370(6516):609–13.
- Harringmeyer OS, Woolfolk ML, Hoekstra HE. Fishing for the genetic basis of migratory behavior. *Cell.* 2021;184(2):303–5.
- Narum SR, Alex DG, Micheletti SJ, Alejandro M. Genomic variation underlying complex life-history traits revealed by genome sequencing in Chinook salmon. *Proc Biol Sci.* 1883;2018(285):20180935.
- Prince DJ, O'Rourke SM, Thompson TQ, Ali OA, Lyman HS, Saglam IK, Hotaling TJ, Spidle AP, Miller MR. The evolutionary basis of premature migration in Pacific salmon highlights the utility of genomics for informing conservation. *Sci Adv.* 2017;3(8):e1603198.
- Duan Z, Dong S, Sun Y, Dong Y, Gao Q. Response of Atlantic salmon (*Salmo salar*) flavor to environmental salinity while culturing between freshwater and seawater. *Aquaculture.* 2021;530:735953.
- Fossmark RO, Attramadal K, Nordoy K, Osterhus SW, Vadstein O. A comparison of two seawater adaptation strategies for Atlantic salmon post-smolt (*Salmo salar*) grown in recirculating aquaculture systems (RAS): nitrification, water and gut microbiota, and performance of fish. *Aquaculture.* 2021;532:735973.
- Takehana Y, Zahm M, Klopp C, Bouchez O, Barrachina C, Journot L, Kawaguchi M, Yasumasu S, Ansai S, Naruse K. Genome sequence of the euryhaline Javafish medaka, *Oryzias javanicus*: a small aquarium fish model for studies on adaptation to salinity. *G3 (Bethesda).* 2020;10(3):g3.4007254002019.

33. Howe K, Clark MD, Torroja CF, Torrance J, Berthelot C, Muffato M, Collins JE, Humphray S, McLaren K, Matthews L. The zebrafish reference genome sequence and its relationship to the human genome. *Nature*. 2013;496:498–503.
34. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. *Nature*. 2011;475(7357):493–6.
35. Harris RM, Hofmann HA. Seeing is believing: dynamic evolution of gene families. *Proc Natl Acad Sci U S A*. 2015;112(5):1252–3.
36. Rodríguez A, Gallardo MA, Gisbert E, Santilari S, Ibarz A, Sánchez J, Castelló-Orvay F. Osmoregulation in juvenile Siberian sturgeon (*Acipenser baerii*). *Fish Physiol Biochem*. 2002;26(4):345–54.
37. Blasco J, Fernández-Borrás J, Marimon I, Requena A. Plasma glucose kinetics and tissue uptake in brown trout in vivo: effect of an intravascular glucose load. *J Comp Physiol B*. 1996;165(7):534–41.
38. Marshall WS, Breves JP, Doohan EM, Tipsmark CK, Kelly SP, Robertson GN, Schulte PM. claudin-10 isoform expression and cation selectivity change with salinity in salt-secreting epithelia of *Fundulus heteroclitus*. *J Exp Biol*. 2018;221(1):jeb168906.
39. Bollinger RJ, Reed JP, Bossus MC, Tipsmark CK. Prolactin and cortisol regulate branchial claudin expression in Japanese medaka. *Gen Comp Endocr*. 2017;240:77–83.
40. Musilova Z, Salzburger W, Cortesi F. The visual opsin gene repertoires of teleost fishes: evolution, ecology, and function. *Annu Rev Cell Dev Bi*. 2021;37:441–68.
41. Bowmaker JK, Semo MA, Hunt DM, Jeffery G. Eel visual pigments revisited: the fate of retinal cones during metamorphosis. *Vis Neurosci*. 2008;25(3):249–55.
42. Polychronopoulos D, King James WD, Nash AJ, Tan G, Lenhard B. Conserved non-coding elements: developmental gene regulation meets genome organization. *Nucleic Acids Res*. 2017;45(22):12611–24.
43. Yin JS, Shen JB, Luan XH. Effect of temperature on the development of dace (*Leuciscus brandti*) embryos in the Suifen River. *Acta Ecol Sin*. 1992;3:232–8.
