PREPRINT

PERTANIKA JOURNALS

TROPICAL AGRICULTURAL SCIENCE

Journal homepage: http://www.pertanika.upm.edu.my/

Ubiquitin Gene Diversity in Different Populations of Javanese Medaka (*Oryzias javanicus*) from Peninsular Malaysia

Chu Mengxuan[1], Nurul Izza Ab Ghani[1,2,3]*, Rosimah Nulit[1], and Ahmad Ismail[3,4]

[1]Department of Biology, Faculty of Science, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia

[2]Laboratory of Sustainable Animal Production and Biodiversity, Institute of Tropical Agriculture and Food Security (ITAFoS), Universiti Putra Malaysia (UPM), 43400 Serdang, Selangor, Malaysia

[3]Persatuan Ekologi Malaysia, Universiti Putra Malaysia (UPM), 43400 Serdang, Selangor, Malaysia

[4]Academy of Sciences Malaysia, Level 20, West Wing, MATRADE Tower, Jalan Sultan Haji Ahmad Shah, off Jalan Tuanku Abdul Halim, 50480 Kuala Lumpur, Malaysia

ABSTRACT

This research examines ubiquitin (*UB*) gene diversity in eight *Oryzias javanicus* populations from Peninsular Malaysia, including industrial, island, and coastal habitats. PCR amplification of genomic DNA from 160 individuals yielded a single band of approximately 300 bp, revealing complete monomorphism across all 121 individuals. Bidirectional Sanger sequencing of 25 individuals revealed structural heterogeneity, while core coding regions remained consistent. The fragment belongs to a ubiquitin-L40 fusion gene and contains a 110-bp intron. Main conclusions were: (1) Coding regions were almost invariant, whereas the intron was polymorphic ($\pi = 0.022$, about 20 times that of the coding regions); (2) A divergent allele was found in one individual from Tanjung Piai (TP7); (3) No population-specific structural variations: the Neighbour-Joining tree produced no node with a bootstrap support of ≥50% (three bootstrap replicates were tested; 1000, 5000 and 10,000) and individuals did not group by population or region; and (4) Exceptional *UB* conservation within *Oryzias* but substantial diversification in other teleosts. Results showed steady purifying selection over a long period rather than a recent positive selective sweep, which would have wiped out variation in the linked intron via genetic hitchhiking on the *UB* core. Flanking structural variation may contribute to adaptation in heterogeneous estuaries. *UB* is therefore strongly conserved, and the intron variation occurs within populations. Thus, multi-locus or genome-wide markers are needed to assess the population structure of *O. javanicus*.

Keywords: conserved gene, environmental pressures, genetic diversity, *Oryzias javanicus*, population genetics, ubiquitin gene

ARTICLE INFO

Article history:
Received: 09 July 2026
Accepted: 22 November 2026
Published: 02 December 2026

DOI: https://doi.org/10.47836/pjtas.49.6.10

E-mail addresses:
gs71523@student.upm.edu.my (Chu Mengxuan)
nurul_izza@upm.edu.my (Nurul Izza Ab Ghani)
rosimahn@upm.edu.my (Rosimah Nulit)
ismailmt56@yahoo.com.my (Ahmad Ismail)
* Corresponding author

e-ISSN: 2231-8542

© Universiti Putra Malaysia Press

INTRODUCTION

The Javanese medaka (*Oryzias javanicus*) is a euryhaline teleost in the family Adrianichthyidae and order Beloniformes. This species inhabits mangrove swamps, estuaries, and marine environments in Southeast Asia, such as Peninsular Malaysia, Thailand, Indonesia, and Singapore (Bleeker, 1854; Magtoon & Termvidehakorn, 2009; Ngamniyom et al., 2022; Roberts, 1998). This species has shown high adaptability to changing salinity levels (2.0 to 30.0 ppt) and dissolved oxygen levels (2.7 to 8.2 mg/L), which can be explained by specific physiological and genetic adaptations (Inoue & Takei, 2003; Yusof et al., 2012a). The Javanese medaka's presence on the western and southern coasts of Peninsular Malaysia has important ecological implications, particularly in mangrove ecosystems, as it contributes to the food web and nutrient cycling, and acts as food for other aquatic organisms (Zulkifli et al., 2012). This species has also been recognised as a bioindicator for assessing mangrove ecosystem health in Southeast Asia (Takehana et al., 2020). This species has been shown to be highly susceptible to various types of pollutants, such as heavy metals (Woo et al., 2006), herbicides (Kamarudin et al., 2020), and microplastics (Usman et al., 2021), and has been recognised as a test fish for ecotoxicological studies (Takehana et al., 2020). For instance, measurable physiological and behavioural responses offer potential early warning signs of environmental degradation (Puspitasari & Lestari, 2018; Viarengo et al., 2007). Hence, *O. javanicus* is an important tool for guiding mangrove-associated biodiversity conservation, a critical concern given the various threats to mangrove ecosystems (Blanton et al., 2024; Gerona-Daga & Salmo, 2022). Recent advances in genomics, including a chromosome-level assembly of 809.7 Mb with an N50 of 6.3 Mb and annotation of 21,454 transcripts, have reinforced the significance of *O. javanicus* as a model species for evolutionary and ecological genomics in the *Oryzias* genus (Takehana et al., 2020). From a phylogenetic perspective, *O. javanicus* is associated with other euryhaline species such as *Oryzias melastigma* and forms a distinct clade from other species of freshwater fish, such as *Oryzias latipes*, thus emphasising its significance to research on genetic adaptations to saline environments (Hilgers & Schwarzer, 2019; Takehana et al., 2020).

Human activities that have resulted in industrial pollution, habitat fragmentation, and climate change (Halpern et al., 2019) pose major risks to the genetic integrity and long-term sustainability of *O. javanicus*. For example, industrial areas around Manjung (oil and gas industry) and Seberang Perai (a manufacturing hub) are known to harbour pollutants that can cause DNA damage and selection pressures (Amal et al., 2018; Yusof et al., 2013). Other areas of interest include major aquaculture sites (Bagan Lalang) and areas around power plants, which can fragment habitats and reduce genetic diversity (Idris, 2018). These environmental heterogeneities thus require detailed evaluations of genetic variation to guide conservation approaches. A population genetic approach is essential for

understanding gene flow, genetic drift, and adaptation, as well as population sustainability under environmental change. Genetic variation can be assessed using heterozygosity, allelic richness, and differentiation (F_{ST}), which are essential for understanding population genetic structure and evolutionary potential, especially in species exposed to environmental change and habitat fragmentation (Luikart et al., 2010; Rousset, 2008).

The ubiquitin gene (*UB*) encodes a highly conserved 76-amino-acid protein and is a key molecular marker because it is essential for maintaining eukaryotic protein homeostasis. *UB* plays a key role in regulating critical eukaryotic cell processes such as protein degradation via the ubiquitin-proteasome pathway, DNA repair, cellular stress response, and cell cycle progression through post-translational modification of target proteins (Braten et al., 2016; Hershko & Ciechanover, 1998). Its high sequence conservation across diverse taxonomic groups makes it well suited for comparative analyses and phylogenetic inferences, while its functional importance suggests it may have undergone adaptive variation (Hochstrasser, 1996; Marín, 2014). In fish, *UB* expression is sensitive to environmental stressors. In the liver of *O. javanicus*, *UB* expression increased dose-dependently after sublethal exposure to silver, cadmium, and copper (Woo et al., 2009). *UB* orthologues are also annotated in the congeners *O. melastigma* (XM_024279027.2) and *O. latipes* (XM_004067926.5). The polyubiquitin gene of rainbow trout (*Oncorhynchus mykiss*) is a *UB* gene similar to that of mammals and is expressed in response to food deprivation (Martin et al., 2002). The *UB* promoter of the zebrafish (*Danio rerio*) is expressed in all tissues and developmental stages (Mosimann et al., 2011). Although it is highly applicable for various analyses, genetic variation in *O. javanicus* natural populations remains unstudied, particularly in Peninsular Malaysia's heterogeneous environments. Although the ubiquitin protein is highly conserved, it is encoded by different gene types: polyubiquitin genes consist of tandem 228-bp coding units (Sharp & Li, 1987), whereas two fusion genes encode a single ubiquitin unit fused to a ribosomal protein, L40 or S27a (Finley et al., 1989). The gene family evolves under strong purifying selection (Nei et al., 2000), so variation among populations is expected mainly in non-coding regions such as introns, which can affect gene regulation (Pai et al., 2015). The role of such variation in adaptation of wild fish populations is largely unstudied.

