



OPEN

DATA DESCRIPTOR

High-quality Japanese flounder genome aids in identifying stress-related genes using gene coexpression network

Xi-wen Xu^{1,2,5}, Weiwei Zheng^{1,3,5}, Yingming Yang^{1,2,5}, Jilun Hou⁴ & Songlin Chen^{1,2,3}✉

The Japanese flounder is one of the most economically important marine flatfish. However, due to the increased frequency of extreme weather events and high-density industrial farming, an increasing number of environmental stresses have become severe threats to the healthy development of the Japanese flounder culture industry. Herein, we produced a high-quality chromosome-scale Japanese flounder genome using PacBio Circular Consensus Sequencing technologies. The assembled Japanese flounder genome spanned 588.22 Mb with a contig N50 size of 24.35 Mb. In total, 105.89 Mb of repetitive sequences and 22,565 protein-coding genes were identified by genome annotation. In addition, 67 candidate genes responding to distinct stresses were identified by gene coexpression network analysis based on 16 published stress-related RNA-seq datasets encompassing 198 samples. A high-quality chromosome-scale Japanese flounder genome and candidate stress-related gene set will not only serve as key resources for genomics studies and further research on the underlying stress responsive molecular mechanisms in Japanese flounder but will also advance the progress of genetic improvement and comprehensive stress-resistant molecular breeding of Japanese flounder.

Background & Summary

Japanese flounder (*Paralichthys olivaceus*, FishBase ID: 1351), naturally distributed in the Western Pacific: Korean Peninsula, Kuril Islands, Japan, Sakhalin to the South China Sea, is now an economically important cosmopolitan marine-culture flatfish species due to its delicious taste and high nutritional value¹. In recent years, however, Japanese flounder have suffered from an increasing number of environmental stresses, such as the increased temperature of global water², a wide range of environmental pollutants derived from agricultural and industrial wastes³⁻⁵, and the outbreak of fish diseases caused by high-density industrial farming⁶, which have severely threatened the healthy development of the Japanese flounder industry and led to enormous economic losses. Hence, acquiring a high-quality chromosome-scale assembled genome of Japanese flounder and obtaining abundant gene resources associated with stress resistance are necessary, which will aid in research on the molecular mechanism of stress resistance in Japanese flounder and provide theoretical support for the subsequent genetic improvement of Japanese flounder germplasm.

To the best of our knowledge, a high-quality chromosome-scale assembled genome is the basis of genetic resource mining and is critical for efficient molecular breeding. To date, two genome assemblies of Japanese flounder including one from our own group have been published in the National Center for Biotechnology Information (NCBI)^{7,8}. However, because of the restriction of sequencing techniques and assembly algorithms, the genome assemblies remain fragmented. Therefore, it is urgent to obtain an improved high-quality Japanese

¹Laboratory for Marine Fisheries Science and Food Production Processes, Qingdao National Laboratory for Marine Science and Technology, Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences, Nanjing Road 106, Qingdao, 266071, China. ²Key Lab of Sustainable Development of Marine Fisheries, Ministry of Agriculture and Rural Affairs, Wenhai Road 1, Qingdao, 266071, China. ³College of Fisheries and Life Science, Shanghai Ocean University, Shanghai, China. ⁴Beidaihe Central Experiment Station, Chinese Academy of Fishery Sciences, Qinhuangdao, Hebei, 066100, China. ⁵These authors contributed equally: Xi-wen Xu, Weiwei Zheng, Yingming Yang. ✉e-mail: chensl@ysfri.ac.cn

flounder genome. The long-read and high-accuracy PacBio single molecule sequencing technique (SMRT) and the emergence of corresponding advanced assembly methods make it possible.

RNA-sequencing (RNA-seq) technology has now become an effective tool for obtaining large-scale genetic information and functional genes, which has promoted the exploration of molecular mechanisms of resistance. Recently, many RNA-seq studies of Japanese flounder have made significant progress in detecting stress responsive genes and molecular mechanisms under various environmental stresses, such as chemical stress^{3,5,9,10}, heat stress^{11,12}, and pathogen stress^{6,13–18}. Although many key pathways and genes have been identified during environmental stresses through conventional transcriptome profiling, little attention has been given to connectivity analysis between genes by complex gene coexpression networks (GCNs). In this study, we reanalyzed multiple RNA-seq datasets from Japanese flounder using GCN analysis, which can explore the system-level functionality of the transcriptome¹⁹. GCN analysis can reveal the interrelation between gene modules and traits and identify key modules and the most central and vital hub genes, which provides an efficacious method to explore the in-depth mechanism of complex traits²⁰. GCN analyses have been used extensively to identify key gene modules and genes in various teleosts, such as *Scophthalmus maximus*^{21,22} and *Nibea albiflora*²³.