44. Deng Y, Meng M, Fang J, Jiang H, Sun N, Lv W, Lei Y, Wang C, Bo J, Liu C, et al. Genome of the butterfly hillstream loach provides insights into adaptations to torrential mountain stream life. *Mol Ecol Resour*. 2021;21(6):1922–35.
45. Han Z, Liu M, Liu Q, Zhai H, Xiao S, Gao T. Chromosome-level genome assembly of burbot (*Lota lota*) provides insights into the evolutionary adaptations in freshwater. *Mol Ecol Resour*. 2021;21(6):2022–33.
46. Wu B, Feng C, Zhu C, Xu W, Yuan Y, Hu M, Yuan K, Li Y, Ren Y, Zhou Y, et al. The genomes of two billfishes provide insights into the evolution of endothermy in teleosts. *Mol Biol Evol*. 2021;38(6):2413–27.
47. Zhang XY, Wen HS, Zhang KQ, Liu Y, Fang X, Li Y. Analysis of the isotonic point and effects of seawater desalination on the Na<sup>+</sup>/K<sup>+</sup>/Cl<sup>-</sup> concentration, Na<sup>+</sup>-K<sup>+</sup>-ATPase activity and relative gene expressions in *Lateolabrax maculatus*. *J Fisher Chin*. 2018;42(8):1199–208.
48. Xu Z, Gan L, Li T, Xu C, Chen K, Wang X, Qin JG, Chen L, Li E. Transcriptome profiling and molecular pathway analysis of genes in association with salinity adaptation in Nile tilapia *Oreochromis niloticus*. *PLoS ONE*. 2015;10(8):e0136506.
49. Zong SB, Li YL, Liu JX. Genomic architecture of rapid parallel adaptation to fresh water in a wild fish. *Mol Biol Evol*. 2020;38(9):1317–29.
50. Allen P, Mitchell Z, DeVries R, Aboagye D, Ciaramella M, Ramee S, Stewart H, Shartau R. Salinity effects on Atlantic sturgeon (*Acipenser oxyrinchus oxyrinchus* Mitchell, 1815) growth and osmoregulation. *J Appl Ichthyol*. 2014;30(6):1229–36.
51. Yang S, Li D, Feng L, Zhang C, Xi D, Liu H, Yan C, Xu Z, Zhang Y, Li Y. Transcriptome analysis reveals the high temperature induced damage is a significant factor affecting the osmotic function of gill tissue in Siberian sturgeon (*Acipenser baerii*). *BMC Genomics*. 2023;24(1):2.
52. Kolosov D, Bui P, Wilkie MP, Kelly SP. Claudins of sea lamprey (*Petromyzon marinus*) – organ-specific expression and transcriptional responses to water of varying ion content. *J Fish Biol*. 2020;96(3):768–81.
53. Bossus MC, Madsen SS, Tipsmark CK. Functional dynamics of claudin expression in Japanese medaka (*Oryzias latipes*): Response to environmental salinity. *Comp Biochem Physiol A Mol Integr Physiol*. 2015;187:74–85.
54. Pca B, Tn A, Ts A, Ac A. Water salinity and postprandial effects on transcription of peptide and amino acid transporters in the kidney of Mozambique tilapia (*Oreochromis mossambicus*). *Aquaculture*. 2021;536:736384.
55. Rasmussen RN, Christensen KV, Holm R, Nielsen CU. Transcriptome analysis identifies activated signaling pathways and regulated ABC transporters and solute carriers after hyperosmotic stress in renal MDCK I cells. *Genomics*. 2018;111(6):1557–65.
56. Tseng YC, Hwang PP. Some insights into energy metabolism for osmoregulation in fish. *Comp Biochem Physiol C Toxicol Pharmacol*. 2008;148(4):419–29.
57. Bonzi LC, Monroe AA, Lehmann R, Berumen ML, Ravasi T, Schunter C. The time course of molecular acclimation to seawater in a euryhaline fish. *Sci Rep*. 2021;11(1):18127.
