

A chromosome-level genome assembly of the Chinese herring (*Ilisha elongata*) uncovered its population dynamics and genes related to lipid metabolism

Tianqin Wu^{1,2†}, Jiantao Hu^{1,2†}, Vanthu Giap^{1,2}, Qian Wang³, Liang Lu^{1,2‡}, and Chenhong Li^{1,2‡}

¹ Shanghai Universities Key Laboratory of Marine Animal Taxonomy and Evolution, Shanghai Ocean University, Shanghai, China 201306;

² Engineering Research Center of Environmental DNA and Ecological Water Health Assessment, Shanghai Ocean University, Shanghai, China 201306;

³ East China Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Shanghai, 200090, China;

[†]Tianqin Wu and Jiantao Hu contributed equally to this work.

[‡]Authors for correspondence: Chenhong Li, chli@shou.edu.cn; Liang Lu, llu@shou.edu.cn

Accepted

Abstract

The Chinese herring (*Ilisha elongata*) is an economically important clupeiform fish, mostly found along the coast of China and Japan in the western Pacific Ocean. Overfishing and climate change have caused a substantial decline in its annual catch, which may have impacted its population size and structure. Here, we present a chromosome-level genome assembly of the Chinese herring, generated using a combination of Nanopore sequencing, Illumina sequencing, and high-throughput chromatin conformation capture (Hi-C) technologies. The total length of this assembly is 802.47 Mbp, anchored to 24 chromosomes, with contig N50 and scaffold N50 values of 13.22 Mbp and 33.10 Mbp, respectively. BUSCO analysis indicated high completeness of this assembly, with 94.9% of universal single-copy orthologs. Analysis of resequencing samples from various coastal regions of China and Japan revealed four distinct genetic populations of Chinese herring along the East Asian coast. The Dandong population (Yellow Sea) diverged earliest, exhibiting unique genetic structures and selective sweep signals, whereas the other three populations expanded from a tropical South China Sea ancestral population to the East China Sea and to Japanese waters. Among 23,366 protein-coding genes, several families related to metabolism of very-long-chain unsaturated fatty acids are particularly contracted compared to the American shad. Additionally, several genes associated with fatty acid synthesis formation have undergone accelerated positive selection. These findings may explain the differential fatty acid ratios in marine versus riverine migratory clupeiforms. This high-quality genome assembly of the Chinese herring may provide valuable resources for comparative study of coastal fishes and benefit the development of a sustainable management strategy for fisheries of the Chinese herring.

Key words: *elongata ilisha*, population structure, the Northwest Pacific, fatty acid, chromosomal assembly, genome annotation

Significance statement

The Chinese herring (*Ilisha elongata*) is economically important species along the coast of the Northwest Pacific. The fishery resource of the Chinese herring is declining in recent years, but its fine population structure has not been studied using genomic data and cryptic diversity may exist. Understanding its population dynamics and diversity using genomic data is critical to conservation and sustainable utilization of this species. A high-quality genome assembly of the Chinese herring also can provide useful resources for comparative study on evolution of interesting traits among coastal fishes, such as migration, fatty acid synthesis, intermuscular bones, etc.

Accepted Manuscript

Introduction

The order Clupeiformes, including anchovies, herrings, sardines, shads, and their relatives, comprises 435 species distributed across a variety of habitats, reflecting significant biodiversity (Fricke et al., 2022). Over half of the fish species within the Clupeiformes are found in pan-saline coastal marine environments, such as the Indo-West Pacific (Nelson et al., 2016). Species of the Clupeiformes predominantly occupy the lower strata of marine food web, serving as essential forage fish for numerous commercially important species, holding substantial ecological significance. Due to overfishing, climate change and other anthropogenic activities, populations of many species are at risk of decline. Despite this, current research predominantly focuses on endangered and commercially significant fish species, resulting in a relative paucity of studies and data concerning clupeiform populations (Xue et al., 2019; Birge et al., 2021).

The Chinese herring, *Ilisha elongata* (Clupeiformes, Pristigasteridae) is a warm-water fish characterized by its coastal distribution and upper-layer migratory behavior in offshore waters. This species is primarily distributed in the western Pacific Ocean, with notable populations along the coasts of China, the Korean Peninsula and Japan, and few in the Sarawak and Java Seas. It is regarded as a regional economic species. The Chinese herring has been an significant edible fish in China since ancient times. Evidence of its consumption dates back to 4,000 years ago, with skeletal remains discovered at the Sanlihe site in Jiaozhou city, Shandong Province of China (Cheng, 1981). In Mandarin, this fish is commonly referred to as Lè yú, a term derived from the sensation of having a fish bone lodged in the throat, which reflects numerous fine, intermuscular bones present in its flesh. Additionally, a regional dialect in the Zhoushan area refers to it as Xiǎng yú, a term indicating a traditional preservation method involving splitting and drying the fish, typically used for fish with higher fat content. This nomenclature suggests that the Chinese herring has a higher fat content compared to other marine edible fish.

According to the Chinese Fisheries Yearbook (2023), the annual production of the Chinese herring has decreased from 83,262 tons in 2013 to 57,833 tons in 2023. The fishing yield and fishery resources of the Chinese herring have undergone a notable decline, with observable

degeneration in biological traits in wild populations of this species in certain regions. This includes phenomena such as early sexual maturation and miniaturization during the maturation phase (Xie et al., 2012). Commercial marine species like the Chinese herring typically exhibit reduced intraspecific genetic variation compared to non-commercial species (Pinsky and Palumbi, 2014). Understanding the population genetic structure of marine fish species, such as the Chinese herring, is crucial for developing effective long-term management and conservation strategies for the fishery stocks. Slatkin and Hudson (1991) established that pairwise comparisons of mitochondrial DNA sequences are essential for understanding population dynamics in both stable and growing populations, providing foundational insights into fish genetic diversity and structure. Norris et al. (2000) extended this by using microsatellite markers to assess parentage and relatedness in farmed Atlantic salmon, highlighting genetic variation in evaluating adaptive capacity and evolutionary potential. Anderson et al. (2024) further advanced this field with a strategic road map for sustainable fisheries through population genomics, showing how detailed genetic data can enhance management practices. Together, these studies underscore the critical importance of genetic diversity assessments for ensuring sustainable and resilient fish populations, offering a robust scientific basis for effective fisheries management.

In early studies of the Chinese herring populations, they were categorized into geographic subgroups based on morphological characteristics, growth traits, and specific locations, without genetic validation. Morphological research indicated differences in the edge index of scales between the Chinese herring from the Korean Peninsula and the East China Sea, leading to the conclusion that these two regions should not be considered part of the same population group (Zhang and Takita, 2007; Kim et al., 2007).

Recent studies have analyzed the population diversity of the Chinese herring using mitochondrial DNA sequences. Mitochondrial D-loop analysis revealed genetic differentiation among populations along the Chinese coast, with the Qingdao population exhibiting notable uniqueness (Wu et al., 2009). A study using 16S rDNA marker further confirmed these findings, highlighting genetic diversity along the Chinese coast and suggesting population abundance (Lv et al., 2010). This contrasts with a study suggesting a decline in diversity in the East China Sea

due to overfishing and environmental factors (Li et al., 2013). Wang et al. (2016) extended this research by analyzing D-loop sequences from 16 locations, identifying two primary clades: one including a single Yalu River estuary population (Dandong), and another showing gene flow between 15 other sites. They proposed that ancestral populations were located in the South China Sea and suggested that current distribution patterns resulted from population expansion or colonization events. Particularly, Wang et al. (2016) were the first to identify significant mitochondrial sequence differences between the Dandong Yalu River estuary population and other populations.

However, these studies, including those by Wu et al. (2009), Lv et al. (2010), and Li et al. (2013), are limited by their reliance on mitochondrial markers, which may not fully capture the population structure due to their maternal inheritance and lack of recombination. To achieve a comprehensive understanding of the demographic history and evolutionary forces shaping the Chinese herring, it is essential to reevaluate it using genome-level data, which would provide a more detailed and integrated perspective on its genetic diversity and population structure in the Northwestern Pacific.

In a previous study, Wang et al. (2022) collected nuclear sequence data from 4,434 single-copy protein-coding loci obtained via a gene-capture method to determine the phylogenetic positions of ten families within the order Clupeiformes. The *Ilisha elongata* belongs to the family Pristigasteridae within this order. While the Pristigasteridae comprises five species, none currently have available genomic data. In terms of phylogenetic relationships, the closest sister clade to the Pristigasteridae consists of three families: Dussumieriidae, Chirocentridae, and Clupeidae. The nearest relative with a published genome to the *Ilisha elongata* is *Clupea harengus* from the family Clupeidae, which diverged from the Pristigasteridae approximately 57.98 million years ago (Mya). A more distantly related species with an available genome is *Alosa alosa* from the family Alosidae; this species diverged along with four other families approximately 65.16 Mya.

In this study, we used third-generation Nanopore sequencing technology with high-resolution chromosome conformation capture (Hi-C) techniques to assemble and annotate the

first chromosome-level genome of the Chinese herring. We then use resequencing data to elucidate the differences in genetic loci among populations inhabiting different habitats, particularly the Dandong population. Furthermore, we aim to investigate the genetic mechanism related to abundant Omega-3 fatty acids found in the Chinese herring. Previous studies have suggested that the ELOVL gene family plays a primary role in the synthesis of very long-chain unsaturated fatty acids. Studies on ELOVL genes have demonstrated that the synthesis of unsaturated fatty acids in fish requires the coordinated action of fatty acid desaturases (FADs) and ELOVLs (Jakobsson et al. 2006; Guillou et al. 2010), with research showing a positive correlation between the expression of ELOVL genes and the synthesis of long-chain polyunsaturated fatty acids through knockout and overexpression experiments in various bony fish species (Wang et al. 2020; Zhao et al. 2020; Wang et al. 2022). Fat metabolism studies highlight the importance of fatty acids for fish physiology, noting that β -oxidation efficiency differs between saturated and unsaturated fatty acids (Langdon and Waldock, 1981). Through the analysis of gene families and positive selection, we seek to uncover the relationship between the life histories of various clupeiform species and the differences in their fatty acid profiles. These findings will contribute to further research on the biology and genetic evolution of this species and have significant implications for the sustainable management of the Chinese herring fisheries.

Result

Genome assembly and annotation

A sample of the Chinese herring was caught by hand rod and line in the vicinity of Goji Berry Island, Shengsi County, Zhoushan City, Zhejiang Province (coordinates: E122° , N30°) in October 2023. The individual is a male fish and weighted for 1,216 gram (Fig. 1 A). Liver tissue from this individual was harvested and frozen in liquid nitrogen for Nanopore sequencing, while muscle sample was collected to construct the Hi-C library. Nine tissues (heart, liver, brain,

kidney, skin, eye, muscle, swim bladder, gill) were kept in RNA later for transcriptomic sequencing.

To assemble a chromosome-level genome, we obtained 115.62 GB data from Nanopore sequencing, providing substantial coverage of the whole genome. Additionally, we sequenced 63.81 GB of paired-end Illumina data to correct errors in the Nanopore data. We also sequenced 94.05 GB of Hi-C data to anchor scaffolds to chromosomes. For genome annotation, we sequenced 58.91 GB of transcriptome data of various tissue samples.