In the present study, we assembled a high-quality chromosome-scale Japanese flounder genome, of which the contig N50 size reached 24.35 Mb, using PacBio High fidelity (HiFi) reads. Based on this high-quality genome assembly, we further improved the annotation of the repetitive sequences and protein-coding genes in Japanese flounder. Moreover, we reanalyzed all published stress-related RNA-seq datasets from Japanese flounder from the NCBI Sequence Read Archive (SRA) database using GCN analysis, and identified multiple gene modules and candidate genes that respond to various environmental stresses in Japanese flounder. In summary, our study produced abundant and valuable genetic resources at the genome and gene levels, which can be used for further research on the functional genome and stress response mechanisms of Japanese flounder and can also provide theoretical support for the development of molecular breeding technology for resistant Japanese flounder varieties.

Methods

Japanese flounder samples and genome sequencing. Genomic DNA was isolated from fresh muscle samples of a female Japanese flounder. The degree of DNA degradation and RNA protein contamination were analyzed using 1% agarose gel electrophoresis, and the concentration was accurately quantified using Qubit.

Then, qualified genomic DNA was sheared into fragments, and the size-selection of fragments was performed using BluePippin. After the two ends of size-selected fragments were repaired, a poly-A tail and ligated adaptors were added, and the DNA library was constructed following the PacBio manufacturing protocols. Next, the library was sequenced using the PacBio Sequel II platform in circular consensus sequencing (CCS) mode following the manufacturer's instructions. As a result, we obtained a total of 49.17 Gb (~83X) PacBio HiFi reads. The average and N50 lengths of the HiFi reads were 14.00 Kb and 14.23 Kb, respectively.

Genome assembly and annotation. To obtain high-quality Japanese flounder genome sequences, 49.17 Gb PacBio HiFi reads were generated and assembled by Hifiasm (v0.15.1)²⁴ with default parameters. After removing redundant sequences using purge_dups (v1.2.5)²⁵, a total of 588.22 Mb genome sequences were obtained, and the length of contig N50 was 24.35 Mb, including 46 contigs. Then, RagTag (v2.1.0)²⁶ with default parameters was used to scaffold 46 contigs into 25 scaffolds, and therein, 24 chromosome-length scaffolds (Fig. 1) covered 99.99% of the assembled genome sequences, demonstrating a much higher quality genome assembly compared to other published Japanese flounder genomes^{7,8} (Table 1). For genome annotation, we first used EDTA (v2.00)²⁷ (--sensitive 1, --anno 1) to annotate genome transposable elements. Approximately 105.88 Mb of repetitive sequences were identified, as shown in Fig. 1, which is much higher than that in previously reported Japanese flounder genome (Table 2). Then, we used Liftoff (v1.6.1)²⁸ to map the gene structure annotation of *P. olivaceus* from Refseq (GCF_001970005.1) onto our newly assembled genome. Liftoff used minimap2 to align the gene sequences from Refseq annotation of *P. olivaceus* to the newly assembled genome. On the premise of ensuring transcript and gene structure, Liftoff found the optimal alignment of exons. Then Liftoff integrated information to map existing annotations to the newly assembled genome and generated annotation file. Finally, a total of 22,565 protein-coding genes were identified. The distributions of these genes on each chromosome are shown in Fig. 1.

Gene coexpression network construction and module-trait association analysis. To identify key genes and modules in response to various stresses by GCN analysis in Japanese flounder, we collected all of the published stress-related RNA-seq datasets, including 16 independent experiments from 3 distinct stresses (i.e., chemical, heat and pathogen stress) encompassing a total of 198 samples (Table 3), until March 31, 2022 from the NCBI SRA database. RNA-seq datasets downloaded in SRA format were first converted into FASTQ format utilizing the fastq-dump tool in SRAToolkit (v3.0.0)²⁹. Then, all sequenced reads were mapped to the high-quality Japanese flounder genome using STAR (v2.7.10a_alpha_220314)³⁰ software with default parameters. Next, gene expression data for all genes normalized in transcripts per million (TPM) were calculated by TPMCalculator (v0.0.2) (-q 1)³¹ with sorted bam files obtained from read alignment. Next, the gene expression data were pre-processed using BioNERO (v1.4.2)³² based on the following steps: I) displacing missing values (NAs) with 0 utilizing the replace_na function; II) eliminating the genes with average gene expression less than 1 using the remove_nonexp function; III) removing outlying samples using the ZKfiltering function; and IV) adjusting for confounding artifacts using the PC_correction function to make every gene follow an approximate normal distribution. After step-by-step data preprocessing, we obtained a normalized gene expression matrix comprising 13,138 genes whose medial expression values were ≥ 1 from 184 RNA sequencing samples.