58. Hill A, Shachar-Hill Y. What are aquaporins for? [Review]. *J Membr Biol*. 2004;197(1):1–32.
59. Zhang X, Wen H, Wang H, Ren Y, Zhao J, Li Y. RNA-Seq analysis of salinity stress-responsive transcriptome in the liver of spotted sea bass (*Lateolabrax maculatus*). *PLoS ONE*. 2017;12(3):e0173238.
60. Kumari A, Koner D, Lanong AS, Khongmawloh E, Snaitarang R, Saha N. Changes in the expression of certain osmosensitive channel and transporter genes in primary hepatocytes of air-breathing catfish, *Clarias magur*: a strategy to adapt under osmotic stress. *Aquaculture*. 2022;562:738756.
61. Shen Y, He Y, Bi Y, Chen J, Zhao Z, Li J, Chen X. Transcriptome analysis of gill from *Lateolabrax maculatus* and *aqp3* gene expression. *Aquacult Fish*. 2019;4(6):247–54.
62. Cowan KJ, Storey KB. Mitogen-activated protein kinases: new signaling pathways functioning in cellular responses to environmental stress. *J Exp Biol*. 2003;206(Pt 7):1107.
63. Kultz D. Molecular and evolutionary basis of the cellular stress response. *Annu Rev Physiol*. 2005;67(1):225–57.
64. Michael K. Mitogen-activated protein kinase cascades as regulators of stress responses. *Ann N Y Acad Sci*. 1998;851(1):139–46.
65. Li J, Ferraris JD, Yu D, Singh T, Izumi Y, Wang G, Gucek M, Burg MB. Proteomic analysis of high NaCl-induced changes in abundance of nuclear proteins. *Physiol Genomics*. 2012;44(21):1063–71.
66. Zhang TG, Chen QQ, Wang N, Xia XH, Wang J, Chang Y, Yang YL, Yang N, Sun WC. Molecular cloning and characterization of the mitogen-activated protein kinase kinase gene (MKK4) and its promoter sequence from oilseed rape (*Brassica campestris* L.). *Plant Cell Tissue Organ Cult PCTOC*. 2013;115(3):341–53.
67. Pickford GE, Phillips JG. Prolactin, a factor in promoting survival of hypophysectomized killifish in fresh water. *Science*. 1959;130(3373):454–5.
68. Wang FY, Fu WC, Wang IL, Yan HY, Wang TY. The giant mottled eel, *Anguilla marmorata*, uses blue-shifted rod photoreceptors during upstream migration. *PLoS ONE*. 2014;9(8):e103953.
69. Wang FY, Tang MY, Yan HY. A comparative study on the visual adaptations of four species of moray eel. *Vision Res*. 2011;51(9):1099–108.
70. Cottrill PB, Davies WL, Semo MA, Bowmaker JK, Hunt DM, Jeffery G. Developmental dynamics of cone photoreceptors in the eel. *BMC Dev Biol*. 2009;9(1):1–9.
71. Zhang H, Futami K, Horie N, Okamura A, Utoh T, Mikawa N, Yamada Y, Tanaka S, Okamoto N. Molecular cloning of fresh water and deep-sea rod opsin genes from Japanese eel *Anguilla japonica* and expressional analyses during sexual maturation. *FEBS Lett*. 2000;469(1):39–43.
72. Dann SG, Ted Allison W, Levin DB, Taylor JS, Hawryshyn CW. Salmonid opsin sequences undergo positive selection and indicate an alternate evolutionary relationship in *Oncorhynchus*. *J Mol Evol*. 2004;58:400–12.
73. Mclean CY, Reno PL, Pollen AA, Bassan AI, Capellini TD, Guenther C, Indjeian VB, Lim X, Menke DB, Schaar BT. Human-specific loss of regulatory DNA and the evolution of human-specific traits. *Nature*. 2011;471(7337):216–9.
74. Navratilova P, Fredman D, Hawkins TA, Turner K, Lenhard B, Becker TS. Systematic human/zebrafish comparative identification of cis-regulatory activity around vertebrate developmental transcription factor genes. *Dev Biol*. 2009;327(2):526–40.