This study aims to fill in the knowledge gap in terms of *UB* gene diversity in eight geographically diverse populations of *O. javanicus* in Peninsular Malaysia, ranging from Kuala Perlis in Perlis in the north to Tanjung Piai in Johor in the south, and including an island population at Sungai Kilim (Langkawi) and industrial populations in Manjung and Seberang Perai. Its objectives include: (i) screening *UB* length variation in eight populations by PCR; (ii) characterising the structure and nucleotide variation of the *UB* locus by sequencing; (iii) examining relationships among individuals, populations and across eight teleosts with Neighbour-Joining trees; and (iv) exploring whether *UB* variation corresponds to environmental and anthropogenic differences among sites. This study provides baseline

data on *UB* variation in *O. javanicus* for population genetic and ecotoxicological studies and for conserving mangrove biodiversity in Southeast Asia.

MATERIALS AND METHODS

Sampling Locations

Javanese medaka populations were collected from eight sites across Peninsular Malaysia. Sampling sites included Kuala Perlis in Perlis (the northern part of Peninsular Malaysia), Sungai Kilim in Langkawi (an island population), Sungai Merbok in Kedah (famous for its freshwater and wetland ecosystems), Kuala Sepetang in Perak (near Matang Mangrove Forest Park), Manjung in Perak (an area known for its oil and gas industries), Bagan Lalang in Selangor (an important area for aquaculture activities with a power plant nearby), Seberang Perai in Penang (an industrial area), and Tanjung Piai in Johor (the southern tip of Peninsular Malaysia and a RAMSAR site) (Figure 1). Mangrove trees (*Rhizophora* spp.) are present at all sampling sites. General water parameters of each locality are given in Supplementary Table 1.

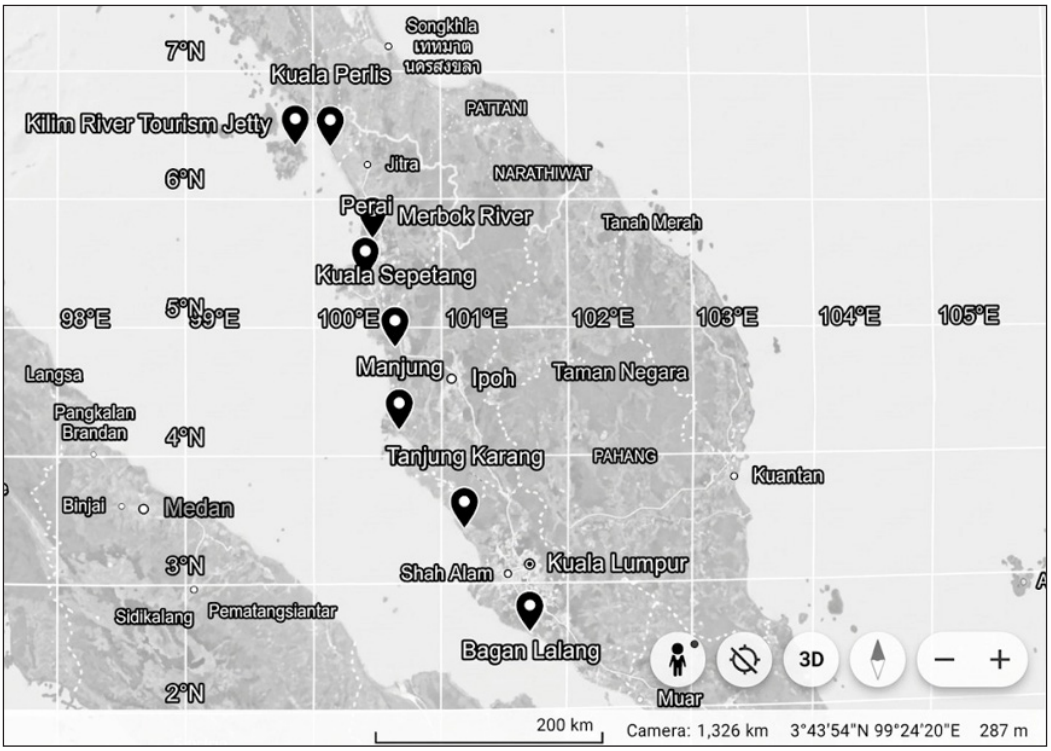


Figure 1. The location of sampling sites of eight populations (Kuala Perlis in Perlis, Sungai Kilim in Langkawi, Sungai Merbok in Kedah, Kuala Sepetang in Perak, Manjung in Perak, Bagan Lalang in Selangor, Seberang Perai in Penang, and Tanjung Piai in Johor) in Peninsular Malaysia (Google Maps)

To avoid damaging samples during collection, the hand-netting method was used to collect specimens. Each sample was stored in a 50-mL sampling bottle with 70% ethanol and assigned a unique identification number based on the sampling location and month of collection. The sampling process was conducted between 2018 and 2020 in cooperation with local fishermen.

The collection, handling, and treatment of the fish followed approval from the Institutional Animal Care and Use Committee, Universiti Putra Malaysia (AUP No.: R006/2016).

Data Collection

General water quality parameters (DO, dissolved oxygen; BOD, biochemical oxygen demand; COD, chemical oxygen demand; and EC, electrical conductivity) at the eight sampling localities were measured using YSI meters (YSI MultiLab and YSI 910 COD Colorimeter).

Genomic DNA was isolated from caudal fin tissues of individual Javanese medaka (*Oryzias javanicus*) using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany), with extractions performed for 20 individuals per population. The DNA quality and purity were checked using a DeNovix DS-11 Spectrophotometer and agarose gel electrophoresis to measure DNA concentration, detect contaminants, assess degradation, and determine the optimal DNA concentration for subsequent Polymerase Chain Reaction (PCR) amplification.

The ubiquitin (*UB*) gene (GenBank DQ660326.1) was amplified by PCR in a total volume of 15 µL. The PCR master mix contained 1.0 µL of forward primer, 1.0 µL of reverse primer, 3.0 µL of DNA sample, 7.0 µL of MyTaq™ RedMix, and 3.0 µL of ddH_2O. The primers were 5'-CTCCTGATCAGCGTCTGATTT-3' (forward) and 5'-ATCATCTTGTCGCAGTTGTATTTC-3' (reverse), amplifying a 282-bp fragment. After mixing all components, the mixture was gently inverted 10 times to ensure thorough mixing. The mixture was then subjected to 40 cycles of PCR in a Bio-Rad thermocycler with the following parameters: 95 °C for 300 seconds at the beginning of the first cycle, 95 °C for 60 seconds, 56.3 °C for 60 seconds, and 72 °C for 90 seconds for all 40 cycles, followed by a final extension at 72 °C for 300 seconds.

After amplification, the products were visualised on a 2.0% agarose gel and observed under UV light. Genotyping was performed manually by observing the bands and comparing them to a 100 bp molecular weight ladder. One band at about 300 bp was designated as one electrophoretic phenotype. A single band does not indicate homozygosity or allelic identity, so sequence variation was assessed by Sanger sequencing, as agarose gels resolve only length differences. Individuals that failed to amplify after three independent PCR reactions were excluded.

For sequence validation, three to five individuals per population (28 in total) were selected for bidirectional Sanger sequencing at 1st BASE (Selangor, Malaysia), based on DNA concentration and PCR results. Three reactions failed, leaving 25 sequences (three or four per population). Quality-trimmed reads (modified Mott algorithm, error limit 0.01) and forward and reverse reads were assembled using custom Python scripts; the consensus between the primers was kept. A site was considered heterozygous (IUPAC code) if a secondary peak was ≥35% of the primary peak in both reads. The sequences have been deposited in GenBank (accession nos. PX869318-PX869321 and PX890231-PX890251).