Next, the gene coexpression network (GCN) shown in Fig. 2 was constructed using BioNERO with normalized gene expression matrix as input. To make the GCN satisfy the scale-free topology, of which the scale-free

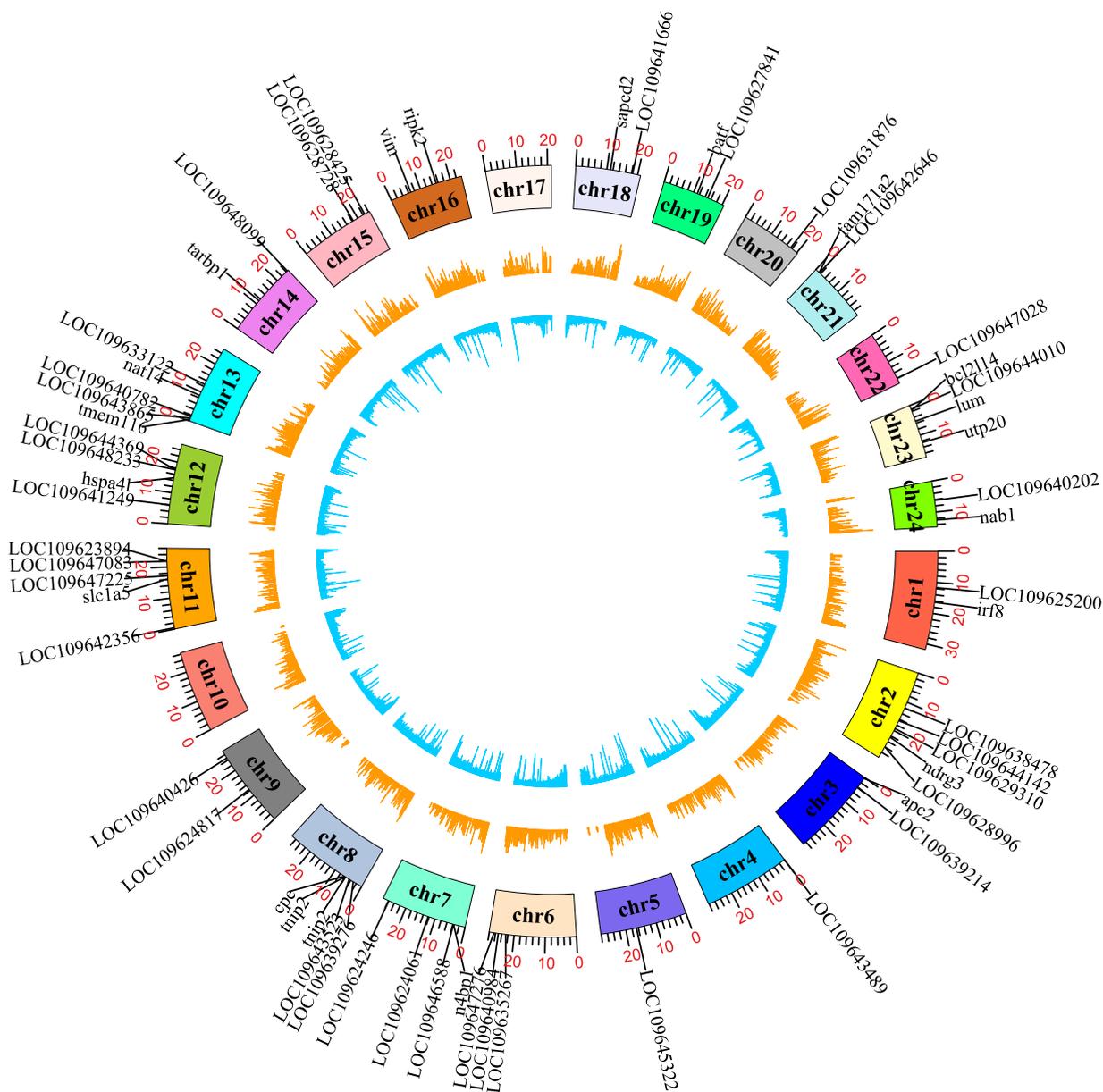


Fig. 1 A circos plot of 24 chromosomes of *P. olivaceus* genome. The tracks from inside to outside are: the distributions of transposable element, bar plot for gene density profile, 24 chromosomes, tag labels for candidate key hub genes related to various stresses.

Genome assembly	This study	Shao <i>et al.</i> ⁸	Wei <i>et al.</i> ⁷
Scaffold N50 (Mb)	26.05	23.21	10.55
Contig N50 (Mb)	24.35	0.031	0.036
Total scaffold number	25	6,969	9,525
Total contig number	46	38,614	32,584
Total length (Mb)	588.22	545.78	643.91
GC Content (%)	41.54	41.44	41.64

Table 1. Comparative statistic of the *P. olivaceus* genome assembly with old ones.

topology fit index (R^2) reaches 0.8 and the mean connectivity tends to 0, we identified the best β power of 9 with the function of SFT_fit. Then, the GCN was inferred using the exp2gcn function with power 9. Eventually, as shown in Fig. 2, a total of 25 coexpression modules were identified, of which the number of genes per module varied from 31 (pink) to 5,947 (tan).