75. Visel A, Prabhakar S, Akiyama JA, Shoukry M, Lewis KD, Holt A, Plajzer-Frick I, Afzal V, Rubin EM, Pennacchio LA. Ultraconservation identifies a small subset of extremely constrained developmental enhancers. *Nat Genet*. 2008;40:158–60.
76. Liu B, Gao Q, Liu B, Song C, Sun C, Liu M, Liu X, Liu Y, Li Z, Zhou Q, et al. Application of transcriptome analysis to understand the adverse effects

- of hypotonic stress on different development stages in the giant freshwater prawn *Macrobrachium rosenbergii* post-larvae. *Antioxidants*. 2022;11(3):440.
77. Taugbøl A, Solbakken MH, Jakobsen KS, Vøllestad LA. Salinity-induced transcriptome profiles in marine and freshwater threespine stickleback after an abrupt 6-hour exposure. *Ecol Evol*. 2022;12(10):e9395.
  78. Liu B, Shi Y, Yuan J, Hu X, Zhang H, Li N, Li Z, Chen Y, Mu D, Fan W: Estimation of genomic characteristics by analyzing k-mer frequency in *de novo* genome projects. In: *ArXiv*. 2013: 1308.2012v1301.
  79. Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*. 2017;27(5):722–36.
  80. Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. *Nat Biotechnol*. 2019;37(5):540–6.
  81. Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res*. 2017;27(5):737–46.
  82. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS ONE*. 2014;9(11):e112963.
  83. Langmead B. Aligning short sequencing reads with Bowtie. *Curr Protoc Bioinformatics*. 2010;32(1):11.17.11–11.17.14.
  84. Burton JN, Adey A, Patwardhan RP, Qiu R, Kitzman JO, Shendure J. Chromosome-scale scaffolding of *de novo* genome assemblies based on chromatin interactions. *Nat Biotechnol*. 2013;31(12):1119–25.
  85. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*. 2015;31(19):3210–2.
  86. Tarailo-Graovac M, Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinformatics*. 2009;25(1):4.10.11–14.10.14.
  87. Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Buell CR, Wortman JR. Automated eukaryotic gene structure annotation using EvidenceModeler and the Program to Assemble Spliced Alignments. *Genome Biol*. 2008;9(1):1–22.
  88. Wang Y, Lu Y, Zhang Y, Ning Z, Li Y, Zhao Q, Lu H, Huang R, Xia X, Feng Q, et al. The draft genome of the grass carp (*Ctenopharyngodon idellus*) provides insights into its evolution and vegetarian adaptation. *Nat Genet*. 2015;47(6):625–31.
  89. Jian J, Yang L, Gan X, Wu B, Gao L, Zeng H, Wang X, Liang Z, Wang Y, Fang L, et al. Whole genome sequencing of silver carp (*Hypophthalmichthys molitrix*) and bighead carp (*Hypophthalmichthys nobilis*) provide novel insights into their evolution and speciation. *Mol Ecol Resour*. 2021;21(3):912–23.
  90. Liu H, Chen C, Gao Z, Min J, Gu Y, Jian J, Jiang X, Cai H, Ebersberger I, Xu M. The draft genome of blunt snout bream (*Megalobrama amblycephala*) reveals the development of intermuscular bone and adaptation to herbivorous diet. *Gigascience*. 2017;6(7):gix039.
  91. Warren WC, Boggs TE, Borowsky R, Carlson BM, Ferrufino E, Gross JB, Hillier L, Hu Z, Keene AC, Kenzior A. A chromosome-level genome of *Astyanax mexicanus* surface fish for comparing population-specific genetic differences contributing to trait evolution. *Nat Commun*. 2021;12(1):1447.
  92. Chen X, Zhong L, Bian C, Xu P, Qiu Y, You X, Zhang S, Huang Y, Li J, Wang M. High-quality genome assembly of channel catfish, *Ictalurus punctatus*. *Gigascience*. 2016;5(1):s13742–13016–10142–3745.