Data Analysis

For data on the ubiquitin (*UB*) locus, band-based genotype data for analysis using GenePop 4.7 (Rousset, 2008) were prepared. The standard population genetic parameters were only calculated if polymorphism was detected. These included per-population deviations from Hardy-Weinberg Equilibrium (HWE) evaluated by exact tests (Guo & Thompson, 1992), pairwise genetic differentiation measured via F_{ST} (Weir & Cockerham, 1984), within-population inbreeding coefficients (F_{IS}) calculated according to Wright (1965) and expected heterozygosity (H_E) as a measure of genetic diversity (Nei, 1973). For loci that were monomorphic, a standard protocol (Luikart et al., 2010) that reported a fixed allele frequency of 1.0 and an effective number of alleles of 1.0, without calculating F_{ST} and F_{IS} due to mathematical constraints, was adopted. Because only three or four individuals per population were sequenced, nucleotide diversity (π; Nei, 1987) and observed (H_O) and expected (H_E) heterozygosity were calculated for the pooled sample with custom Python scripts, counting heterozygous sites as two alleles. Exact tests (Guo & Thompson, 1992) were used to test Hardy-Weinberg proportions at four common variable sites (positions 94, 109, 110 and 117) with the algorithm of Wigginton et al. (2005).

Furthermore, all trimmed sequences were run with BLASTn (BLAST+ 2.17.0) against the NCBI core nucleotide database (core_nt) (https://www.ncbi.nlm.nih.gov/genbank/) to confirm nucleotide sequence identity. Percentage similarity, E-value, and query cover were used to confirm *UB* gene identity. The GenBank accession numbers were also noted. TP7 and SPP32 were also searched against the *O. javanicus* genome assembly (WGS project RWID01; Takehana et al., 2020).

Then, evolutionary analyses were performed using MEGA12 (Kumar et al., 2024). Global pairwise alignment was used to align sequences to the common haplotype, and gaps were examined manually. Two Neighbour-Joining trees (Saitou & Nei, 1987) with Jukes-Cantor distances (Jukes & Cantor, 1969) and pairwise deletion were constructed: (i) among the studied *O. javanicus* populations, and (ii) including the studied *O. javanicus* populations and eight teleost fish species. Node support was assessed with 1000, 5000 and 10,000 bootstrap replicates for the first tree and 10,000 for the second (Felsenstein,

1985). TP10 sample was excluded because its intron was not resolved. The first tree was built by aligning the 253-bp region and midpoint-rooted without an outgroup, because the available *O. javanicus* reference sequence, such as DQ660326.1 (Woo et al., 2009), is mRNA and lacks the intron. A second tree was built in the same way from the 127-bp coding region by aligning our samples with DQ660326.1 and other teleost ubiquitin-L40 sequences (Supplementary Table 3), and rooted with *Dunckerocampus dactyliophorus*, the species most distant from *Oryzias* (Betancur-R et al., 2017).

RESULTS AND DISCUSSION

No length variation was detected at the ubiquitin (*UB*) locus among 121 *Oryzias javanicus* individuals, precluding band-based population genetic analyses. Sequencing showed almost invariant coding regions and intron variation shared among populations, and the Neighbour-Joining tree showed no grouping by population or region.

Genotyping of the Ubiquitin Locus

Genomic DNA samples were collected from 160 *O. javanicus* individuals representing eight distinct populations (Kuala Perlis, Sungai Kilim, Sungai Merbok, Kuala Sepetang, Manjung, Bagan Lalang, Seberang Perai, and Tanjung Piai). DNA concentrations were below the theoretical maximum required for high-throughput applications (> 50 ng/μL). Despite this, the DNA samples were suitable for subsequent molecular investigations, provided appropriate precautions were followed to ensure reliable results (Asahida et al., 1997; Green & Sambrook, 2012; Hansen et al., 2007).

PCR amplification of a 282-bp fragment of the ubiquitin (*UB*) gene yielded scorable results in 121 *O. javanicus* (75.6% of sampled specimens), representing eight distinct and geographically isolated populations. Agarose gel electrophoresis consistently showed a sharp band migrating close to 300 bp in all successful amplification reactions. Crucially, electrophoretic variation was not observed in any of the 121 successfully amplified samples, with all samples showing an identical electrophoretic pattern, i.e., a single band of approximately 300 bp (Table 1). This pattern was consistent across all successfully amplifiable samples in all populations, i.e., Bagan Lalang (20 samples), Sungai Kilim (19 samples), Sungai Merbok (17 samples), Manjung (16 samples), Tanjung Piai (16 samples), Seberang Perai (15 samples), Kuala Sepetang (10 samples), and Kuala Perlis (8 samples).

The lack of length variation at this gene locus precluded all population genetics investigations based on genotypic variation. As a result, all band-based estimates of genetic diversity remained constant since allele frequency = 1.0 (for the 300-bp allele) and effective number of alleles = 1.0 in all populations. Population differentiation indices F_{ST}, inbreeding coefficient F_{IS}, and Hardy-Weinberg equilibrium tests are mathematically undefined in this dataset since there is a complete lack of genotypic variation. Thirty-nine

Table 1
Ubiquitin locus genotyping summary by population and individual

Population	Individual	Ubiquitin Locus	Population	Individual	Ubiquitin Locus
Kuala Perlis	KP4	-	Sungai Kilim	SK1	300, 300
	KP5	-		SK3	300, 300
	KP9	-		SK4	300, 300
	KP16	-		SK5	300, 300
	KP19	-		SK6	300, 300
	KP24	300, 300		SK7	300, 300
	KP26	300, 300		SK8	300, 300
	KP28	-		SK10	300, 300
	KP29	-		SK11	300, 300
	KP31	300, 300		SK12	300, 300
	KP33	300, 300		SK13	300, 300
	KP36	-		SK16	300, 300
	KP38	-		SK17	300, 300
	KP39	-		SK22	300, 300
	KP41	-		SK32	300, 300
	KP42	-		SK36	300, 300
	KP44	300, 300		SK40	300, 300
	KP47	300, 300		SK43	300, 300
	KP48	300, 300		SK48	300, 300
	KP49	300, 300		SK50	-
Sungai Merbok	SMK5	300, 300	Kuala Sepetang	KSP1	300, 300
	SMK7	300, 300		KSP5	-
	SMK9	300, 300		KSP7	-
	SMK10	300, 300		KSP8	300, 300
	SMK14	300, 300		KSP9	-
	SMK20	300, 300		KSP15	-
	SMK25	300, 300		KSP16	300, 300
	SMK26	300, 300		KSP17	300, 300
	SMK30	300, 300		KSP18	-
	SMK31	-		KSP21	-
	SMK32	-		KSP23	300, 300
	SMK34	300, 300		KSP24	-
	SMK35	-		KSP32	-
	SMK37	300, 300		KSP34	-
	SMK38	300, 300		KSP38	300, 300
	SMK39	300, 300		KSP43	300, 300
	SMK40	300, 300		KSP44	300, 300
	SMK44	300, 300		KSP45	-
	SMK45	300, 300		KSP48	300, 300
	SMK46	300, 300		KSP49	300, 300

Table 1 (*Continued*)

Population	Individual	Ubiquitin Locus	Population	Individual	Ubiquitin Locus
Manjung	MP1	300, 300	Bagan Lalang	BL1	300, 300
	MP2	-		BL2	300, 300
	MP4	-		BL4	300, 300
	MP5	-		BL5	300, 300
	MP10	-		BL7	300, 300
	MP14	300, 300		BL8	300, 300
	MP18	300, 300		BL9	300, 300
	MP20	300, 300		BL11	300, 300
	MP21	300, 300		BL12	300, 300
	MP22	300, 300		BL13	300, 300
	MP23	300, 300		BL18	300, 300
	MP25	300, 300		BL19	300, 300
	MP26	300, 300		BL20	300, 300
	MP39	300, 300		BL22	300, 300
	MP40	300, 300		BL23	300, 300
	MP41	300, 300		BL24	300, 300
	MP43	300, 300		BL25	300, 300
	MP45	300, 300		BL26	300, 300
	MP47	300, 300		BL30	300, 300
	MP48	300, 300		BL32	300, 300
Seberang Perai	SPP4	300, 300	Tanjung Piai	TP4	300, 300
	SPP6	-		TP5	300, 300
	SPP8	300, 300		TP6	300, 300
	SPP9	-		TP7	300, 300
	SPP12	300, 300		TP10	300, 300
	SPP13	-		TP11	300, 300
	SPP14	300, 300		TP12	300, 300
	SPP16	-		TP14	300, 300
	SPP19	300, 300		TP16	300, 300
	SPP22	300, 300		TP22	300, 300
	SPP27	300, 300		TP26	300, 300
	SPP32	300, 300		TP29	300, 300
	SPP35	300, 300		TP30	300, 300
	SPP38	300, 300		TP33	300, 300
	SPP39	300, 300		TP34	-
	SPP42	300, 300		TP37	300, 300
	SPP44	300, 300		TP39	300, 300
	SPP47	300, 300		TP45	-
	SPP49	300, 300		TP48	-
	SPP50	-		TP50	-

Note. “300, 300”: a single band of approximately 300 bp; “-”: no amplification after three PCR attempts. A single band does not establish homozygosity

samples (24.4%) failed to amplify in any of the three independent PCR trials conducted under optimal PCR conditions and were excluded from this dataset in accordance with standard PCR protocols. Failures were concentrated in Kuala Perlis (12 of 20) and Kuala Sepetang (10 of 20). They may reflect low DNA quality or null alleles at primer-binding sites, so *UB* diversity may be underestimated in these populations.