Class		This study		Shao <i>et al.</i> ⁸		Wei <i>et al.</i> ⁷	
		Length (bp)	% in Genome	Length (bp)	% in Genome	Length (bp)	% in Genome
LTR	Copia	112,330	0.02%	—	—	—	—
	Gypsy	3,786,094	0.64%	975,109	0.18%	237,541	0.05%
	unknown	17,000,615	2.89%	4,362,453	0.82%	3,202,793	0.61%
TIR	CACTA	23,074,608	3.92%	21,956,463	4.11%	22,514,061	4.32%
	Mutator	10,131,242	1.72%	6,863,910	1.28%	5,790,854	1.11%
	PIF_Harbinger	2,410,258	0.41%	1,888,730	0.35%	1,688,125	0.32%
	Tc1_Mariner	845,440	0.14%	930,093	0.17%	427,982	0.08%
	hAT	4,617,190	0.78%	2,578,715	0.48%	2,368,980	0.45%
	polinton	74,664	0.01%	—	—	14,991	0.00%
nonLTR	DIRS_YR	12,352	0.00%	—	—	—	—
	LINE_element	2,852,254	0.48%	787,838	0.15%	718,077	0.14%
nonTIR	helitron	7,640,001	1.30%	4,848,794	0.91%	4,868,953	0.93%
	repeat_region	33,330,298	5.67%	8,190,805	1.53%	8,657,096	1.66%
Total		105,887,346	18.00%	53,382,910	9.99%	50,489,453	9.68%

Table 2. Comparative statistic of the *P. olivaceus* repeat sequences with old ones.

To identify modules in response to various stresses, the module-trait Spearman correlation coefficients were calculated with the function `module_trait_cor` in the BioNERO software package. Modules extremely significantly (p value < 0.01) positively or negatively correlated with particular traits (stresses) were found, among which 2, 5 and 5 modules were extremely significantly related to chemical, heat, and pathogen stress, respectively. The results are shown in Fig. 3. Interestingly, among these extremely significant modules, one module, the cyan module, was extremely significantly associated with all three distinct stresses. All of these results provide abundant resources for further study of stress-responsive molecular mechanisms in Japanese flounder.

Module function enrichment analysis. To detect biological processes or pathways to which the genes in the modules that were extremely significantly associated with each stress were related, GO and KEGG function enrichment analyses were conducted using TBtools (v1.09852)³³. Significantly enriched GO terms and KEGG pathways (corrected p value (BH method) < 0.5) were identified based on a hypergeometric cumulative distribution function test.

For chemical stress, GO function enrichment results demonstrated that cellular response to chemical stress, cellular response to chemical stimulus, and response to toxic substance were the significantly enriched terms in modules that were extremely significantly related to chemical stress. In addition, KEGG function enrichment analyses showed that chemical carcinogenesis-receptor activation, various signaling pathways related to environmental information processing (such as JAK-STAT, MAPK, and NF-kappa), and signal transduction were significantly enriched pathways in the same modules.

For heat stress, GO function enrichment results illustrated that cellular response to heat, heat shock protein binding, response to heat, lipid catabolic process and catalytic activity were the significantly enriched terms in modules that were extremely significantly correlated with heat stress. Furthermore, KEGG enrichment analyses were conducted on the same modules, and the results showed that cortisol synthesis and secretion and multiple signaling pathways related to environmental information processing (such as NOD-like receptor, JAK-STAT, NF-kappa B, and MAPK) were significantly enriched pathways.

For pathogen stress, GO function enrichment results illustrated that regulation of response to biotic stimulus, immune response, defense response, regulation of immune system process, T-cell activation and alpha-beta T-cell differentiation were the significantly enriched terms in modules that were extremely significantly associated with pathogen stress. Furthermore, KEGG functional enrichment analyses indicated that the immune system, antigen processing and presentation, a variety of signaling pathways (TNF, Toll-like receptor, NOD-like receptor, NF-kappa B, B-cell receptor, T-cell receptor, chemokine, MAPK, etc.), CD molecules and phagosomes were significantly enriched pathways.

It is worth emphasizing that the cyan module, which was extremely significantly associated with all three kinds of stresses, was also significantly enriched in stress-related GO terms, such as stress-activated MAPK cascade, regulation of stress-activated MAPK cascade, regulation of response to stress, p38 MAPK cascade, response to external stimulus, ERK1/2 cascade, and regulation of JNK cascade. Meanwhile, KEGG enrichment results indicated that the cyan module was mainly enriched in environmental information processing-related pathways, such as the MAPK/AMPK/p53 signaling pathway and signal transduction. These results suggested that the cyan module was an important module responsive to different stresses and is worth further study.