  93. Gong G, Dan C, Xiao S, Guo W, Huang P, Xiong Y, Wu J, He Y, Zhang J, Li X. Chromosomal-level assembly of yellow catfish genome using third-generation DNA sequencing and Hi-C analysis. *GigaScience*. 2018;7(11):gij120.
  94. Terekhanova NV, Logacheva MD, Penin AA, Neretina TV, Barmintseva AE, Bazykin GA, Kondrashov AS, Mugue NS. Fast evolution from precast bricks: genomics of young freshwater populations of threespine stickleback *Gasterosteus aculeatus*. *PLoS Genet*. 2014;10(10):e1004696.
  95. Aparicio S, Chapman J, Stupka E, Putnam N, Chia JM, Dehal P, Christoffels A, Rash S, Hoon S, Smit A. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science*. 2002;297(5585):1301–10.
  96. Conte MA, Gammerdinger WJ, Bartie KL, Penman DJ, Kocher TD. A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. *BMC Genomics*. 2017;18(1):1–19.
  97. Naruse K, Fukamachi S, Mitani H, Kondo M, Matsuoka T, Kondo S, Hanamura N, Morita Y, Hasegawa K, Nishigaki R. A detailed linkage map of medaka, *Oryzias latipes*: comparative genomics and genome evolution. *Genetics*. 2000;154(4):1773–84.
  98. Schartl M, Walter RB, Shen Y, Garcia T, Catchen J, Amores A, Braasch I, Chalopin D, Volff JN, Lesch KP. The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nat Genet*. 2013;45(5):567–72.
  99. Lin Q, Fan S, Zhang Y, Xu M, Zhang H, Yang Y, Lee AP, Woltering JM, Ravi V, Gunter HM. The seahorse genome and the evolution of its specialized morphology. *Nature*. 2016;540(7633):395–9.
  100. You X, Bian C, Zan Q, Xu X, Liu X, Chen J, Wang J, Qiu Y, Li W, Zhang X. Mudskipper genomes provide insights into the terrestrial adaptation of amphibious fishes. *Nat Commun*. 2014;5(1):5594.
  101. Braasch I, Gehrke AR, Smith JJ, Kawasaki K, Manouaki T, Pasquier J, Amores A, Desvignes T, Batzel P, Catchen J, et al. The spotted gar genome illuminates vertebrate evolution and facilitates human-teleost comparisons. *Nat Genet*. 2016;48(4):427–37.
  102. Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. 2004;5:1792–7.
  103. Alexandros S. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*. 2014;30(9):1312–3.
  104. Rannala YB. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. *Mol Biol Evol*. 2006;23(1):212–26.
  105. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–60.
  106. De Bie T, Cristianini N, Demuth JP, Hahn MW. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics*. 2006;22(10):1269–71.
  107. Yang Z. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol*. 2007;24(8):1586–91.
  108. Finn RD, Coghill P, Eberhardt RY, Eddy SR, Mistry J, Mitchell AL, Potter SC, Punta M, Qureshi M, Sangrador-Vegas A. The Pfam protein families database: towards a more sustainable future. *Nucleic Acids Res*. 2016;44(D1):D279–85.
  109. Kielbasa SM, Wan R, Sato K, Horton P, Frith MC. Adaptive seeds tame genomic sequence comparison. *Genome Res*. 2011;21(3):487–93.
  110. Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AF, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res*. 2004;14(4):708–15.
  111. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Spieth J, Hillier LW, Richards S. Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res*. 2005;15(8):1034–50.
  112. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. PHAST: a fast phage search tool. *Nucleic Acids Res*. 2011;39(suppl\_1):W347–52.
  113. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164–e164.
  114. Frazer KA, Pachter L, Poliakov A, Rubin EM, Dubchak I. VISTA: computational tools for comparative genomics. *Nucleic Acids Res*. 2004;32(suppl\_2):W273–9.
  115. *Pseudaspius brandtii* and *P. hakonensis* sequence. NCBI BioProject accession PRJNA980574. 2023. <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA980574>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.