Our genotyping results show no detectable length variation among individuals, making it impossible to perform other population genetic analyses such as genotype frequencies, Hardy-Weinberg equilibrium, and F_{ST}-based population differentiation. These findings are consistent with the well-documented fact that ubiquitin (*UB*), the protein to which our results apply, is under strong evolutionary constraint because it is essential for proper eukaryotic cell function (Zuin et al., 2014). The high sequence similarity among species for ubiquitin, such as 96% sequence identity between yeast and humans, also reflects the strong evolutionary constraint on ubiquitin (Mavor et al., 2016). Intense purifying selection imposes this constraint on the ubiquitin gene family: although DNA sequences diverge significantly, protein sequences are virtually identical among fungi, plants, and animals (Nei et al., 2000). Moreover, this strong purifying selection extends beyond nonsynonymous sites to synonymous sites in functionally critical genes, such as those in *Drosophila* (Lawrie et al., 2013). Ubiquitin's critical, non-redundant role in the ubiquitin-proteasome pathway, which regulates protein degradation and other cellular functions, also strongly constrains sequence divergence (Schmidt et al., 2021; Upadhyay & Joshi, 2024). In fact, studies on ubiquitin have shown that it tolerates mutations well under normal conditions but shows reduced tolerance under different chemical conditions, indicating that the sequence is optimised for proper function (Mavor et al., 2016). In other words, the absence of detectable length variation at the ubiquitin locus observed in our study, which makes it impossible to perform other population genetic analyses, is entirely consistent with previous studies on this critical and non-redundant protein. However, sequencing (see below) showed that this single band masked intron variation. Nevertheless, this finding indicates strong conservation of *UB* across these eight populations and supports using *UB* as a stable marker in an ecotoxicological study of *O. javanicus* (Woo et al., 2009).

UB Gene Validation

Sequences were obtained from 25 individuals (Supplementary Table 2); TP10 gave low-quality reads in two runs. The 237-bp region between the primers (251 bp in TP7) comprises 57 bp encoding ubiquitin residues 45-64, a 110-bp intron (124 bp in TP7) with canonical GT-AG splice sites, and 70 bp encoding ubiquitin residues 64-76 and the N-terminus of ribosomal protein L40. Thus, the locus is a ubiquitin-L40 fusion gene. The apparent difference after CAGAAAG is because the mRNA reference DQ660326.1 lacks the intron.

Table 2
Polymorphic sites in the ubiquitin (UB) fragment among 23 Oryzias javanicus individuals

Position	Region	Alleles	n	Heterozygous individuals	Homozygous for the less common allele
30	Exon 1 (Thr55Pro)	A/C	23	-	TP16
59	Intron	T/C	23	KSP8	-
94	Intron	A/T	22	KP33, KP38, SK22, SK32, SMK14, KSP8, KSP48, TP16	SMK26, SMK39, SPP32, KSP23, BL7
105	Intron	T/C	21	KSP8	-
109	Intron	T/C	19	SK32, SMK14, SMK26, KSP48, BL11, BL22	KP44, SPP35, SPP42, MP25, MP41
110	Intron	G/T	20	SK32, SMK14, SMK26, KSP48, BL11, BL22	KP44, SPP35, SPP42, MP25, MP41, MP45
111	Intron	T/C	21	MP41	-
117	Intron	T/C	20	KP44, SK22, SMK14, SMK26, SMK39, KSP48	SK32, SPP32, KSP8, KSP23, MP25, BL7, TP16
128	Intron	G/A	22	-	TP16
140	Intron	G/A	23	KP44	-
142	Intron	T/C	23	SPP42, BL11	-
190	Exon 2 (synonymous)	C/T	19	BL7	-

Note. Positions are numbered from the first base after the forward primer; n, individuals scored. TP7 and TP10 are excluded; sites 109 and 110 are linked. KP, Kuala Perlis; SK, Sungai Kilim; SMK, Sungai Merbok; KSP, Kuala Sepetang; MP, Manjung; SPP, Seberang Perai; BL, Bagan Lalang; TP, Tanjung Piai

Except for TP7, 12 polymorphic sites in 23 sequences were identified, 10 of which occurred in the intron (Table 2). BL7 was heterozygous at a synonymous site in the coding region, while TP16 had C at position 30, replacing Thr55 with Pro in ubiquitin; this residue is conserved in eukaryotes, and the single-individual variant warrants confirmation. Common intron variants were found at positions 94 (A/T), 109-110 (CT/TG) and 117 (C/T); heterozygotes were found in all these areas and homozygotes in several populations. The 23 sequences were the same length. The pooled sample had a nucleotide diversity of 0.0216 in the intron and 0.0011 in the coding regions (0.0106 overall). At the four common sites, H_O = 0.30-0.36 was lower than H_E = 0.49-0.51, but not significantly (exact tests, P = 0.08-0.37).

The common haplotype was found in 23 individuals, while TP7 had a divergent homozygous haplotype differing from the common haplotype by 19 substitutions (including 15 intronic and four synonymous) and indels that lengthened its intron to 124 bp (Jukes-Cantor distance 0.088-0.105, versus ≤0.022 among the remaining 23 individuals). The TP7 gene was found to have a divergent allele in the *O. javanicus* genome assembly: TP7

and the common haplotype were both located at the same single locus in the genome (TP7 scaffold OJAV_Scaf_054, GenBank RWID01000054.1, 84.2% identity; common haplotype scaffold OJAV_Scaf_054, GenBank RWID01000054.1, 99.6% identity), and no other copy of TP7 with an intron was located.

Validation of the *UB* gene sequence shows strong functional constraints, as indicated by the strict conservation of the open reading frames (ORFs) despite variation in the intron. This reflects strong purifying selection on the functional protein product. This pattern has been observed in other important genes, such as the ovine keratin-associated protein (*KAP*) gene, which is subject to purifying selection that conserves the open reading frames while allowing variability in flanking regions (Gong et al., 2016; Li et al., 2019). Similarly, analysis of the ovine *KRTAP1-2* gene has shown several DNA sequences with significant variability outside the critical coding regions, emphasising the functional constraints on the protein product (Gong et al., 2011; Zhou et al., 2018). This has been shown to be the result of the significant evolutionary conservation of the coding sequences, which has been shown to be part of the general genomic organisation, as indicated by the comparative analysis of the mammalian orthologues, which has shown that the evolution of the housekeeping genes, which are responsible for the fundamental cellular processes, occurs much more slowly compared to the evolution of the tissue-specific genes, which are indicated by the lower rates of nonsynonymous substitutions and Ka/Ks ratios (Zhang & Li, 2004). In contrast, variability in non-coding regions may serve as a reservoir for evolutionary change in gene regulation. Standing variation in non-coding regions is well understood to serve as a reservoir for rapid adaptation through gene regulation without compromising protein-coding regions (Closser et al., 2021; Pai et al., 2015). This aligns with the neutral theory of molecular evolution, which proposes that functional regions tend to be conserved while neutral regions tend to vary (Kimura, 1983; Wang et al., 2020). In *O. javanicus*, *UB* variation is largely confined to introns and is shared among populations. However, our data do not test whether these variants affect *UB* expression or fitness, so their adaptive role remains hypothetical.

Intron diversity is about 20 times that of coding regions and exceeds the genome-wide heterozygosity of *O. javanicus* (0.96%, Takehana et al., 2020), indicating that introns are not unusually conserved. This locus is expected to be under strong constraint in the coding regions because the fusion protein is cleaved to ubiquitin and a ribosomal protein essential for ribosome biogenesis (Finley et al., 1989), and because the gene is highly expressed as a ribosomal protein gene (Baker & Board, 1991). The slight heterozygote deficit may reflect conservative heterozygote calling, allele dropout at variable primer sites (which would also explain the frequent amplification failures noted above), or pooling of individuals from different populations; the small sample size does not allow these causes to be separated.