Identification of key hub genes. In this study, candidate key hub genes were defined as the intersections of the hub gene set and differentially expressed gene (DEG) set related to the same stress. First, we established the hub gene set. According to BioNERO's algorithm, hub genes were selected as the top 10% of genes with the highest k_{Within} (i.e., the degree of connectivity of the edge of one gene to all other genes under the same module) whose module membership (MM) (i.e., correlation of a gene to its module eigengene) > 0.8. The function `get_hubs_gcn`

Stress		SRA Study	Number of Experiments	Number of Individuals	Platform (Illumina)	Size (GB)	References
Chemicals	bisphenol S- and benzo[a]pyrene	SRP126397	8	22	HiSeq 3000	55.69	³
	nickel and cobalt	SRP238940	6	180	HiSeq 2000	48.42	⁵
	contaminated sediment	SRP313148	4	200	HiSeq 2500	91.86	⁹
	copper yrithione and zinc pyrithione	SRP342241	5	650	HiSeq 3000	29.22	¹⁰
Heat	—	SRP266623	12	36	HiSeq X Ten	249.20	¹¹
		SRP330366	12	30	NovaSeq 6000	94.97	¹²
		SRP317999					
		SRP330365					
SRP324580							
Pathogens	<i>Edwardsiella tarda</i>	SRP109689	10	130	HiSeq 4000	143.49	¹⁵
		PRJNA796380 Un-published	6	—	NovaSeq 6000	—	⁴³
		SRR8293419	6	18	HiSeq 4000	125.8	¹⁸
		SRR8290049					
		SRR8284024					
		SRR8271825					
		SRR8270369					
	SRR8269382						
	<i>Hirame novirhabdovirus</i>	SRP162413	2	—	HiSeq X Ten	14.58	—
		SRP162514	4	320	HiSeq X Ten	27.96	¹⁶
		SRP299209	18	18	HiSeq 2500	117.18	⁴⁴
	<i>Vibrio anguillarum</i>	SRP214405	18	18	HiSeq 4000	202.72	⁶
	<i>Megalocytivirus</i>	SRP297629	18	72	HiSeq 4000	264.17	¹⁷
<i>Viral hemorrhagic septicemia virus</i>	SRP102673	36	36	HiSeq 2500	202.76	⁴⁵	
	SRP105302	36	40	HiSeq 2500	202.58	¹⁴	
Total	—	—	168	—	—	1870.6	

Table 3. Overview of the RNA-seq datasets used in this study.

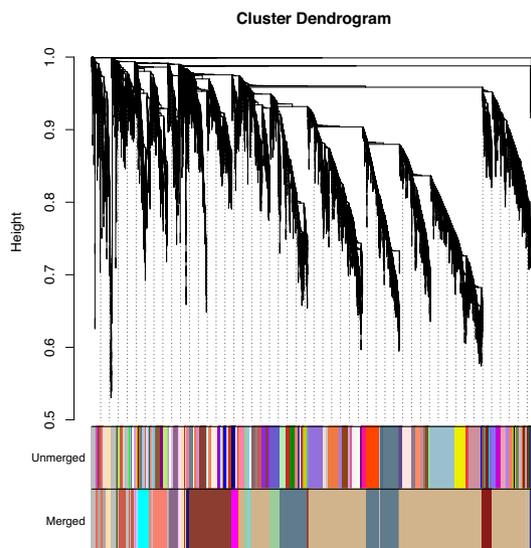


Fig. 2 Cluster dendrogram of genes and modules. The branches and color bands represent the assigned module. The tips of the branches represent genes.

was used to identify hub genes in each module. Next, hub genes contained in modules extremely significantly associated with the same stress were selected as hub gene set for this stress. Second, the DEG set was established. The featureCounts (-p)³⁴ software program in the Subread (v2.0.3)³⁵ package was first used to build read count matrixes. Next, edgeR (v3.14)³⁶ software was used to identify DEGs with the criteria of false discovery rate (FDR) <0.05 and $|\log_2(\text{fold change})| > 2$. DEGs associated with the same stress were merged into the DEG set for