The final assembled genome size is 802.47 Mbp, with 98.87% of the initial contigs anchored to 24 chromosomes through Hi-C methods. The Scaffold N50 length is 33.10 Mbp, and the Contig N50 length is 13.22 Mbp, with a GC content of 43.7%. BUSCO analysis indicated that 94.9% (3456/3640) of the BUSCOs are complete, including 3400 single-copy and 56 duplicated BUSCOs. RepeatMasker results revealed that 33.46% of the sequences were predicted as repetitive elements, with retroelements TE comprising 5.88%, LTRs 1.89%, and DNA transposons 7.76% being the predominant repetitive types (Table 1). The high mapping rate (98.04%) of the NGS re-sequencing data, calculated after excluding supplementary alignments (372,583,887 total reads; 370,992,679 mapped), combined with a properly paired rate of 96.38% and minimal singletons (0.15%), strongly supports the high completeness and accuracy of the assembled genome, with no significant gaps or misassemblies affecting read alignment. We predicted a total of 23,366 protein-coding genes based on homology and RNA-seq data. The chromosome contact map of the entire genome is presented using Juicerbox (Fig. 1 C). Visualization of chromosome-level synteny among the Chinese herring and *Clupea harengus* and *Alosa alosa*, demonstrates a high degree of synteny among clupeiform genomes (Fig. 2 A). Additionally, Figure 2 B shows various genomic features, including intra-genomic synteny, GC content heatmap, gene density distribution, LTR transposon distribution, and TE transposon distribution.

Table 1. Summary of genome assembly data of the Chinese herring (*Ilisha elongata*)

assembly	data
Total length (Mbp)	802.47
GC (%)	43.7
Scaffold N50 length (Mbp)	33.10
Contig N50 length (Mbp)	13.22
Scaffold N50 number	219
Contig N50 number	332
Gene number	23366
Heterozygosity (%)	0.98556
Total BUSCO (%)	94.9
Repeat content (%)	33.46
Hi-C anchoring rate (%)	98.87
Number of chromosomes	24
Percent gaps (%)	0.007

A



B



C

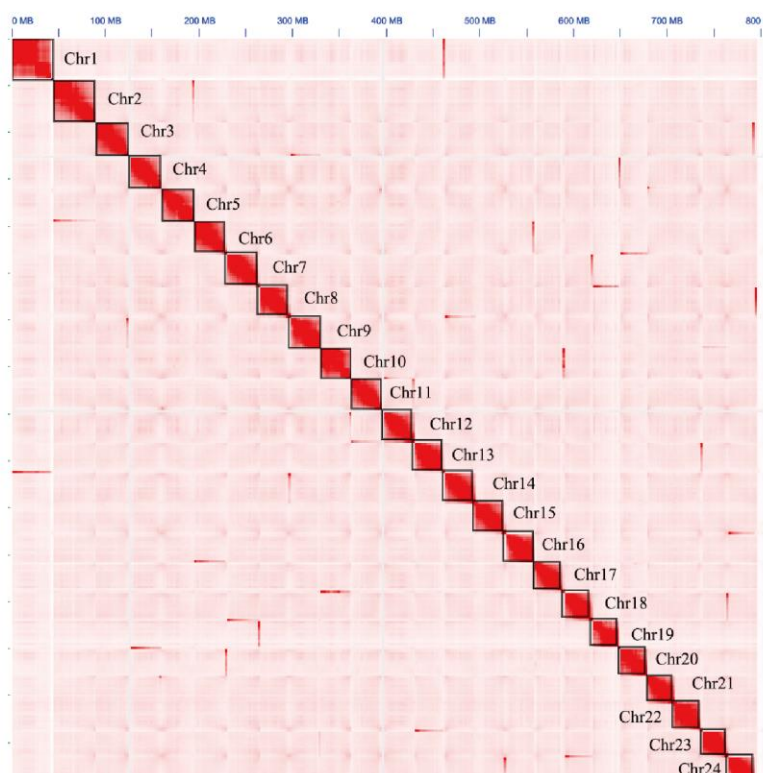
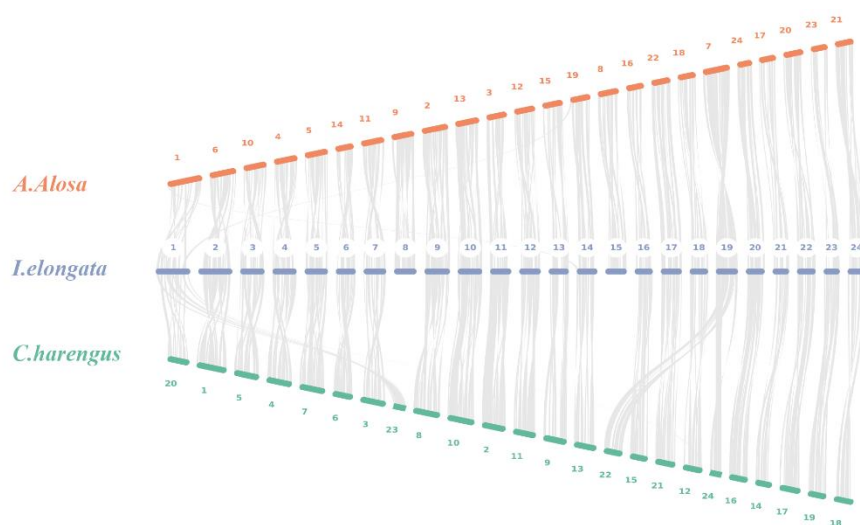


Fig. 1 **A.** The specimen of the Chinese herring (*Ilisha elongata*) used for genome sequencing. **B.** Sampling locations of the Chinese herring for genome resequencing. **C.** Chromosome contact map of the Chinese herring.

Accepted Manuscript

A



B

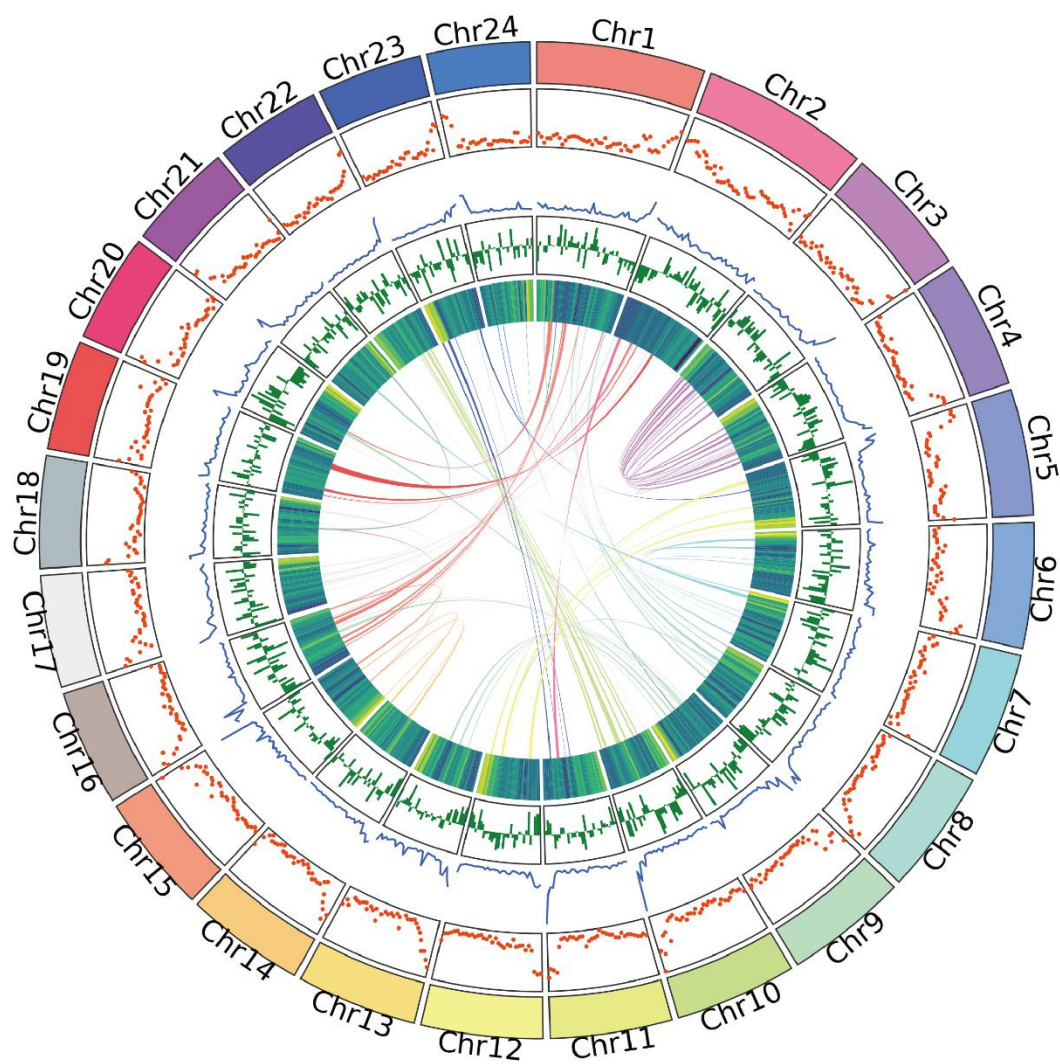


Fig. 2. **A.** Synteny plot showing the genomic collinearity between the Chinese herring, *Clupea harengus* and *Alosa alosa*. **B.** Genomic features of the Chinese herring. From the inner to the outer layer: genome-wide synteny, GC content heatmap, gene density distribution, LTR transposon distribution, and TE transposon distribution.

Population structure

The paired-end resequencing data for 51 Chinese herring samples totaled approximately 1.7 TB. Following SNP calling with GATK, around 20 million SNPs were identified, of which approximately 5 million were filtered for subsequent analysis. PCA results clearly illustrate that individuals from the four populations are distinct, with the Dandong population being the most divergent from the other three, which are relatively closer to each other (Fig. 3 A). This indicates significant genomic differentiation between the Dandong population and the others.

Admixture analysis showed cross-validation errors as follows: CV error for K=2 was 0.48056, K=3 was 0.47427, K=4 was 0.51460, and K=7 was 0.68776, with lower values indicating the optimal number of clusters was K=3, under which the populations were divided into Dandong, Guangxi-Zhoushan (representing the East China Sea and South China Sea populations), and Japan. Particularly, the Zhoushan and Japan populations exhibited some genetic exchange (Fig. 3 C).