The TP7 allele was the most divergent sequence found. It shares the same locus as the common haplotype, but differs at 19 sites and in intron length, and is more similar to the

O. latipes sequence (83.07% identity compared to 78.35-80.31%). This is an extremely deep allele that could represent an old lineage persisting in a large population, or it could result from introgression from a congener like *O. melastigma* (*O. dancena*), which is also present in the coastal waters of Malaysia (Takehana et al., 2020). These hypotheses could be tested by sequencing this intron in *O. melastigma* and in more individuals from Tanjung Piai. Because TP7 also differs at four synonymous sites, primers or probes based on the common sequence may be less efficient for amplification in expression studies.

Comparative Analysis of *UB* Gene

In BLASTn searches (Supplementary Table 3), all 25 sequences matched best the *O. javanicus* ubiquitin mRNA DQ660326.1 (98.61-100% identity) and the ubiquitin-L40 mRNAs of *O. melastigma* (98.25-98.61%) and *O. latipes* (96.49-98.25%). Query cover for these mRNAs was only 51-55% because they lack introns. Other teleosts, such as *Nematolebias whitei*, *Amphiprion melanopus* and *Echeneis naucrates*, showed lower identity (89.09-94.44%). All sequences, including TP7, aligned over their full length to a single ubiquitin-L40 locus on *O. latipes* chromosome 4 (HF933210.1), with 78.35-80.31% identity for the common haplotype and 83.07% for TP7.

Population analysis of *O. javanicus* shows extremely high sequence identity among the species (≥96.49% in the coding regions), indicating that the gene is stable and conserved. This high conservation also aligns with the use of conserved genes for phylogenetic analysis in teleost species, as observed in other fish species (Liang et al., 2020). Additionally, the high similarity with other species belonging to the same genus (*O. melastigma* and *O. latipes*) and the high divergence with other teleost species are also observed, thereby clearly demarcating taxonomic boundaries, similar to those observed for phylogenetic relationships using mitochondrial genome sequence analysis for medaka fish species (Setiamarga et al., 2009). The single band of approximately 300 bp in all 121 individuals agrees with the sequencing results: only TP7 differed in length by 14 bp, which a 2% agarose gel cannot resolve.

Phylogenetic Analysis

In the Neighbour-Joining tree of 24 sequences (Figure 2), individuals other than TP7 were separated by Jukes-Cantor distances of ≤0.022, and no node received ≥50% bootstrap support with 1000, 5000 or 10,000 replicates. Populations and regions did not segregate, with Manjung individuals interspersed with individuals from Seberang Perai, Kuala Perlis, and Sungai Kilim. The long TP7 branch is not population-differentiated, but represents its divergent allele, since the other Tanjung Piai individual (TP16) did not cluster with it.

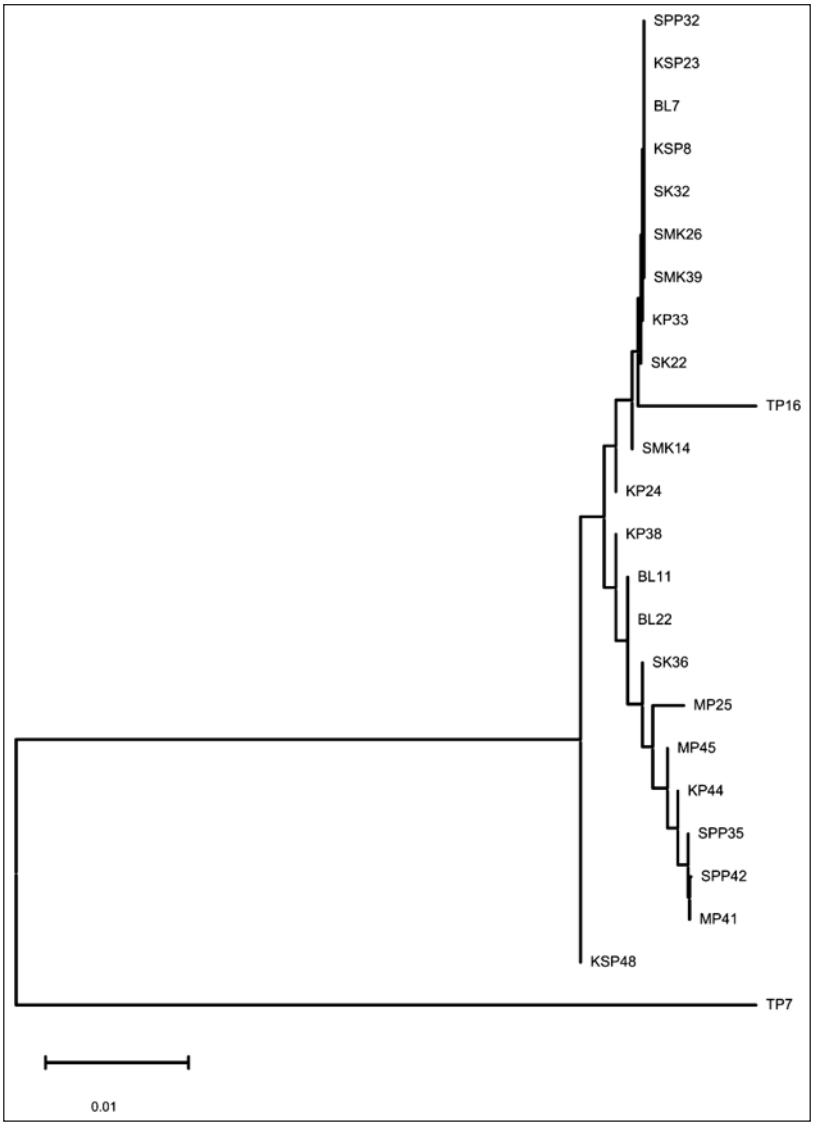


Figure 2. Neighbour-Joining tree of *UB* sequences (region between the primers) from 24 *Oryzias javanicus* samples, based on Jukes-Cantor distances with pairwise deletion (253 aligned positions) and midpoint-rooted. No node received ≥50% bootstrap support with 1000, 5000 and 10,000 bootstrap replicates; the 10,000-replicate tree is shown. Scale bar: 0.01 substitutions per site. Sample codes as in Table 2

In the second Neighbour-Joining tree (Figure 3), the Malaysian sequences grouped with the *O. javanicus* reference DQ660326.1 and *O. melastigma* (63% bootstrap support). Most Malaysian sequences differed from DQ660326.1 only at position 190, a synonymous site at which BL7 was heterozygous, so the reference carries an allele also present in Malaysia. TP7 fell outside this group but within the *Oryzias* cluster (68%), and the three

Oryzias species (65%) grouped with the other two atherinomorphs, *Leuresthes tenuis* and *Nematolebias whitei* (55%). After excluding the intron, the earlier separation of the Malaysian samples from the reference disappeared. Because the 127-bp coding region contains few, mostly synonymous, variable sites, it could not separate *O. javanicus* from *O. melastigma* (which differed from DQ660326.1 at one site), and relationships among the other teleosts were weakly supported, except for the two pomacentrids (97%).