Module–trait relationships

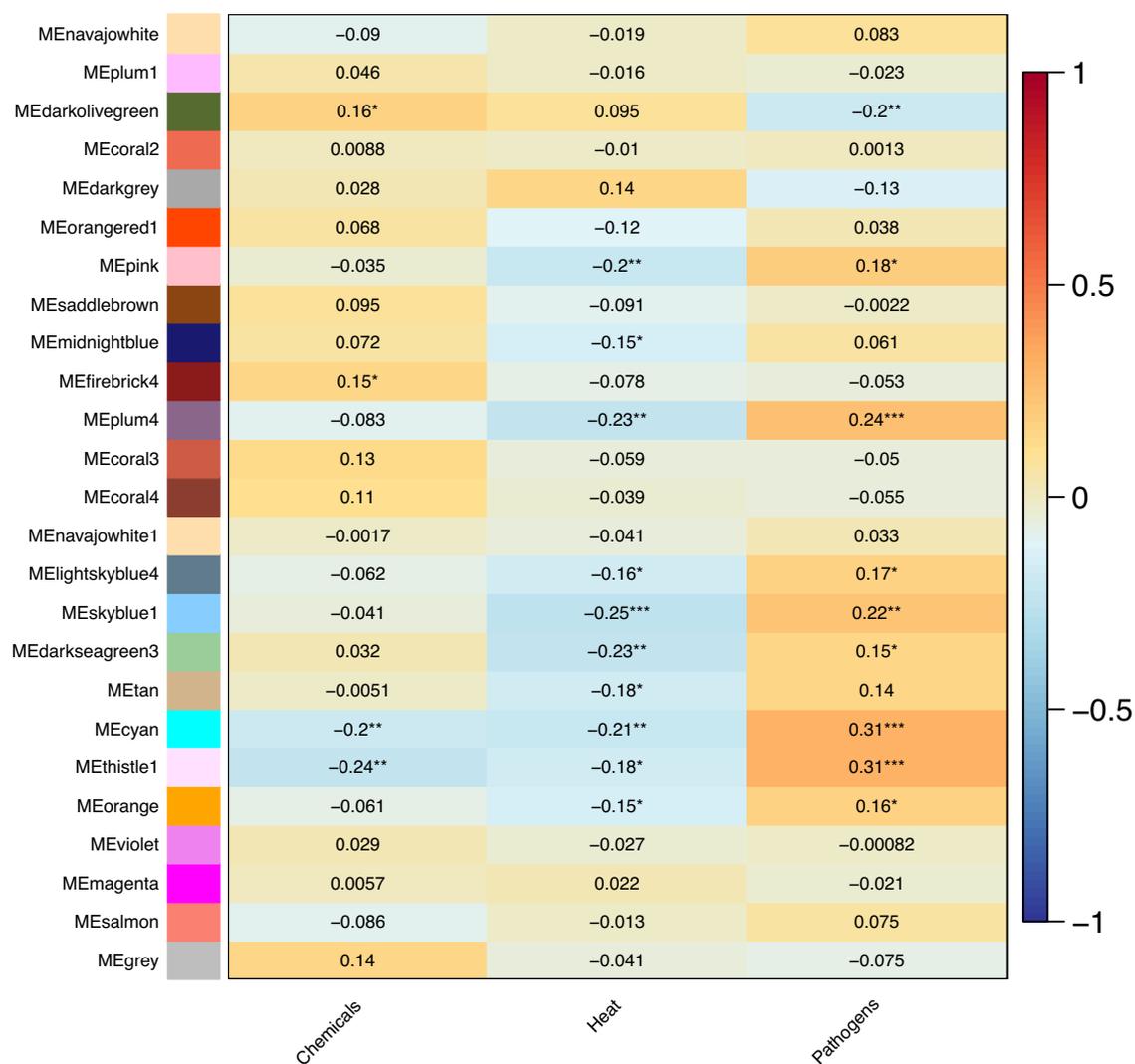


Fig. 3 Correlation between modules and stresses. The value in the box is the correlation coefficients. Correlation coefficients with ** or *** represent extremely significant correlation and significant correlation with *.

this stress. Finally, 5, 4 and 58 candidate key hub genes associated with chemical, heat, and pathogen stress were identified, respectively. The positions of these key hub genes on chromosomes are shown in Fig. 1.

Data Records

The raw PacBio sequencing data of the Japanese flounder genome was deposited at the NCBI SRA database under SRA study accession number SRP378268³⁷. The final assembled Japanese flounder genome has been submitted to NCBI under accession number GCA_024713975.1³⁸. The final genome was also submitted to the National Genomics Data Center with accession number GWHBJWM00000000³⁹. In addition, the annotation file of the *P. olivaceous* genome assembly, the gene expression matrix used for GCN construction, and results from GCN analysis including genes per module, DEGs set, hub genes set, GO and KEGG function enrichment, and candidate key hub genes, have been deposited in Figshare⁴⁰.

Technical Validation

Completeness and quality assessment of the Japanese flounder genome. Two independent methods were used to assess genome assembly completeness and quality. We first employed Benchmarking Universal Single-Copy Orthologs (BUSCO v5.3.2)⁴¹ with the actinopterygii_odb10 database, including 3,640 BUSCOs, to evaluate the completeness of the final assembled genome. BUSCO analysis indicated that 98.5% complete BUSCOs were captured, including 97.6% single-copy and 0.9% duplicated BUSCOs. Then, we used Inspector (v1.0.1)⁴² to estimate the assembly quality by mapping PacBio HiFi reads to the assembled contigs to generate read-to-contig alignment and conduct downstream assembly evaluation. The results revealed that the

quality value (QV) and read-to-contig mapping rate were 44.31 and 99.91%, respectively. The above evaluation results demonstrated the high quality and completeness of the reported genome assembly, which will contribute to further research on the genomics and genetics of Japanese flounder.

Code availability

The data analysis methods, software and associated parameters used in present study are described in the section of Methods. If no detail parameters were described for software used in this study, default parameters were employed. No custom scripts were generated in this work.