PSMC analysis (Fig. 3 B) illuminated significant historical population dynamics in the Chinese herring across different regions, including Guangxi, Zhoushan, Japan, and Dandong. During the Dali Glacial Period (57,000-16,000 years ago), which reduced marine temperatures and challenged the species' survival, notable population declines were observed in Japan, Zhoushan, and Guangxi populations, with subsequent recoveries post-glaciation. Particularly, the Guangxi population experienced the smallest decrease and the earliest recovery, consistent with the refugia hypothesis. This hypothesizes that during glacial periods, the Chinese herring retreated to the warmer waters of the South China Sea, serving as a refuge, and subsequently expanded northward during interglacial periods, giving rise to the Zhoushan and Japan populations. In contrast, the Dandong population exhibited atypically small historical population size fluctuations, suggesting a distinct evolutionary trajectory. Further, the

phylogenetic tree constructed based on SNPs (Fig. 4 A) underscores the unique genetic structure of the Dandong population, which diverged earliest from the others. The sequential divergence of the Guangxi, Zhoushan, and Japan populations supports the South China Sea refuge hypothesis, reinforcing the notion that ancestral populations sought refuge in tropical waters during cold periods and dispersed northward during warmer interglacial intervals. SweeD selection sweep analysis for each population showed that the Dandong population had an unusually high number and score of selected sites compared to the other populations (Fig. 4). The other three populations exhibited increasing numbers and scores of selected sites from south to north on Chr2 and Chr7. We partitioned the genome of each chromosome into distinct loci at a distance of 10 kilobase pairs (Kbp) and calculated the SweeD statistical score for each locus. The magnitude of this score represents the intensity of selection on the corresponding sequence segment. We ranked the scores in descending order and selected approximately 700 loci, representing the top 1% of the highest scores, both from the Dandong population and across all populations studied. Subsequently, we extracted the genes residing within these selected loci based on annotated GFF (General Feature Format) files and subjected these genes to blast analysis using the NCBI database. These selected sites across all populations include genes related to energy synthesis and transport, such as various apolipoproteins, several mitochondrial energy metabolism enzymes such as *DLD* (*dihydrolipoyl dehydrogenase, mitochondrial*), *AGK* (*acylglycerol kinase, mitochondrial*), *IDH* (*isocitrate dehydrogenase [NADP], mitochondrial*), and enzymes involved in very long-chain fatty acid synthesis like *KAR*. This pattern of selection is evidently related to the increased energy metabolism demands associated with higher latitudes where these populations are found. The genes identified within the selected loci of the Dandong population included *PGAM* (*Phosphoglycerate Mutase*) and *TPI* (*Triosephosphate Isomerase A*), enzymes involved in glycolysis and gluconeogenesis, crucial for energy metabolism. In cold environments, the demand for energy metabolism typically increases, suggesting that these enzymes may play pivotal roles in adaptation to cold conditions. Furthermore, *GCL* (*Glutamate cysteine ligase*), a key enzyme in the synthesis of GGC (γ -glutamylcysteine), whose product is a direct precursor of glutathione, an essential antioxidant, was also identified. Glutathione aids organisms in

combating oxidative stress, which is particularly crucial for organisms residing in cold environments, as low temperatures can elevate the risk of oxidative stress (Wang H et al. 2022). We computed pairwise F_{ST} indices between different geographic populations. As shown in Fig. 5A, the results indicate that only the Dandong population exhibits relatively high F_{ST} values (>0.2) when compared to other populations, while F_{ST} values among the remaining populations are generally low. To emphasize the genetic differentiation of the Dandong population, we subsequently analyzed this population against a pooled group comprising the other three geographic populations (Fig. 5B). This comparison reveals genome-wide divergence between the Dandong population and the others, with F_{ST} values exceeding 0.25. Then we further selected the top 1% of genomic loci with the highest F_{ST} values for gene enrichment analysis. Notably, genes such as *ELOVL6* (elongation of very long chain fatty acids protein 6), *AEBP1* (adipocyte enhancer-binding protein 1), and *FASN* (fatty acid synthase) were identified. These genes are associated with lipid synthesis and adipocyte growth regulation, suggesting potential adaptive evolution in the Dandong population due to its prolonged residence in high-latitude cold waters. Additionally, we evaluated nucleotide diversity (π) and Tajima's D within each population, as presented in Fig. 5C and Fig. 5D. The results indicate similar trends in genetic diversity across chromosome positions among the populations.

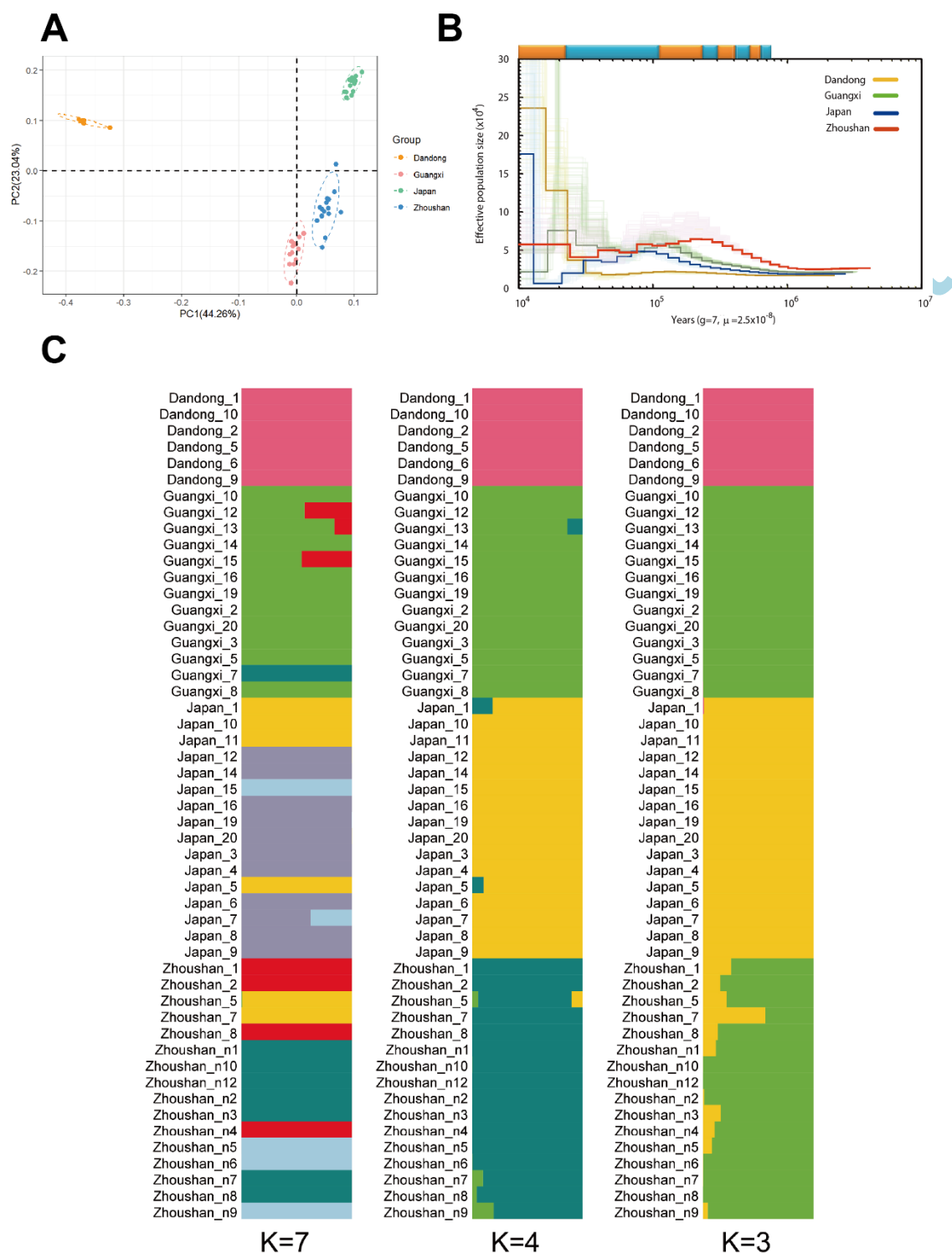


Fig. 3 **A.** PCA analysis of the Chinese herring of the four populations. **B.** PSMC (Pairwise Sequentially Markovian Coalescent) population history plot for the Chinese herring with 100 bootstrap replicates. The top bar indicates glacial periods in blue and interglacial periods in

orange. C. Admixture results for the Chinese herring of the four populations.

A

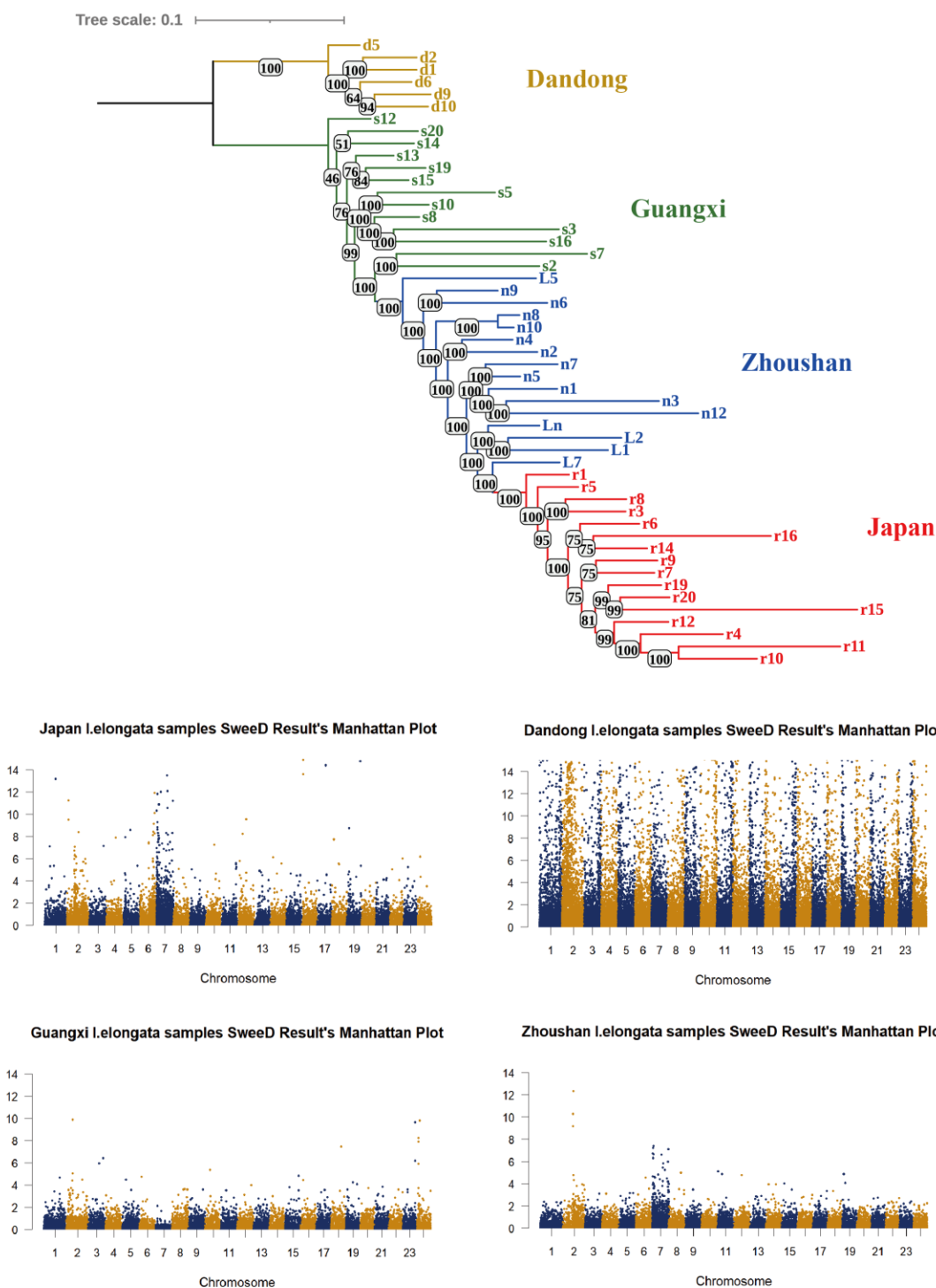


Fig. 4. **A.** Phylogenetic trees constructed using SNP data with IQ-TREE; **B.** Manhattan plots of SweeD results for the four populations of the Chinese herring.