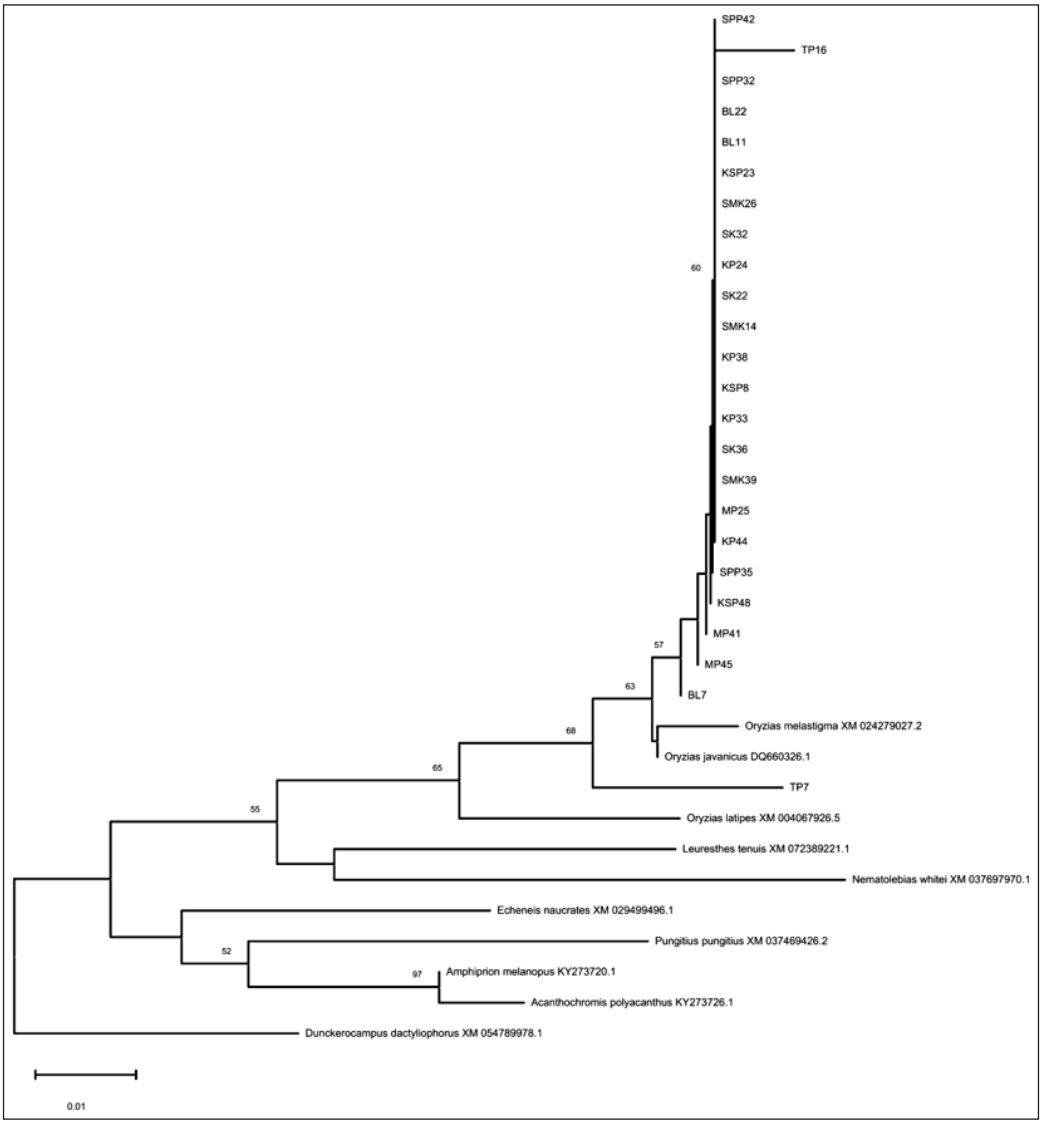


Figure 3. Neighbour-Joining tree of the *UB* coding region (127 bp) from the studied 24 *Oryzias javanicus* samples, the *O. javanicus* reference (DQ660326.1) and other teleosts (GenBank accession numbers shown), based on Jukes-Cantor distances with pairwise deletion and rooted with *Dunckerocampus dactyliophorus*. Bootstrap values ≥50% (10,000 replicates) are shown. Scale bar: 0.01 substitutions per site

Figure 3 further supports the strong conservation of the *UB* coding region. *O. javanicus* and *O. melastigma* belong to the same species group (Takehana et al., 2005) and both species occur together in estuaries along the west coast of Peninsular Malaysia from Kuala Gula to Tanjung Piai (Yusof et al., 2012b); *O. melastigma* differed from most Malaysian sequences at one or two coding sites, and the tree was unable to separate the two species. Thus, the coding region is conserved even between species, and differences at this locus occur within, not between, populations (Figure 2). Classifying *Oryzias* in the same group as the other two atherinomorphs is consistent with current teleost classification (Betancur-R et al., 2017). The introgression of *O. melastigma* (proposed above) is not supported, as TP7 did not group with *O. melastigma*, although the short coding region is not very informative.

The *UB* locus showed no phylogeographic structure, but this does not mean that the populations are panmictic. Nuclear loci have a larger effective population size than mitochondrial DNA and therefore sort into population-specific lineages about four times more slowly (Palumbi et al., 2001), so populations separated for thousands of generations can still share most intron alleles. A single locus with few variable sites also has little power, and a tree of unphased diploid sequences is not designed for population inference; allele-frequency methods such as F_{ST} applied to multiple loci would be more informative. Individuals from industrial sites and from sites with poorer water quality (WQI class IV; Supplementary Table 1) carried the same alleles as those from cleaner sites, so *UB* sequence variation does not track local conditions. For ecotoxicology, *UB* is therefore better used as a conserved reference sequence and an expression marker (Woo et al., 2009) than as a marker of population history.

Pleistocene glacial peaks lowered sea levels on the Sunda Shelf by up to approximately 120 m, uncovering the Strait of Malacca and altering mangrove distribution, which influenced mangrove genetic diversity in the region (Yang et al., 2017). The northern sites face the Andaman Sea, while Tanjung Piai faces the South China Sea. Genetic differences between the Andaman Sea and the Gulf of Thailand have been reported in *O. javanicus* (Ngamniyom et al., 2022), among sites in Indonesia in *O. javanicus* (Hardianto et al., 2023), and in *Aplocheilus panchax* across Southeast Asia (Beck et al., 2017). Such structure was not found at the *UB* locus in Malaysia. Gene flow between these populations could be limited, but if so, they may be too young to have experienced lineage sorting at a nuclear locus, so mitochondrial and genome-wide markers are required to assess this.

Although *UB* showed no structure, mangrove corridors remain important for maintaining gene flow. This aligns with an increasing body of genetic studies from mangrove ecosystems worldwide. A study of *Avicennia marina* in Iran found low to moderate genetic diversity and high differentiation among mangrove populations, which were attributed to habitat fragmentation, which reduced gene flow and increased the risk of genetic erosion (Malekmohammadi et al., 2023). A study of seven dominant mangrove species along the southern Chinese coast found that human activities and fragmentation

resulted in isolated populations with low contemporary gene flow, again underscoring the importance of conserving corridors in mangrove ecosystems (Geng et al., 2021). A detailed study of *Rhizophora mangle* in Honduras also found high levels of genetic structure in small forest patches, indicating that corridor disruptions increase genetic isolation and drift, particularly in peripheral communities (Canty et al., 2022). These studies show that conserving mangrove corridors is critical in preventing genetic erosion in marine organisms by maintaining essential gene flow between mangrove populations. In addition, our results provide insights into how environmental gradients and anthropogenic influences impact gene diversity. For instance, the decreased DNA quality and increased A260/230 ratios in DNA from industrial estuaries suggest that contaminants degrade nucleic acids. Wilfinger et al. (1997) support this association. Other studies support this, showing that air quality components such as PM2.5 induce oxidative stress, thereby damaging and degrading DNA (Ullah et al., 2021). While evidence supports steady purifying selection in coding regions, it also suggests that population genetic diversity is susceptible to environmental stressors (Dukler et al., 2022).

CONCLUSION

The ubiquitin gene of *Oryzias javanicus* is highly conserved among populations from Peninsular Malaysia: the fragment, part of a ubiquitin-L40 fusion gene, showed no length variation and nearly invariant coding sequences, indicating steady purifying selection. In contrast, the intron was polymorphic ($\pi = 0.022$); its variants were shared among populations and did not differ between industrial and less disturbed sites, and one divergent allele was found at Tanjung Piai. The locus showed no phylogeographic structure, which, at a single nuclear locus, does not exclude population differentiation. The baseline data support using *UB* as a conserved reference sequence and expression marker in heavy metal studies, while multi-locus or genome-wide markers are required to assess population structure and local adaptation.

ACKNOWLEDGEMENT

We thank the Gen-Tox Research Laboratory members for their help with sample collection and anonymous reviewers for their helpful comments on an earlier version of this paper. We gratefully acknowledge financial support for this work from the Coca-Cola Foundation, provided through the Managing Plastic Pollution #6383400 and Recycling@U #6388600 Programs.