Received: 26 August 2022; Accepted: 21 October 2022;

References

- Fuji, K. *et al.* Identification of a single major genetic locus controlling the resistance to lymphocystis disease in Japanese flounder (*Paralichthys olivaceus*). *Aquaculture* **254**, 203–210 (2006).
- Kim, W. J. *et al.* Transcriptome profiling of olive flounder responses under acute and chronic heat stress. *Genes Genomics* **43**, 151–159 (2021).
- Jung, J. H. *et al.* Comparative analysis of distinctive transcriptome profiles with biochemical evidence in bisphenol S- and benzo[a]pyrene-exposed liver tissues of the olive flounder *Paralichthys olivaceus*. *Plos One* **13**, e0196425 (2018).
- Anbuselvan, N., Nathan, S. D. & Sridharan, M. J. M. p. b. Heavy metal assessment in surface sediments off Coromandel Coast of India: Implication on marine pollution. *Marine Pollution Bulletin* **131**, 712–726 (2018).
- Sun, Z. *et al.* Toxicity of nickel and cobalt in Japanese flounder. *Environ Pollut* **263**, 114516 (2020).
- Ning, X. & Sun, L. Gene network analysis reveals a core set of genes involved in the immune response of Japanese flounder (*Paralichthys olivaceus*) against *Vibrio anguillarum* infection. *Fish Shellfish Immunol* **98**, 800–809 (2020).
- Xiaolin, W. *et al.* pBACode: a random-barcode-based high-throughput approach for BAC paired-end sequencing and physical clone mapping. *Nucleic Acids Research* **45**, 7 (2017).
- Shao, C. *et al.* The genome and transcriptome of Japanese flounder provide insights into flatfish asymmetry. *Nat Genet* **49**, 119–124 (2017).
- Choi, Y. *et al.* Sediment quality assessment combining chemical and biological (non)target analysis. *Aquat Toxicol* **238**, 105883 (2021).
- Shin, D. *et al.* Comparative toxicity study of waterborne two booster biocides (CuPT and ZnPT) on embryonic flounder (*Paralichthys olivaceus*). *Ecotoxicol Environ Saf* **233**, 113337 (2022).
- Yuan, M. *et al.* Transcriptomics reveals that the caudal neurosecretory system in the olive flounder (*Paralichthys olivaceus*) is more responsive in bold individuals and to chronic temperature change. *Aquaculture* **544**, 737032 (2021).
- Qiao, Y. *et al.* Identification, evolution and expression analyses of mapk gene family in Japanese flounder (*Paralichthys olivaceus*) provide insight into its divergent functions on biotic and abiotic stresses response. *Aquat Toxicol* **241**, 106005 (2021).
- Jeong, J. M., Jeswin, J., Bae, J. S. & Park, C. I. Differential expression of olive flounder (*Paralichthys olivaceus*) transcriptome during viral hemorrhagic septicemia virus (VHSV) infection at warmer and colder temperature. *Data Brief* **19**, 2023–2028 (2018).
- Hwang, J. Y. *et al.* Transcriptome analysis of olive flounder (*Paralichthys olivaceus*) head kidney infected with moderate and high virulent strains of infectious viral hemorrhagic septicemia virus (VHSV). *Fish Shellfish Immunol* **76**, 293–304 (2018).
- Li, Z. *et al.* Transcriptome profiling based on protein-protein interaction networks provides a core set of genes for understanding blood immune response mechanisms against *Edwardsiella tarda* infection in Japanese flounder (*Paralichthys olivaceus*). *Dev Comp Immunol* **78**, 100–113 (2018).
- Wang, H. *et al.* Transcriptome analysis reveals temperature-dependent early immune response in flounder (*Paralichthys olivaceus*) after *Hirame novirhabdovirus* (HIRRV) infection. *Fish Shellfish Immunol* **107**, 367–378 (2020).
- Wu, Q., Ning, X., Jiang, S. & Sun, L. Transcriptome analysis reveals seven key immune pathways of Japanese flounder (*Paralichthys olivaceus*) involved in megalocytivirus infection. *Fish Shellfish Immunol* **103**, 150–158 (2020).
- Li, X. P. & Zhang, J. A live attenuated *Edwardsiella tarda* vaccine induces immunological expression pattern in Japanese flounder (*Paralichthys olivaceus*) in the early phase of immunization. *Comp Biochem Physiol C Toxicol Pharmacol* **239**, 108872 (2021).
- Lu, C. *et al.* Comparative transcriptomics and weighted gene co-expression correlation network analysis (WGCNA) reveal potential regulation mechanism of carotenoid accumulation in *Chrysanthemum x morifolium*. *Plant Physiol Biochem* **142**, 415–428 (2019).
- Ly, L. *et al.* Gene co-expression network analysis to identify critical modules and candidate genes of drought-resistance in wheat. *Plos One* **15**, e0236186 (2020).
- Huang, Z. *et al.* Transcriptome analysis and weighted gene co-expression network reveals potential genes responses to heat stress in turbot *Scophthalmus maximus*. *Comp Biochem Physiol Part D Genomics Proteomics* **33**, 100632 (2020).
- Xu, X.-w *et al.* Identification of stress-related genes by co-expression network analysis based on the improved turbot genome. *Scientific Data* **9**, 374 (2022).
- Zhao, X., Sun, Z., Xu, H., Song, N. & Gao, T. Transcriptome and co-expression network analyses reveal the regulatory pathways and key genes associated with temperature adaptability in the yellow drum (*Nibea albiflora*). *J Therm Biol* **100**, 103071 (2021).
- Cheng, H., Concepcion, G. T., Feng, X., Zhang, H. & Li, H. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature Methods* **18**, 170–175 (2021).
- Guan, D. *et al.* Identifying and removing haplotypic duplication in primary genome assemblies. *Bioinformatics* **36**, 2896–2898 (2020).
- Alonge, M. *et al.* RaGOO: fast and accurate reference-guided scaffolding of draft genomes. *Genome Biology* **20**, 224 (2019).
- Ou, S. *et al.* Benchmarking transposable element annotation methods for creation of a streamlined, comprehensive pipeline. *Genome Biology* **20**, 275 (2019).
- Shumate, A. & Salzberg, S. L. LiftOff: accurate mapping of gene annotations. *Bioinformatics* **37**, 1639–1643 (2021).
- Team, S. T. D. SRAToolkit version 2.11.0. (2021).
- Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).
- Vera Alvarez, R., Pongor, L. S., Marino-Ramirez, L. & Landsman, D. TPMCalculator: one-step software to quantify mRNA abundance of genomic features. *Bioinformatics* **35**, 1960–1962 (2019).
- Almeida-Silva, F., Venancio, T. M. J. F. & Genomics, I. BioNERO: an all-in-one R/Bioconductor package for comprehensive and easy biological network reconstruction. *Functional & Integrative Genomics* **22**, 131–136 (2022).
- Chen, C. *et al.* TBtools: an integrative toolkit developed for interactive analyses of big biological data. *Molecular Plant* **13**, 1194–1202 (2020).
- Liao, Y., Smyth, G. K. & Shi, W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