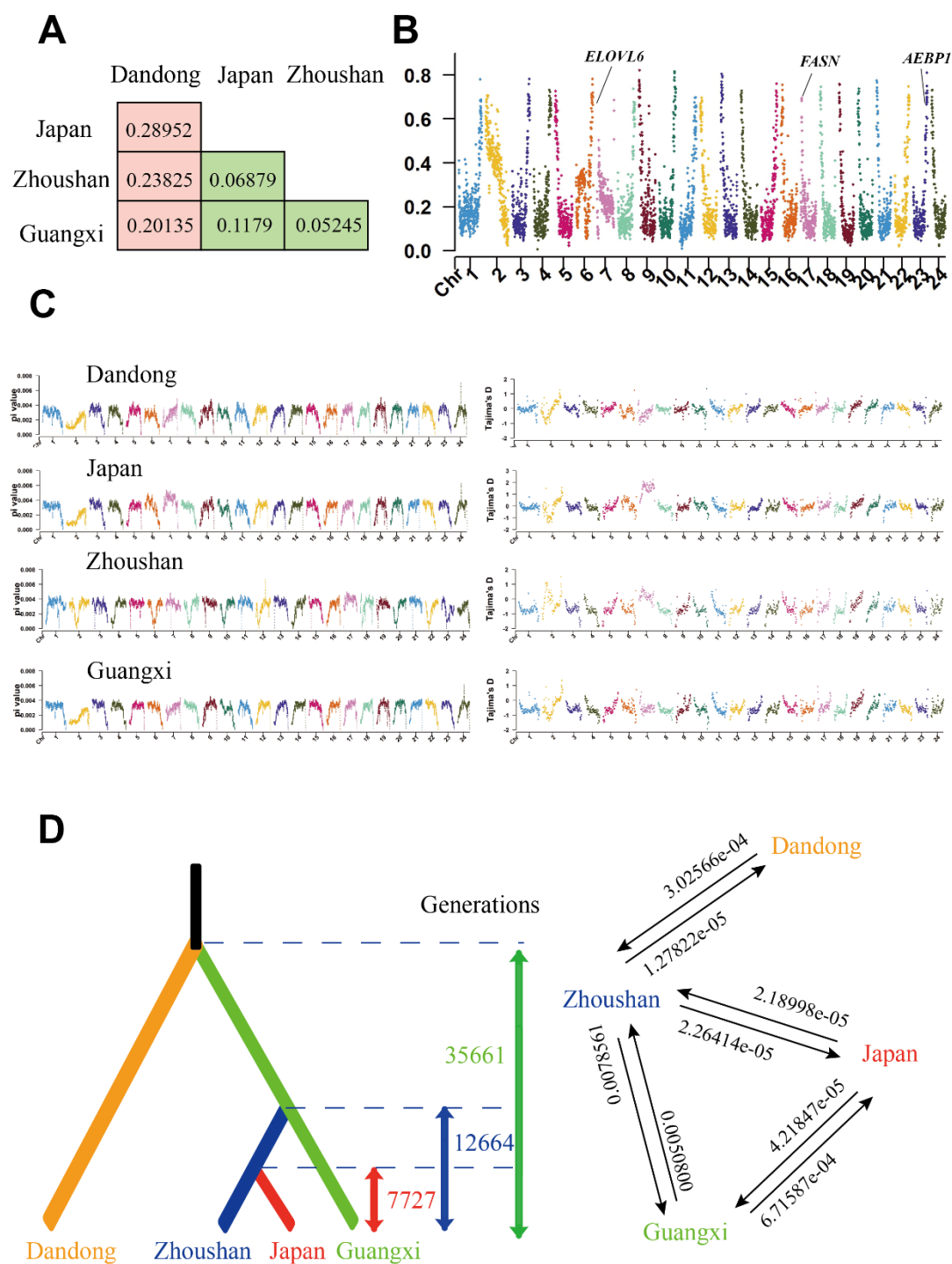


Fig. 5. **A.** The genome-wide Weir and Cockerham weighted F_{ST} estimate values between every two of the four populations of the Chinese herring. **B.** The genome-wide weighted F_{ST} results comparing the Dandong population to all other samples considered as a single population. **C.** The π (π) values and the Tajima's D for each population. **D.** Results of simulate historical

population models (Left) and estimated gene flow between adjacent populations (Right) using Fastsimcoal 2.8.

Comparative genomics analysis

We used OrthoFinder to identify single-copy orthologous genes among nine bony fish species (*Ilisha elongata*, *Alosa alosa*, *Danio rerio*, *Alosa sapidissima*, *Ictalurus punctatus*, *Larimichthys crocea*, *Epinephelus lanceolatus*, *Thunnus albacares*, *Siniperca chuatsi*), identifying a total of 5,528 single-copy orthologous genes. A Venn diagram illustrates the number of unique or shared genes among six of these species (Fig. 6 A).

Gene family expansion is a key mechanism driving phenotypic diversity and evolutionary adaptation. We compared gene families across these nine fish species and found 813 expanded ($p < 0.05$) and 3,727 contracted ($p < 0.05$) gene families in the Chinese herring (Fig. 7 A). Compared to other fish species, the Chinese herring exhibits a higher number of contracted gene families, including the *ELOVL* (*elongation of very long chain fatty acids protein*) and *KAR* (*very-long-chain 3-oxoacyl-CoA reductase*) gene families, which undergone rapid contraction. These gene families are primarily involved in the early steps of very long-chain fatty acid elongation. The rapid contraction of these gene families suggests that the synthesis of various long-chain unsaturated fatty acids may be suppressed in the Chinese herring compared to the American shad.

Fat accumulation is a polygenic trait influenced by various physiological pathways, including hormonal regulation, energy metabolism, and muscle growth. To investigate the impact of selective pressure on fat accumulation, we conducted dN/dS analysis using PAML software, comparing single-copy orthologous genes among the nine fish species. We identified 1,273 positively selected genes in three clupeiform species like the Chinese herring and the American shad. These positively selected genes include those involved in the very long-chain fatty acid synthesis process (Fig. 7 B), such as *ELOVL1*, *KAR*, and *HACD*, *HSL* (*hormone-sensitive triglyceride lipase*), as well as key genes related to intermuscular spine formation, including *Runx2b*. Additionally, these positively selected genes were enriched in pathways related to signal transduction (Fig. 6 B, 6 C), digestive system, and the metabolism of fatty acids,

carbohydrates, and amino acids, suggesting a connection to the unique fat accumulation traits observed in clupeiform fishes.

Accepted Manuscript

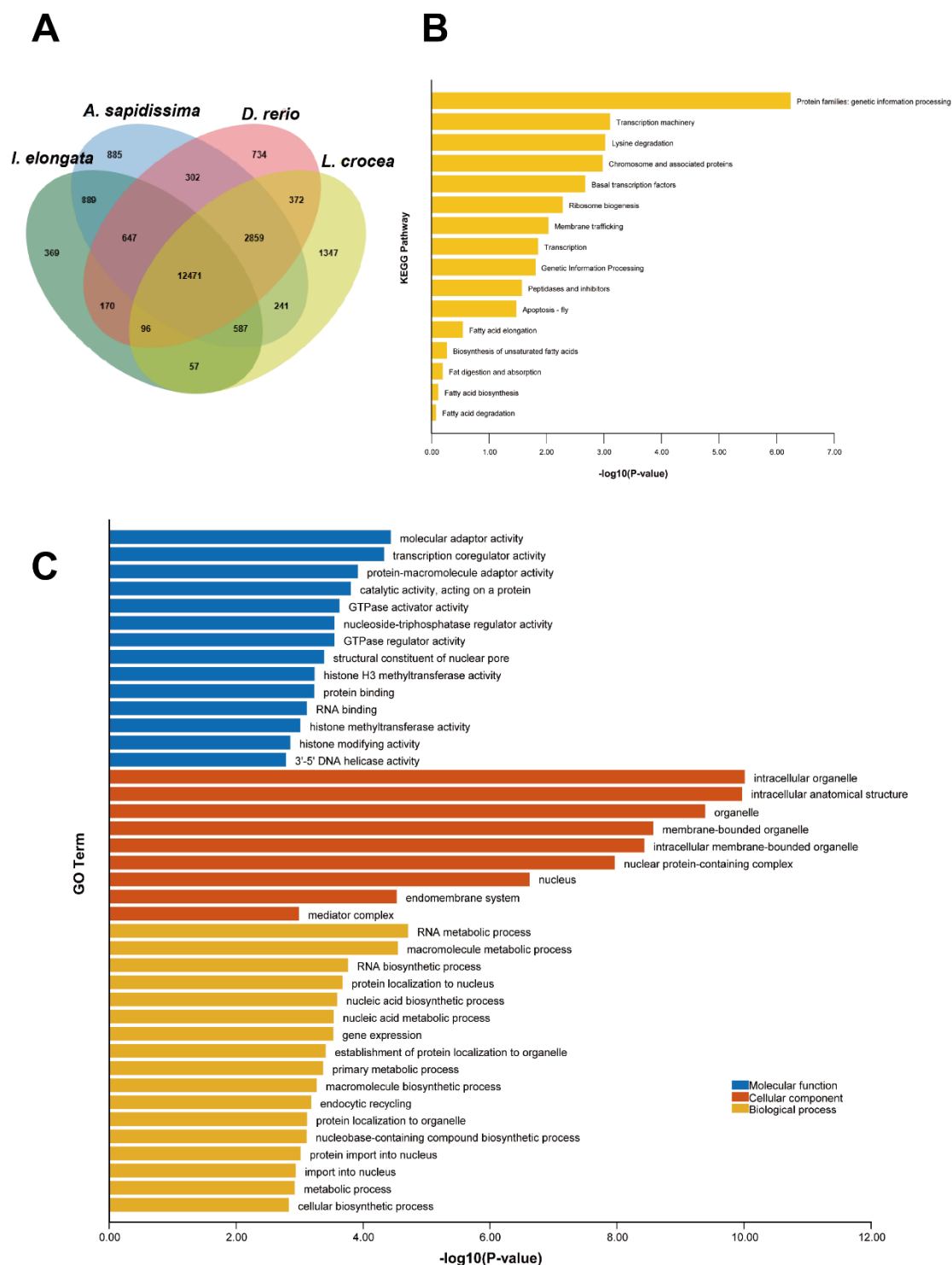


Fig. 6 **A**. Venn diagram illustrating the homologous genes shared among the Chinese herring and three other fish species. **B**. Bar chart presenting the KEGG pathway enrichment results for positively selected genes. **C**. Bar chart showing the Gene Ontology (GO) enrichment results for positively selected genes.