REFERENCES

Amal, M. N. A., Zarif, S. T., Suhaiba, M. S., Aidil, M. R. M., Shaqinah, N. N., Zamri-Saad, M., & Ismail, A. (2018). The effects of fish gender on susceptibility to acute *Streptococcus agalactiae* infection in

Javanese medaka *Oryzias javanicus*. *Microbial Pathogenesis*, *114*, 251-254. https://doi.org/10.1016/j.micpath.2017.11.069

Asahida, T., Yamashita, Y., & Kobayashi, T. (1997). Identification of consumed stone flounder, *Kareius bicoloratus* (Basilewsky), from the stomach contents of sand shrimp, *Crangon affinis* (De Haan) using mitochondrial DNA analysis. *Journal of Experimental Marine Biology and Ecology*, *217*(2), 153-163. https://doi.org/10.1016/s0022-0981(97)00039-7

Baker, R. T., & Board, P. G. (1991). The human ubiquitin-52 amino acid fusion protein gene shares several structural features with mammalian ribosomal protein genes. *Nucleic Acids Research*, *19*(5), 1035-1040. https://doi.org/10.1093/nar/19.5.1035

Beck, S. V., Carvalho, G. R., Barlow, A., Rüber, L., Hui Tan, H., Nugroho, E., Wowor, D., Mohd Nor, S. A., Herder, F., Muchlisin, Z. A., & de Bruyn, M. (2017). Plio-Pleistocene phylogeography of the Southeast Asian Blue Panchax killifish, *Aplocheilus panchax*. *PLoS ONE*, *12*(7), Article e0179557. https://doi.org/10.1371/journal.pone.0179557

Betancur-R, R., Wiley, E. O., Arratia, G., Acero, A., Bailly, N., Miya, M., Lecointre, G., & Ortí, G. (2017). Phylogenetic classification of bony fishes. *BMC Evolutionary Biology*, *17*, Article 162. https://doi.org/10.1186/s12862-017-0958-3

Blanton, A., Ewane, E. B., McTavish, F., Watt, M. S., Rogers, K., Daneil, R., Vizcaino, I., Gomez, A. N., Arachchige, P. S. P., King, S. A. L., Galgamuwa, G. A. P., Peñaranda, M. L. P., Al-Musawi, L., Montenegro, J. F., Broadbent, E. N., Zambrano, A. M. A., Hudak, A. T., Swangjang, K., Valasquez-Camacho, L. F., . . . Mohan, M. (2024). Ecotourism and mangrove conservation in Southeast Asia: Current trends and perspectives. *Journal of Environmental Management*, *365*, Article 121529. https://doi.org/10.1016/j.jenvman.2024.121529

Bleeker, P. (1854). Ichthyologische waarnemingen, gedaan op verschillende reizen in de residentie Banten [Ichthyological observations made during various journeys in the Residency of Banten]. *Natuurkundig Tijdschrift voor Nederlandsch-Indië, 7*, 309-326.

Braten, O., Livneh, I., Ziv, T., Admon, A., Kehat, I., Caspi, L. H., Gonen, H., Bercovich, B., Godzik, A., Jahandideh, S., Jaroszewski, L., Sommer, T., Kwon, Y. T., Guharoy, M., Tompa, P., & Ciechanover, A. (2016). Numerous proteins with unique characteristics are degraded by the 26S proteasome following monoubiquitination. *Proceedings of the National Academy of Sciences*, *113*(32), E4639-E4647. https://doi.org/10.1073/pnas.1608644113

Canty, S. W. J., Kennedy, J. P., Fox, G., Matterson, K., González, V. L., Núñez-Vallecillo, M. L., Preziosi, R. F., & Rowntree, J. K. (2022). Mangrove diversity is more than fringe deep. *Scientific Reports*, *12*(1), Article 1695. https://doi.org/10.1038/s41598-022-05847-y

Closser, M., Guo, Y., Wang, P., Patel, T., Jang, S., Hammelman, J., De Nooij, J. C., Kopunova, R., Mazzoni, E. O., Ruan, Y., Gifford, D. K., & Wichterle, H. (2021). An expansion of the non-coding genome and its regulatory potential underlies vertebrate neuronal diversity. *Neuron*, *110*(1), 70-85.e6. https://doi.org/10.1016/j.neuron.2021.10.014

Dukler, N., Mughal, M. R., Ramani, R., Huang, Y., & Siepel, A. (2022). Extreme purifying selection against point mutations in the human genome. *Nature Communications*, *13*(1), Article 43123. https://doi.org/10.1038/s41467-022-31872-6

Felsenstein, J. (1985). Confidence limits on phylogenies: An approach using the bootstrap. *Evolution*, *39*(4), 783-791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x

Finley, D., Bartel, B., & Varshavsky, A. (1989). The tails of ubiquitin precursors are ribosomal proteins whose fusion to ubiquitin facilitates ribosome biogenesis. *Nature*, *338*(6214), 394-401. https://doi.org/10.1038/338394a0

Geng, Q., Wang, Z., Tao, J., Kimura, M. K., Liu, H., Hogetsu, T., & Lian, C. (2021). Ocean currents drove genetic structure of seven dominant mangrove species along the coastlines of southern China. *Frontiers in Genetics*, Article *12*. https://doi.org/10.3389/fgene.2021.615911

Gerona-Daga, M. E. B., & Salmo, S. G. (2022). A systematic review of mangrove restoration studies in Southeast Asia: Challenges and opportunities for the United Nation's Decade on Ecosystem Restoration. *Frontiers in Marine Science*, *9*. https://doi.org/10.3389/fmars.2022.987737

Gong, H., Zhou, H., Forrest, R., Li, S., Wang, J., Dyer, J., Luo, Y., & Hickford, J. (2016). Wool keratin-associated protein genes in sheep—A review. *Genes*, *7*(6), 24. https://doi.org/10.3390/genes7060024

Gong, H., Zhou, H., Yu, Z., Dyer, J., Plowman, J. E., & Hickford, J. (2011). Identification of the ovine keratin-associated protein KAP1-2 gene (*KRTAP1-2*). *Experimental Dermatology*, *20*(10), 815-819. https://doi.org/10.1111/j.1600-0625.2011.01333.x

Green, M. R., & Sambrook, J. (2012). *Molecular cloning: A laboratory manual* (4th ed.). Cold Spring Harbor Laboratory Press.

Guo, S. W., & Thompson, E. A. (1992). Performing the exact test of Hardy-Weinberg proportion for multiple alleles. *Biometrics*, *48*(2), 361-372. https://doi.org/10.2307/2532296

Halpern, B. S., Frazier, M., Afflerbach, J., Lowndes, J. S., Micheli, F., O'Hara, C., Scarborough, C., & Selkoe, K. A. (2019). Recent pace of change in human impact on the world's ocean. *Scientific Reports*, *9*(1), Article 11609. https://doi.org/10.1038/s41598-019-47201-9

Hansen, H. M., Wiemels, J. L., Wrensch, M., & Wiencke, J. K. (2007). DNA quantification of whole genome amplified samples for genotyping on a multiplexed bead array platform. *Cancer Epidemiology Biomarkers & Prevention*, *16*(8), 1686-1690. https://doi.org/10.1158/1055-9965.epi-06-1024

Hardianto, E., Hirayama, M., Wijayanti, D. P., & Imai, H. (2023). Molecular ecology of the Javanese ricefish, *Oryzias javanicus* (Bleeker): Genetic divergence along the Indonesian Archipelago. *Marine and Freshwater Research*, *74*(15), 1314-1323. https://doi.org/10.1071/MF23129

Hershko, A., & Ciechanover, A. (1998). The ubiquitin system. *Annual Review of Biochemistry*, *67*, 425-479. https://doi.org/10.1146/annurev.biochem.67.1.425

Hilgers, L., & Schwarzer, J. (2019). The untapped potential of medaka and its wild relatives. *eLife*, *8*, Article e46994. https://doi.org/10.7554/eLife.46994

Hochstrasser, M. (1996). Ubiquitin-dependent protein degradation. *Annual Review of Genetics*, *30*(1), 405-439. https://doi.org/10.1146/annurev.genet.30.1.405

Idris, N. S. U. (2018). *Assessment of metal variations in water, sediment, and fish species (*Pangasius *sp. and* Hemibagrus *sp.) from Selangor River, Pahang River and nearby aquaculture ponds* [Doctoral dissertation, University of Malaya]. UM Students' Repository. http://studentsrepo.um.edu.my/8808/

Inoue, K., & Takei, Y. (2003). Asian medaka fishes offer new models for studying mechanisms of seawater adaptation. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*, *136*(4), 635-645. https://doi.org/10.1016/S1096-4959(03)00204-5

Jukes, T. H., & Cantor, C. R. (1969). Evolution of protein molecules. In H. N. Munro (Ed.), *Mammalian protein metabolism* (pp. 21-132). Academic Press.