35. Liao, Y., Smyth, G. K. & Shi, W. The Subread aligner: fast, accurate and scalable read mapping by seed-and-vote. *Nucleic Acids Research* **41**, e108 (2013).
36. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
37. NCBI Sequence Read Archive <https://identifiers.org/insdc.sra:SRP378268> (2022).
38. NCBI Assembly, https://identifiers.org/ncbi/insdc.gca:GCA_024713975.2 (2022).
39. National Genomics Data Center <https://ngdc.cncb.ac.cn/search/?dbId=gwh&q=GWHBJWM00000000> (2022).
40. Xu, X. & Zheng, W. A high-quality chromosome-scale genome assembly of Japanese flounder (*Paralichthys olivaceus*). *figshare* <https://doi.org/10.6084/m9.figshare.20331282.v2> (2022).
41. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular Biology and Evolution* **38**, 4647–4654 (2021).
42. Chen, Y., Zhang, Y., Wang, A. Y., Gao, M. & Chong, Z. Accurate long-read de novo assembly evaluation with Inspector. *Genome Biol* **22**, 312 (2021).
43. Wang, L. *et al.* Transcriptome analysis and protein-protein interaction in resistant and susceptible families of Japanese flounder (*Paralichthys olivaceus*) to understand the mechanism against *Edwardsiella tarda*. *Fish & Shellfish Immunology* **123**, 265–281 (2022).
44. Tang, X. *et al.* The influence of temperature on the antiviral response of mIgM(+) B Lymphocytes against HIRAME Novirhabdovirus in flounder (*Paralichthys olivaceus*). *Frontiers in Immunology* **13**, 802638 (2022).
45. Hwang, J. Y. *et al.* Temperature-dependent immune response of olive flounder (*Paralichthys olivaceus*) infected with viral hemorrhagic septicemia virus (VHSV). *Genes & Genomics* **40**, 315–320 (2018).

Acknowledgements

This work was supported by the Special Scientific Research Funds for Central Non-profit Institutes, Yellow Sea Fisheries Research Institute (20603022021001), the Key Research and Development Project of Shandong Province (2021LZGC028), Central Public-interest Scientific Institution Basal Research Fund, CAFS (2020TD20), and Shandong Taishan Scholar Climbing Project.

Author contributions

S.C. and X.X. applied, designed and supervised the project. X.X., W.Z. analyzed the data. X.X., Y.Y. and J.H. prepared the samples for whole genome sequencing and conducted the experiments. X.X., W.Z., S.C., Y.Y. and J.H. wrote and revised the manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to S.C.

Reprints and permissions information is available at www.nature.com/reprints.

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



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022