Discussion

The Chinese herring is one of the vital economic species in China. In recent years, the catches of Chinese herring in China have shown varying degrees of decline, with fish size decreasing and overfishing becoming increasingly evident, leading to a 31% reduction in production over the past decade. A clear understanding of the genetic diversity and population dynamics of the Chinese herring in the western Pacific, particularly along China's coast, is crucial for the effective formulation of relevant policies and the sustainable development of the Chinese herring resources. Previous studies based on nuclear and mitochondrial genes in the family Pristigasteridae have revealed its close genetic proximity to species in the family Alosidae, such as *Alosa sapidissima* (Wang et al., 2016). The collinearity plot presented in this study further supports this observation (Figure 2A). The Chinese herring, a warm-water fish, spawns upstream in estuarine areas near the coast and is highly sensitive to temperature. During colder seasons, it generally avoids more northern waters but exhibits high tolerance to low salinity. The spawning season typically coincides with the rise in temperature, occurring around April to June, and spawning is only likely when temperatures exceed 19°C, indicating a stringent temperature requirement. Chinese resources of the Chinese herring can be divided into several distinct populations based on their migration routes, reproductive habits, and fishing grounds. For instance, the populations in the Yellow Sea and Bohai Sea primarily migrate through Haizhou Bay, the southern coast of Shandong Peninsula, northern Yellow Sea, and the three bays of Bohai Sea, spawning from May to June annually. Those distributed outside the Yangtze River Estuary and in the Zhoushan fishing grounds in the East China Sea migrate mainly through southern Yellow Sea, Yangtze River Estuary, and the Zhoushan Islands, spawning from May to July. Meanwhile, populations in the South China Sea migrate primarily through the Gulf of Tonkin, Leizhou Peninsula, and areas near the Pearl River Estuary, spawning from March to May annually. Dandong, surrounded by the Liaodong Peninsula and the Korean Peninsula, has a closed sea area with unique hydrological conditions. Estuaries here, such as the Yalu River Estuary, experience lower temperatures compared to those of other populations during the

same period, leading to differences in spawning and migration patterns among the local populations.

The unique genetic structure of the Dandong population can be elucidated through both geographic and biological factors. The Dandong region, situated in the northeastern part of East Asia near the Korean Peninsula, is characterized by a temperate climate that is markedly different from the habitats of other populations. This environmental difference may have driven the Dandong population to develop specific climatic adaptations, particularly to the cold conditions prevalent in winter. Such climatic pressures have led to the evolution of distinct genetic traits in the Dandong population compared to other populations. The cold environment imposes specific selective pressures on the physiological mechanisms of the fish, and the food resources available may also differ from those encountered by other populations. This adaptive evolution likely involves genetic changes related to cold tolerance and metabolic rate, such as *PGAM* and *TPI* genes identified in the SweeD analyses.

Furthermore, the migratory pathways between Dandong and other populations are obstructed by the Liaodong Peninsula and the Korean Peninsula, resulting in restricted gene flow between the Dandong population and others throughout their evolutionary history. The breeding period for the Dandong population occurs annually from April to June, during which individuals migrate from overwintering grounds to shallow estuarine areas with low salinity for reproduction. This breeding behavior restricts genetic exchange primarily to a few estuarine locations, such as the Yalu River estuary, thereby reducing gene flow with other populations and leading to reproductive isolation and genetic differentiation.

Based on the best-fit model derived from Fastsimcoal simulations, the ancestral lineage of the Chinese herring diverged into two distinct lineages approximately 35,661 generations ago (Fig.5 D). One lineage gave rise to the Dandong population, while the other lineage contributed to the formation of the remaining three populations. This ancient divergence contrasts sharply with the more recent separation of the Japanese and northern populations, which occurred approximately 7,727 generations ago from their shared ancestral lineage. In terms of gene flow, the analysis revealed that the Guangxi population and the Zhoushan population exhibit the

highest levels of genetic exchange among the studied groups, indicating significant ongoing or historical interactions between these two populations.

The differences in F_{ST} values among populations also corroborate the distinctiveness of the Dandong population. The F_{ST} value differences between the Dandong population and other populations are significantly greater than those observed within other populations. This finding aligns with the results from Pairwise Sequentially Markovian Coalescent (PSMC) analysis, which indicates that while historical population expansions of other groups follow similar trajectories, the Dandong population exhibits a markedly different pattern.

Phylogenetic tree analysis further reveals that the Dandong population is the most ancient lineage of the Chinese herring, having diverged earliest from the ancestral lineage. This branch remained at the Dandong's Yalu River estuary, unable to expand due to climatic and geographical barriers. In contrast, another branch migrated southward to avoid the cold of glacial periods, eventually settling in the South China Sea, forming the Guangxi population, and gradually expanding northward to ultimately establish the Japanese population.

At first glance, the Dandong population appears geographically central, suggesting potential gene flow with other populations. However, this perspective overlooks the spawning habits of Chinese herring, which have specific temperature requirements for reproduction. They spawn exclusively in the brackish waters of familiar river estuaries from April to June, rather than in any arbitrary marine area. Although the Dandong population may encounter other populations through migration or ocean currents, gene exchange would only occur during their breeding season in the Yalu River estuary.

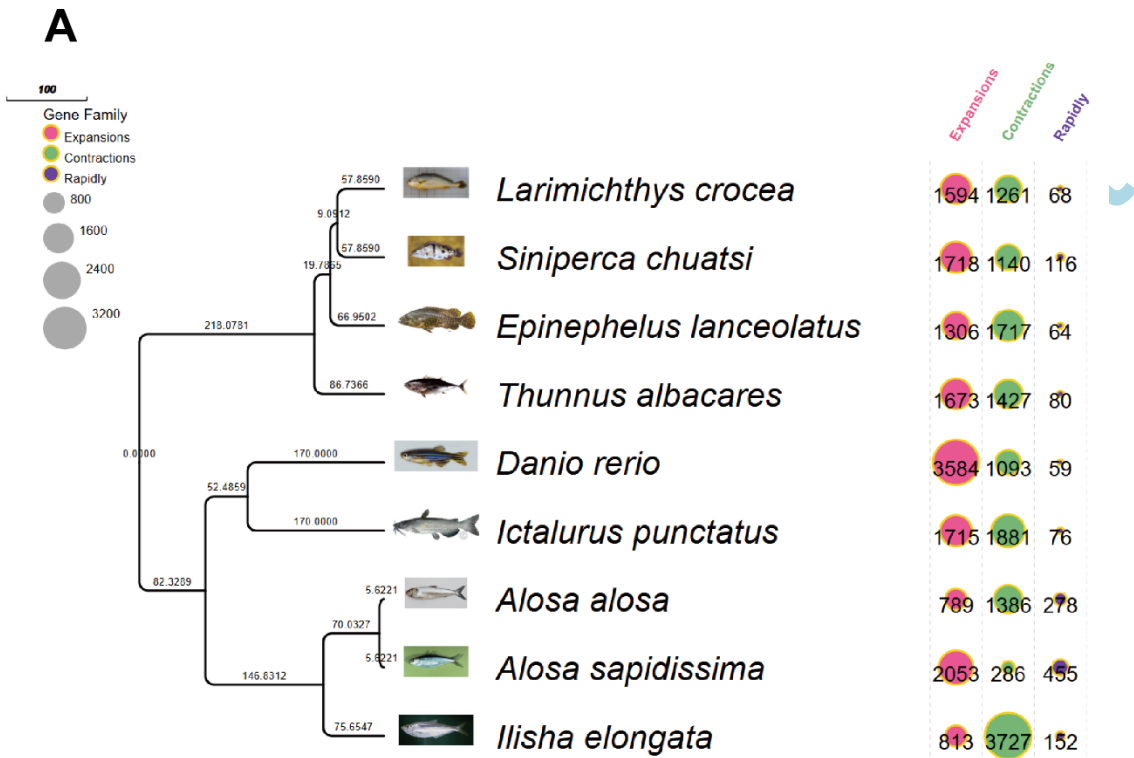
According to Zhang and Takita's research (2007), they propose that populations from Japan and the South China Sea develop in similar environments. Ocean currents may also play a significant role in population divergence. Populations from the South China Sea to Japan migrate along the warm Kuroshio Current of the western Pacific, whereas the Dandong population, blocked by the Korean Peninsula, is influenced only by internal circulation within the Yellow Sea. This geographic isolation impedes gene flow between the two populations.

In his book "Occasional Enjoyment in a Free Mood," Qing Dynasty author Li Yu (circa 1667) noted the distinct flavors between Chinese herring from the northern Yellow Sea and Bohai Sea and Reeves shad, highlighting early Chinese recognition of differences among fish species despite the absence of modern classification techniques. Recent studies have focused on the fatty acid composition of these fish species. Both Chinese herring and various shads, including *Alosa* and *Tenualosa*, belong to the clupeiformes order, known for being rich in fatty acids that can supplement ω -3s not synthesized by the human body. However, environmental changes such as overfishing and hydrological projects in the Yangtze River have led to the functional extinction of Reeves shad in that region, while Chinese herring populations have declined but not to the same extent (Blaber et al., 1998).

Comparative analysis has revealed variations in the ELOVL gene family between Chinese herring and American shad (*Alosa sapidissima*), which may explain differences in their fatty acid profiles (ELOVL gene family study). The ω -3 PUFAs, particularly DHA and EPA, are crucial for human nutrition and health, with Chinese herring being especially rich in these fatty acids (Dessi et al., 2013; Jiang et al., 2016). Studies have quantified the fatty acid composition in the muscle tissue of Chinese herring and American shad, showing significant differences in PUFA profiles (Liu et al., 2011; Huang et al., 2013; Hong et al., 2013; Gu et al., 2007).

While both species are large-bodied clupeiforms with substantial fat reserves, the American shad exhibits higher overall PUFA levels dominated by ω -6 linoleic acid but lower levels of ω -3 EPA and DHA compared to the Chinese herring. This difference may be attributed to the expansion of the ELOVL gene family in the American shad, enabling greater synthesis of PUFAs, with EPA and DHA possibly serving as precursors for elongation into very long-chain LC-PUFAs (ELOVL genes and fatty acid synthesis study). Moreover, research indicates that fatty acid metabolism and conversion between saturated and unsaturated fatty acids are vital for many fish, especially in cold environments (He et al., 2015; Sun et al., 2019). Distinct migratory behaviors likely influence fatty acid proportions, with long-distance migrating shads possibly storing more PUFAs resistant to β -oxidation, while estuary-spawning Chinese herring may have evolved alternative storage strategies. These above analysis still requires further experimental validation. In conclusion, although both Chinese herring and American shad are large, fat-storing

clupeiforms, their fatty acid profiles significantly differ, potentially due to genetic distinctions in the ELOVL gene family and environmental factors like migration patterns and life history traits.



B

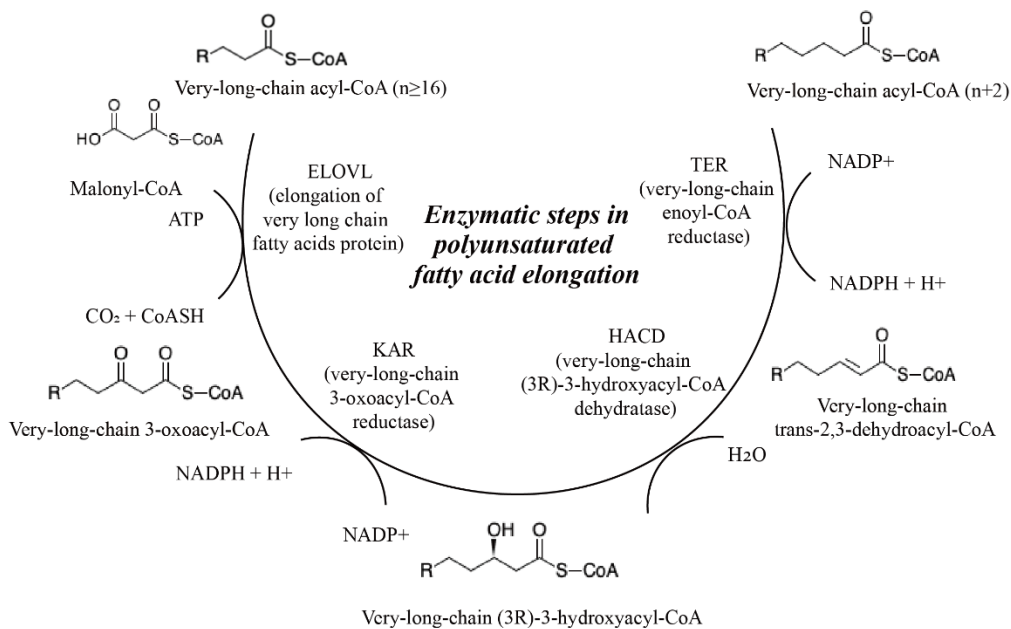


Fig. 7 **A.** Phylogenetic tree of *Ilisha elongata* and other fishes showing gene family expansion and contraction. **B.** Role of various genes in metabolic pathways during the process of very long-chain fatty acid elongation.

Conclusion

This study presents the first chromosome-level genome assembly of the Chinese herring. Population genetic analysis reveals four distinct populations, with the Dandong population diverged the earliest. Comparative genomic analysis identified a rapid contraction of gene families associated with very-long-chain fatty acid elongation, and genes linked to fatty acid metabolism and intermuscular bone formation are under positive selection. These findings elucidate the mechanisms underlying the differential fatty acid ratios in riverine versus marine migratory fishes within the Clupeiformes, highlighting evolutionary adaptations of the Chinese herring to its ecological niche.

Materials and Method

Chromosome-level genome assembly and annotation

Genomic DNA was extracted from liver samples using the CTAB method according to the QIAGEN Genomic DNA Extraction Kit (Cat#13343, QIAGEN). Approximately 3-4 µg of DNA was used for the construction of a DNA library, which was then sequenced on the Nanopore PromethION sequencer (Oxford Nanopore Technologies, UK) at Novogene. An Illumina sequencing library was prepared using a modified method where DNA was sheared to 250 bp, as described by Meyer and Kircher (2010). This library was sequenced on the Illumina HiSeq platform at Novogene. The sequence data were used to polish the assemblies derived from the nanopore sequencing results. Hi-C library construction and sequencing were performed at Novogene, following the pipeline established by Belton et al. (2012). For RNA library

construction and sequencing, one microgram of RNA was extracted from each tissue (heart, liver, brain, kidney, skin, eye, muscle, swim bladder, gill) and used for RNA library preparation. Next-generation sequencing library preparation was conducted according to the manufacturer's protocol. The sequencing results were then used for genome annotation purposes.

Genomescope2.0, a tool for estimating genome size based on k-mer frequency distributions was employed to analyze the sequencing data (Vurture et al., 2017). The 21 k-mer distribution from the Nanopore paired-end sequencing files was first calculated using jellyfish, and the output was subsequently used with Genomescope2.0 to statistically determine the genome size of the Chinese herring. NextDenovo, a tool for assembling long-read data from Pacbio and Nanopore third-generation sequencing was used to generate an initial draft genome from the Nanopore paired-end sequencing files with the estimated genome size (Hu et al., 2024). To enhance the quality of the genome assembly, a two-step error correction process was performed: NextPolish was used to correct errors by Illumina data in the genome resulting from noisy long reads (Hu et al., 2020), followed by Purge-Haplotigs to split heterozygous regions of the genome (Roach et al., 2018).

HiC-Pro was used for quality control of the Hi-C raw data (Servant et al., 2019). Bowtie2 was used to align the raw reads to the assembled draft sequence, filtering out low-quality reads (Langmead et al., 2009). Juicer, an open-source tool for Hi-C dataset analysis (Durand et al., 2016) and 3D-dna were then employed to anchor the primary scaffolds to 24 chromosomes (Dudchenko et al., 2017). The JuiceBox visualization system was used for further analysis (<https://github.com/aidenlab/juicebox>, accessed by June 2024). The integrity of the genome assembly was assessed by aligning it with the actinopterygii_odb10 database using Benchmarking Universal Single-Copy Orthologues (BUSCO) (Simão et al., 2015).

Prior to gene prediction, repeat elements within the Chinese herring genome were annotated. RepeatModeler (Flynn et al., 2020) and Repeatmasker (Tarailo-Graovac et al. 2009) were used in concert to align discovered repeat sequence models with the genomic sequence, marking the positions of repeats within the genome. Based on the genome sequence with repeats masked, gene prediction for the Chinese herring was conducted. The prediction was

based on the transcriptome data of the Chinese herring from nine tissues and protein sequences from closely related species. Trinity was used to assemble transcriptomic paired-end data from the nine tissues (Grabherr et al., 2011), which was then deduplicated using cd-hit (<https://github.com/weizhongli/cdhit/releases>). Protein annotation results from *Danio rerio* and other clupeiform fishes and related species, downloaded from NCBI were used as a database of closely related homologous species. Gene prediction was conducted three rounds with Maker (Cantarel et al., 2008) on repeat-masked assembly, transcriptomic data, and homologous protein sequences. Augustus (Stanke et al., 2008) and SNAP (Leskovec et al., 2016) models were trained with artificial training sets for the Chinese herring gene prediction processing. Augustus training data were derived from GFF file's mRNA annotations by extracting $\pm 1,000$ bp flanking sequences, adjusted for genome boundaries. The dataset was refined via BUSCO v5 (actinopterygii_odb10 lineage) to ensure evolutionary gene conservation, generating a species-specific model. For SNAP, GFF file was filtered (maker2zff; coverage/score thresholds=0.8), partitioned into training/validation sets (fathom), and processed into an HMM (forge). Blastp was used for homology search in public gene databases such as NCBI-NR, KEGG, Pfam, and SwissProt to functionally annotate the predicted genes in the Chinese herring genome (e-value $\leq 1e-05$). InterProScan (Jones et al., 2014) was used to obtain protein domain annotations from the Interpro database and "Gene Ontology" (GO). The most consistent functional annotations from various databases were selected as the final consensus gene annotation results. Upon completing the genome assembly, PyCircos (<https://github.com/ponnhide/pyCircos>) was used to display the genomic features of the chromosomes. Additionally, we employed Jcvi (<https://github.com/tanghaibao/jcvi>) to illustrate the genomic collinearity between the Chinese herring and *A. alosa* as well as *C. harengus*.

Resequencing and population genomic analysis

Fifty one samples from four sites (Liaoning Province, Dandong City; Zhejiang Province, Zhoushan City, Ariake Sea, Japan; and Guangxi Province, Beihai City, Fig. 1 B) of the Chinese herring were collected and preserved in 95% ethanol and subsequently stored in a refrigerator

at 4°C. All experimental procedures involving fish sampling were approved by the Animal Ethics Committee of Shanghai Ocean University, China (2020). DNA extraction was performed using the Ezup Column Animal Genomic DNA Purification Kit according to the manufacturer's instructions (Sangon, Shanghai, China). The concentration of the purified DNA was measured using a NanoDrop 3300 Fluorospectrometer (Thermo Fisher Scientific, Wilmington, DE, United States) and visualized by agarose gel electrophoresis. The purified DNA was fragmented to ~250 bp using the Xinzhi Bio-Ultra DNA Fragmentator (Scientz, Ningbo, China), which involved 22 cycles of ultrasonic treatment for 20 seconds followed by a 20-second pause, with a processing power of 300W.

The preparation of the resequencing library was modified according to the method of Meyer and Kircher (Meyer et al., 2010), where each sample was labeled with a pair of inline indexes during the ligation step of library preparation to reduce the potential risk of data cross-contamination between samples (Wang et al., 2022). The library was then amplified with a pair of P5 and P7 primers, each with an 8 bp DNA index. The amplified products were mixed in equal proportions, concentrated, and size-selected using the QIAquick Gel Extraction Kit (QIAGEN China, Shanghai, China), retaining fragments between 250 and 1000 bp. The purified products were sequenced on the Illumina HiSeq 2500 sequencer by Novogene.

The raw sequencing reads were demultiplexed based on their 8 bp P5, P7 indexes and 6 bp inline indexes. Adapter sequences and reads with low-quality bases ($Q < 20$) were trimmed using Trim_galore v0.6.7 (<https://github.com/FelixKrueger/TrimGalore>) and Cutadapt v1.2.1 (Martin et al., 2011) with default parameters. Data trimming was performed using a Perl script (Yuan et al., 2019), Variant calling was performed using GATK v4.2.6.0 (McKenna et al., 2010) following the Best Practices workflow. Raw SNPs were filtered through GATK's VariantFiltration with hard thresholds recommended for diploid genomes: quality score (QUAL) < 30.0 , quality depth (QD) < 2.0 , Fisher strand bias (FS) > 60.0 , and strand odds ratio (SOR) > 4.0 , which were then visualized with Plink (version 1.9; Purcell et al., 2007) and admixture (Alexander et al., 2013) and analyzed for population selection signals using SweeD (Pavlidis et al., 2013). Maximum likelihood phylogenetic trees for each resequenced individual were reconstructed using IQ-TREE v2.2.0 (Nguyen et al., 2015) under the optimal substitution model (TVM+F+R4)

selected by ModelFinder, with filtered SNP datasets retaining variants showing $\leq 10\%$ missing genotypes ($--geno < 0.1$) and minor allele frequency $\geq 5\%$ ($--maf < 0.05$). To quantify genetic differentiation between populations, we employed VCFtools (Danecek, P. et al., 2011) to calculate the genome-wide weighted average F_{ST} values for each pairwise population comparison. Specifically, for the Dandong population and all other populations, we computed weighted average F_{ST} values using a sliding window approach with a step size of 150k bp across the chromosomal genomes. Furthermore, we estimated nucleotide diversity (π) and Tajima's D for each population as measures of within-population genetic variation and neutrality testing. The results from these analyses were visualized using the R package CMplot. Simultaneously, we selected one individual from each of the four populations, Dandong, Zhoushan, Japan, and Guangxi, and processed their resequencing data at thirty times genome coverage. These data were used to analyze the historical dynamics of each population using a HMM-based Pairwise Sequentially Markovian Coalescent (PSMC) approach (Li et al. 2011). Using the `psmc_plot.pl` script with the available options, we reconstructed the population history. Based on the generation time of 3 years (<https://fishbase.mnhn.fr/>), and applying the formula $g = a + [s/(1-s)]$, The g parameter was estimated to be 7. To ensure the reliability of the results, we performed 100 bootstrap replicates for the PSMC analysis of each population. Based on our population clustering and geographical distribution analysis, fish specimens from four distinct localities (Guangxi, Zhoushan, Japan, Dandong) were grouped into four principal lineages, designated as populations 0 through 3. We proposed four models to elucidate the evolutionary history of these populations: Models 1 to 3 posit varying sequences of divergence where population 3 splits significantly earlier than the others, with population 2 diverging either from the common ancestor of populations 0 and 1 or independently from one of them. Model 4 extends the best-fitting scenario among the first three by incorporating potential historical migrations between geographically proximate groups. To evaluate these hypotheses, we utilized FASTSIMCOAL 2.8 (fast sequential Markov coalescent simulation of genomic data under complex evolutionary models) (Excoffier et al., 2021) for demographic inference using the joint site frequency spectrum (JSFS) as a summary statistic. The JSFS was derived from VCF data via easySFS.py (<https://github.com/isaacovercast/easySFS>). For each model, we conducted 100 independent

point estimation runs, each comprising 100,000 coalescent simulations and 30 expectation-conditional maximization (ECM) iterations. Starting from random seeds in the parameter file, we compared models using the Akaike information criterion (AIC), assessed the confidence intervals of optimal estimates through bootstrapped SFS, and analyzed the likelihoods to gauge model fit. A minimal discrepancy between observed and estimated maximum likelihoods indicates superior model adequacy for the population's evolutionary narrative.

Comparative genomic analysis

To identify genes under potential positive selection and those that may be evolving rapidly, a set of single-copy orthologous genes common to nine species of teleost fish (*Ilisha elongata*, *Alosa alosa*, *Danio rerio*, *Alosa sapidissima*, *Ictalurus punctatus*, *Larimichthys crocea*, *Epinephelus lanceolatus*, *Thunnus albacares*, *Siniperca chuatsi*) were generated. Initially, multiple sequence alignments for each gene family were conducted using Orthofinder (Emms Et al., 2019). Jvenn (Bardou et al. 2014) was used to display the Venn diagram of homologous genes between the Chinese herring and other fish species.

The positive selection analysis used an integrated pipeline with PAML (Yang, 1997) to detect positive selection across orthogroups. Coding sequences were preprocessed by selecting the longest protein isoforms and removing in-frame stop codons through translation frame validation. OrthoFinder identified single-copy orthogroups using reciprocal sequence similarity searches, followed by codon-aware alignment processing: protein sequences were aligned with MAFFT-LINSI, back-translated to codon alignments, and refined using trimAl to remove low-quality regions (>95% gaps). The species tree topology from OrthoFinder was converted to PAML-compatible format by replacing taxon names with numerical IDs and annotating foreground branches. Site-specific selection was tested using CodeML under two evolutionary models—one allowing variable ω values (alternative hypothesis) and another enforcing neutral evolution ($\omega=1$, null hypothesis). Likelihood ratio tests quantified selection signals (χ^2 test, $2\Delta\ln L = 2[\ln L_{alt} - \ln L_{null}]$) with statistical significance determined at $\alpha=0.05$ after multiple-test correction. Key computational safeguards included identifier consistency checks, reading frame preservation during alignment, and automated topology conversion, ensuring reproducibility

through standardized parameters. Gene Ontology (GO) terms and KEGG pathway identifiers associated with positively selected genes were extracted for subsequent analysis. To elucidate the reasons why clupeiform fish exhibit higher proportion of fat and more intermuscular spines compared to other bony fish, the fish species were categorized into two groups: the clupeiform species (*A. alosa*, *A. sapidissima*, *I. elongata*), which are characterized by substantial fat storage and prominent intermuscular bones, and other comparative fish species, which lack these features. Additionally, Gene Ontology (GO) and KEGG enrichment analyses were used to highlight the functional enrichment of positively selected genes (p value $< 1 \times 10^{-5}$) identified through PAML. Then, we used TBtools (Chen et al. 2020) to visualize the Gene Ontology (GO) and KEGG enrichment results.

Expansion of gene families is considered one of the most important mechanisms contributing to phenotypic diversity and evolutionary adaptation to the environment (Rayna et al., 2015). To determine gene families in the *Ilisha elongata* genome, the genomes of 8 other vertebrates were selected and compared. Genes were categorized into orthologous genes, paralogous genes, and single-copy orthologous genes. CAFÉ (De Bie et al., 2006) was used to identify amplified and contracted gene families among the clupeiform species. We downloaded the exon protein annotation files and GFF files for these species from NCBI, and performed an all-by-all BLAST analysis using blastp for each protein. The BLAST results were subjected to MCL clustering, and categories with significant copy number variation were excluded. The species tree was inferred using OrthoFinder and SeqKit (<https://github.com/shenwei356/seqkit>) to quantify the alignment files on which the tree construction depended. Subsequently, ultrametric trees were constructed using r8s (<https://sourceforge.net/projects/r8s/>) and finally ran CAFÉ to obtain gene family results. Evolview (<http://www.evolgenius.info/evolview/#/treeview>) was used to visualize the expansions, contractions, and rapid changes at each node and made a time-calibrated phylogenetic tree to illustrate the patterns of gene family expansion and contraction across species.

Author contribution

TW, LL and CL conceived the research plan. TW, VJ, QW and JH did experiments and collected data. TW and LL analyzed the data. TW and CL wrote the draft. All authors edited and approved the final version of the manuscript.

Funding

This work was supported by the Science and Technology Commission of Shanghai Municipality (19410740500; 19050501900) to CL.

Data availability

The scaffold-level genome assemblies of the Chinese herring (SAMN43154019) have been deposited in NCBI (<https://www.ncbi.nlm.nih.gov/>) genome database. The Chromium-linked Illumina reads, RNA-Seq reads (Illumina HiSeq), and Nanopore reads used for the genome assembly have been deposited to NCBI genome database (SRA) under the BioProject accession number PRJNA1147427.

The other raw data and scripts utilized in this study have been deposited in the Dryad repository under the DOI:

http://datadryad.org/share/LP_NTJf5W07puQRVytuUCUIUjHeLK6zr6_xlpPxopBo. All other relevant data are available from the corresponding authors upon request.

Conflict of interest: the authors declare that there is no conflict of interest.

Reference

- Alexander, D.H., J. Novembre, and K. Lange, Fast model-based estimation of ancestry in unrelated individuals. *Genome Res*, 2009. 19(9): p. 1655-64.
- Andersson L, Bekkevold D, Berg F, et al. How Fish Population Genomics Can Promote Sustainable Fisheries: A Road Map. *Annu Rev Anim Biosci*. 2024;12:1-20. doi:10.1146/annurev-animal-021122-102933 *Aquaculture*, 185(3-4), 367-377.
- Bayse SM, Regish AM, McCormick SD. Proximate composition, lipid utilization and validation of a non-lethal method to determine lipid content in migrating American shad *Alosa sapidissima*. *J Fish Biol*. 2018 Jun;92(6):1832-1848. doi: 10.1111/jfb.13624. Epub 2018 Apr 24. PMID: 29603209.
- Bernatchez, L. (2016). On the maintenance of genetic variation and adaptation to environmental change: Considerations from population genomics in fishes. *Journal of Fish Biology*, 89(6), 2519–2556.
- Belton JM, McCord RP, Gibcus JH, Naumova N, Zhan Y, Dekker J. Hi-C: a comprehensive technique to capture the conformation of genomes. *Methods*. 2012 Nov;58(3):268-76. doi: 10.1016/j.ymeth.2012.05.001. Epub 2012 May 29. PMID: 22652625; PMCID: PMC3874846.
- Birge, T. L., Ralph, G. M., Di Dario, F., Munroe, T. A., Bullock, R. W., Maxwell, S. M., Santos, M. D., Hata, H., & Carpenter, K. E. (2021). Global conservation status of the world's most prominent forage fishes (Teleostei: Clupeiformes). *Biological Conservation*, 253, 108903.
- Blaber S J M, Staunton-Smith J, Milton D A, Fry G, Van der Velde T, Pang J, Wong P, Boon-Teck O. The biology and life-history strategies of *Ilisha* (Teleostei:Pristigasteridae) in the coastal waters and estuaries of Sarawak. *Estuarine, Coastal and Shelf Science*, 1998, 47(4):499-511.
- Cantarel, BL; Korf, I; Robb, SMC; et al. MAKER: An easy-to-use annotation pipeline designed for emerging model [J].*GENOME RESEARCH*, 2008. DOI: 10.1101/gr.6743907.

- Chen C., et al., TBtools: An Integrative Toolkit Developed for Interactive Analyses of Big Biological Data, *Molecular Plant*, Volume 13, Issue 8, 2020.
- Cheng Q. Analysis and Research on Marine Fish Unearthed in China 5,000 Years Ago [J]. *Marine Sciences*, 1981, (04): 1-5.
- Danecek, P. et al. (2011). The variant call format and VCFtools. *Bioinformatics*, 27, 2156–2158.
- De Bie, T., Cristianini, N., Demuth, J. P., & Hahn, M. W. (2006). CAFE: A computational tool for the study of gene family evolution[J]. *Bioinformatics*, 22(10), 1269-1271.
- Dessi M, Noce A, Bertucci P, et al. Atherosclerosis, dyslipidemia, and inflammation: the significant role of polyunsaturated Fatty acids[J]. *ISRN Inflamm*, 2013(3): 191823
- Dudchenko O, Batra S S, Omer A D, et al. De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds[J]. *Science*, 2017, 356(6333): 92-95.
- Durand, N., Shamim, M., Machol, I., Rao, S. P., Huntley, M., Lander, E., & Aiden, E. L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. [J]. *Cell Systems*, 3, 95-98.
- Emms, DM; Kelly, S. OrthoFinder: phylogenetic orthology inference for comparative genomics [J]. *GENOME BIOLOGY* 2019. DOI: 10.1186/s13059-019-1832-y.
- Excoffier, L. et al. 2021. fastsimcoal2: demographic inference under complex evolutionary scenarios. *Bioinformatics*, 37, 4882-4885.
- Flynn, JM; Hubley, R; Goubert, C; Rosen, J; Clark, AG; Feschotte, C; Smit, AF. RepeatModeler2 for automated genomic discovery of transposable element families. *PNAS*. DOI: 10.1073/pnas.1921046117.
- Fricke, R., Eschmeyer, W.N., Fong, J.D., 2022. Eschmeyer's catalog of fishes: genera/ species by family/subfamily, in: Fricke, R., Eschmeyer, W. N. & R. van der Laan (Eds.), Eschmeyer's catalog of fishes: genera, species, references.asp. Electronic version accessed 6 June 2022.

- Grabherr, MG; Haas, BJ; Yassour, M; Levin, JZ; et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome [J]. NATURE BIOTECHNOLOGY 2011. DOI: 10.1038/nbt.1883.
- Guillou H, Zadavec D, Martin P G P, et al. The key roles of elongases and desaturases in mammalian fatty acid metabolism: Insights from transgenic mice [J]. Progress in Lipid Research, 2010, 49(2): 186-199.
- Gu, R. B., Zhang, C. X., Xu, G. C., et al. (2007). Analysis and evaluation of nutritional components in muscle of American shad. Journal of Fisheries of China, (02), 40-46.
- He, J. Qiang H. Yang, P. Xu, Z.X. Zhu, R.Q. Yang. Changes in the fatty acid composition and regulation of antioxidant enzymes and physiology of juvenile genetically improved farmed tilapia *Oreochromis niloticus* (L.), subjected to short-term low temperature stress. J. Therm. Biol., 53 (2015), pp. 90-97
- Hong, X., Xie, W., Zhu, X., et al. (2013). Comparison of nutritional components in muscle of American shad and Indian shad. Acta Nutrimenta Sinica, 35(02), 206-208. doi:10.13325/j.cnki.acta.nutr.sin.2013.02.003.
- Huang, F. Y., He, X. M., Zhang, X. H. (2013). Determination of fatty acids in muscle of *Ilisha elongata* by accelerated solvent extraction-gas chromatography/mass spectrometry. Fisheries Science & Technology Information, 34(07), 34-36.J.
- Hu, J; Fan, JP; Sun, ZY; Liu, SL. NextPolish: a fast and efficient genome polishing tool for long-read assembly. [J]. BIOINFORMATICS 2020. DOI: 10.1093/bioinformatics/btz891.
- Hu, J; Wang, Z; Sun, ZY; Hu, BX; Ayoola, AO; Liang, F; et al. NextDenovo: an efficient error correction and accurate assembly tool for noisy long reads. [J]. GENOME BIOLOGY 2024. DOI: 10.1186/s13059-024-03252-4.
- Jakobsson A, Westerberg R, Jakobsson A. Fatty acid elongases in mammals: Their regulation and roles in metabolism [J]. Progress in Lipid Research, 2006, 45(3): 237-249.

- Jiang, Y., Xiong, W. K., Yin, J. L., et al. (2016). Research progress on the intake of ω -3 and ω -6 polyunsaturated fatty acids in diet and cardiovascular health. *Cereals & Oils*, 29(11), 1-5.
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., ... Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, 30, 1236-1240.
- Kim, J.-K., Choi, O.-I., Kim, J.-I., Chang, D.-S. & Park, K.-D. 2007. Age and growth of the elongate ilisha *Ilisha elongata*. *Fisheries and Aquatic Sciences*, 10, 30-36.
- Langdon C.J., Waldock M.J.. The effect of algal and artificial diets on the growth and fatty acid composition of *crassostrea gigas* spat. *J. Mar. Biol. Assoc. U. K.*, 61 (2) (1981), pp. 431-448
- Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biology*, 10, R25.
- Leskovec, J., & Sasic, R. (2016). SNAP: a general purpose network analysis and graph mining library. *ACM transactions on intelligent systems and technology*, 8, 1.
- Li H., Durbin R. (2011). Inference of human population history from individual whole-genome sequences. *Nature*. 475, 493–496. doi: 10.1038/nature10231
- Li P. et al. Genetic Variation Study of *Clupea harengus* in the East China Sea Based on COI Gene and D-loop Sequences [J]. *Journal of Zhejiang Ocean University (Natural Science Edition)*, 2013, 32(02): 93-98.
- Liu, H. H., Chang, K. M., Zheng, X. L. (2011). Nutritional component analysis of *Ilisha elongata* from Zhoushan sea area. *Acta Nutrimenta Sinica*, 33(06), 624-626. doi:10.13325/j.cnki.acta.nutr.sin.2011.06.003.
- Lv, Z., Xu, Y., Wu, C., Fan, Z. & Zhang, J. 2010. Genetic variation in different populations of *Ilisha elongata* in China coastal water based on 16 S rRNA gene analysis. *Journal of Fishery Sciences of China/Zhongguo Shuichan Kexue*, 17.
- Martin, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet. journal*, 17, 10-12.

- McKenna, A., Hanna, M., Banks, E., Sivachenko, A., Cibulskis, K., Kernytsky, A., Garimella, K., Altshuler, D., Gabriel, S., Daly, M., & DePristo, M. A. (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data.
- Meyer, M. & Kircher, M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor Protocols, 2010, pdb. prot5448.
- Nelson, J.S., Grande, T.C., Wilson, M.V.H., 2016. Fishes of the World, fifth ed. John Wiley & Sons.
- Nguyen, H.A. Schmidt, A. von Haeseler, and B.Q. Minh (2015) IQ-TREE: A fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol. Biol. Evol., 32, 268-274.
- Norris AT, Bradley DG and Cunningham EP. 2000. Parentage and relatedness determination in farmed Atlantic salmon (*Salmo salar*) using microsatellite markers, Aquaculture 182, 73-83.
- Pavlidis, P; Zivkovic, D; Stamatakis, A; Alachiotis, N. SweeD: Likelihood-Based Detection of Selective Sweeps in Thousands of Genomes. [J]. MOLECULAR BIOLOGY AND EVOLUTION DOI:10.1093/molbev/mst112.
- Philippe B. et al. jvenn: an interactive Venn diagram viewer. BMC Bioinformatics 2014, 15:293 doi:10.1186/1471-2105-15-293
- Pinsky, M. L. & Palumbi, S. R. 2014. Meta - analysis reveals lower genetic diversity in overfished populations. Molecular ecology, 23, 29-39
- Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M. A., Bender, D., Maller, J., Sklar, P., de Bakker, P. I., Daly, M. J., & Sham, P. C. (2007). PLINK: A tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics, 81(3), 559–575
- Rayna, M. H., & Hans, A. H. (2015). Seeing is believing: Dynamic evolution of gene families [J]. PNAS, 112(5), 1252-1253.

- Roach, M. J., Schmidt, S. A., & Borneman, A. R. (2018). Purge Haplotigs: Allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*, 19(1), 1–10
- Servant, N., Varoquaux, N., Lajoie, B. R., Viara, E., Chen, C. J., Vert, J. P., Heard, E., Dekker, J., & Barillot, E. (2015). HiC-Pro: an optimized and flexible pipeline for Hi-C data processing. *Genome Biology*, 16, 259.
- Simão, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V., & Zdobnov, E. M. (2015). BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, 31(19), 3210–3212.
- Slatkin M and Hudson RR. 1991. Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genet* 129, 555-562.
- Stanke, M; Keller, O; Gunduz, I; Hayes, A; Waack, S; Morgenstern, B. AUGUSTUS:: ab initio prediction of alternative transcripts [J]. *NUCLEIC ACIDS RESEARCH*, 2008. DOI: 10.1093/nar/gkl200.
- Tarailo-Graovac, M., & Chen, N. (2009). Using RepeatMasker to identify repetitive elements in genomic sequences. *Current Protocols in Bioinformatics*, Chapter 4, Unit 4.10.
- Vurture, GW; Sedlazeck, FJ; Nattestad, M; Underwood, CJ; Fang, H; Gurtowski, J; Schatz, MC. GenomeScope: fast reference-free genome profiling from short reads [J]. *BIOINFORMATICS* 2017. DOI:10.1093/bioinformatics/btx153.
- Wang H, Wang Y, Niu M, Hu L, Chen L. Cold Acclimation for Enhancing the Cold Tolerance of Zebrafish Cells. *Front Physiol*. 2022 Jan 28;12:813451. doi: 10.3389/fphys.2021.813451. PMID: 35153820; PMCID: PMC8832062.
- Wang, Q., Zhang, J., Matsumoto, H., Kim, J.-K. & Li, C. 2016. Population structure of elongate ilisha *Ilisha elongata* along the Northwestern Pacific Coast revealed by mitochondrial control region sequences. *Fisheries science*, 82, 771-785.

- Wang Q et al. Molecular phylogenetics of the Clupeiformes based on exon-capture data and a new classification of the order, *Molecular Phylogenetics and Evolution*, 2022, <https://doi.org/10.1016/j.ympev.2022.107590>.
- Wang, Y., Yuan, H., Huang, J. & Li, C. 2022b. Inline index helped in cleaning up data contamination generated during library preparation and the subsequent steps. *Molecular Biology Reports*, 1-8.
- Wang Y., Sun S., He Jiaxin, et al. Effects of elovl8 knockout on survival rate and lipid metabolism under cold stress in zebrafish [J]. *Journal of Hydrobiology*, 2022, 46(03): 303-315.
- Wang Y., Yang C., Zhao Y., et al. Construction of Transgenic Zebrafish Expressing Long-chain Fatty Acid Elongase 5 (Elov15) Gene from Carp Using CRISPR/Cas9 Technology [J]. *Journal of Agricultural Biotechnology*, 2020, 28(07): 1165-1176.
- Wu C., Xu Y., Liu Z., et al. Population Genetic Structure of *Ilisha elongata* in Chinese Coastal Waters Based on D-loop Gene [J]. *Oceanologia et Limnologia Sinica*, 2009, 40(03): 330-337.
- Xie Y., Huang L., Li J., et al. Evaluation of Clupeiform Fish Resources in the Coastal Waters of Fujian [J]. *Marine Fisheries*, 2012, 34(03): 285-294. DOI: 10.13233/j.cnki.mar.fish.2012.03.003.
- Xue, D. X., Yang, Q. L., Li, Y. L., Zong, S. B., Gao, T. X., & Liu, J. X. (2019). Comprehensive assessment of population genetic structure of the overexploited Japanese grenadier anchovy (*Coilia nasus*): Implications for fisheries management and conservation. *Fisheries Research*, 213, 113–120.
- Xu H. et al., Are fish what they eat? A fatty acid's perspective, *Progress in Lipid Research* (2020), <https://doi.org/10.1016/j.plipres.2020.101064>
- Yang, Z. (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer Applications in the Biosciences* Cabios, 13, 555.
- Yuan, H., Atta, C., Tornabene, L. & Li, C. 2019. Assexon: assembling exon using gene capture data. *Evolutionary Bioinformatics*, 15, 1176934319874792.

Zhang, J. & Takita, T. 2007. Age and growth of *Ilisha elongata* (Teleostei: Pristigasteridae) in Ariake Sound, Japan: comparison among populations in western North Pacific Ocean. Fisheries Science, 73, 971-978.

Zhao Y. Functional Study of Elovl4 in Zebrafish for Highly Unsaturated Fatty Acid Synthesis [D]. Huazhong Agricultural University, 2020. DOI: 10.27158/d.cnki.ghznu.2020.001143.

Accepted Manuscript