Kamarudin, N. A., Zulkifli, S. Z., Azmai, M. N. A., Aziz, F. Z. A., & Ismail, A. (2020). Herbicide diuron as endocrine disrupting chemicals (EDCs) through histopathalogical analysis in gonads of Javanese medaka (*Oryzias javanicus*, Bleeker 1854). *Animals*, *10*(3), Article 525. https://doi.org/10.3390/ani10030525

Kimura, M. (1983). *The neutral theory of molecular evolution*. Cambridge University Press. https://doi.org/10.1017/CBO9780511623486

Kumar, S., Stecher, G., Suleski, M., Sanderford, M., Sharma, S., & Tamura, K. (2024). MEGA12: Molecular Evolutionary Genetic Analysis version 12 for adaptive and green computing. *Molecular Biology and Evolution*, *41*(12), Article msae263. https://doi.org/10.1093/molbev/msae263

Lawrie, D. S., Messer, P. W., Hershberg, R., & Petrov, D. A. (2013). Strong purifying selection at synonymous sites in *D. melanogaster*. *PLoS Genetics*, *9*(5), Article e1003527. https://doi.org/10.1371/journal.pgen.1003527

Li, S., Zhou, H., Gong, H., Zhao, F., Wang, J., Liu, X., Hu, J., Luo, Y., & Hickford, J. G. (2019). Identification of the ovine keratin-associated protein 21-1 gene and its association with variation in wool traits. *Animals*, *9*(7), 450. https://doi.org/10.3390/ani9070450

Liang, P., Saqib, H. S. A., Ni, X., & Shen, Y. (2020). Long-read sequencing and *de novo* genome assembly of marine medaka (*Oryzias melastigma*). *BMC Genomics*, *21*(1), 640. https://doi.org/10.1186/s12864-020-07042-7

Luikart, G., Ryman, N., Tallmon, D. A., Schwartz, M. K., & Allendorf, F. W. (2010). Estimation of census and effective population sizes: The increasing usefulness of DNA-based approaches. *Conservation Genetics*, *11*(2), 355-373. https://doi.org/10.1007/s10592-010-0050-7

Magtoon, W., & Termvidehakorn, K. (2009). First record of the Java medaka, *Oryzias javanicus* (Bleeker, 1854) from Thailand. *Tropical Natural History*, *9*(1), 63-72.

Malekmohammadi, L., Sheidai, M., Ghahremaninejad, F., Danehkar, A., & Koohdar, F. (2023). Studies on genetic diversity, gene flow and landscape genetic in *Avicennia marina*: Spatial PCA, Random Forest, and phylogeography approaches. *BMC Plant Biology*, *23*(1), Article 459. https://doi.org/10.1186/s12870-023-04475-6

Marín, I. (2014). The ubiquilin gene family: Evolutionary patterns and functional insights. *BMC Evolutionary Biology*, *14*(1), Article 63. https://doi.org/10.1186/1471-2148-14-63

Martin, S. A. M., Blaney, S., Bowman, A. S., & Houlihan, D. F. (2002). Ubiquitin-proteasome-dependent proteolysis in rainbow trout (*Oncorhynchus mykiss*): Effect of food deprivation. *Pflügers Archiv - European Journal of Physiology*, *445*(2), 257-266. https://doi.org/10.1007/s00424-002-0916-8

Mavor, D., Barlow, K., Thompson, S., Barad, B. A., Bonny, A. R., Cario, C. L., Gaskins, G., Liu, Z., Deming, L., Axen, S. D., Caceres, E., Chen, W., Cuesta, A., Gate, R. E., Green, E. M., Hulce, K. R., Ji, W., Kenner, L. R., Mensa, B., . . . Fraser, J. S. (2016). Determination of ubiquitin fitness landscapes under different chemical stresses in a classroom setting. *eLife*, *5,* Article e15802. https://doi.org/10.7554/elife.15802

Mosimann, C., Kaufman, C. K., Li, P., Pugach, E. K., Tamplin, O. J., & Zon, L. I. (2011). Ubiquitous transgene expression and Cre-based recombination driven by the ubiquitin promoter in zebrafish. *Development*, *138*(1), 169-177. https://doi.org/10.1242/dev.059345

Nei, M. (1973). Analysis of gene diversity in subdivided populations. *Proceedings of the National Academy of Sciences of the United States of America*, *70*(12), 3321-3323. https://doi.org/10.1073/pnas.70.12.3321

Nei, M. (1987). *Molecular evolutionary genetics*. Columbia University Press.

Nei, M., Rogozin, I. B., & Piontkivska, H. (2000). Purifying selection and birth-and-death evolution in the ubiquitin gene family. *Proceedings of the National Academy of Sciences*, *97*(20), 10866-10871. https://doi.org/10.1073/pnas.97.20.10866

Ngamniyom, A., Duangjai, W., Racharak, P., Wongroj, W., Karnchaisri, K., Soe, B. K., & Siriwattanarat, R. (2022). Mitochondrial variation of wild populations of two brackish ricefish (Teleostei: Beloniformes) in Thailand. *Agriculture and Natural Resources*, *56*(4), 857-866. https://doi.org/10.34044/j.anres.2022.56.4.20

Pai, A. A., Pritchard, J. K., & Gilad, Y. (2015). The genetic and mechanistic basis for variation in gene regulation. *PLoS Genetics*, *11*(1), Article e1004857. https://doi.org/10.1371/journal.pgen.1004857

Palumbi, S. R., Cipriano, F., & Hare, M. P. (2001). Predicting nuclear gene coalescence from mitochondrial data: The three-times rule. *Evolution*, *55*(5), 859-868. https://doi.org/10.1554/0014-3820(2001)055[0859:PNGCFM]2.0.CO;2

Puspitasari, R., & Lestari, L. (2018). Assessment of heavy metal in Kramat Kebo Estuary, West Java, Indonesia as habitat of *Oryzias javanicus*. *Jurnal Segara*, *14*(2). https://doi.org/10.15578/segara.v14i2.6643

Roberts, T. R. (1998). Systematic observations on tropical Asian medakas or ricefishes of the genus *Oryzias*, with descriptions of four new species. *Ichthyological Research*, *45*(3), 213-224. https://doi.org/10.1007/BF02673919

Rousset, F. (2008). GENEPOP'007: A complete re-implementation of the GENEPOP software for Windows and Linux. *Molecular Ecology Resources*, *8*(1), 103-106. https://doi.org/10.1111/j.1471-8286.2007.01931.x

Saitou, N., & Nei, M. (1987). The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution*, *4*(4), 406-425. https://doi.org/10.1093/oxfordjournals.molbev.a040454

Schmidt, M. F., Gan, Z. Y., Komander, D., & Dewson, G. (2021). Ubiquitin signalling in neurodegeneration: Mechanisms and therapeutic opportunities. *Cell Death and Differentiation*, *28*(2), 570-590. https://doi.org/10.1038/s41418-020-00706-7

Setiamarga, D. H. E., Miya, M., Yamanoue, Y., Azuma, Y., Inoue, J. G., Ishiguro, N. B., Mabuchi, K., & Nishida, M. (2009). Divergence time of the two regional medaka populations in Japan as a new time scale for comparative genomics of vertebrates. *Biology Letters*, *5*(6), 812-816. https://doi.org/10.1098/rsbl.2009.0419

Sharp, P. M., & Li, W.-H. (1987). Ubiquitin genes as a paradigm of concerted evolution of tandem repeats. *Journal of Molecular Evolution*, *25*(1), 58-64. https://doi.org/10.1007/BF02100041

Takehana, Y., Naruse, K., & Sakaizumi, M. (2005). Molecular phylogeny of the medaka fishes genus *Oryzias* (Beloniformes: Adrianichthyidae) based on nuclear and mitochondrial DNA sequences. *Molecular Phylogenetics and Evolution*, *36*(2), 417-428. https://doi.org/10.1016/j.ympev.2005.01.016

Takehana, Y., Zahm, M., Cabau, C., Klopp, C., Roques, C., Bouchez, O., Donnadieu, C., Barrachina, C., Journot, L., Kawaguchi, M., Yasumasu, S., Ansai, S., Naruse, K., Inoue, K., Shinzato, C., Schartl, M., Guiguen, Y., & Herpin, A. (2020). Genome sequence of the euryhaline Javafish medaka, *Oryzias javanicus*: A small aquarium fish model for studies on adaptation to salinity. *G3: Genes, Genomes, Genetics*, *10*(3), 907-915. https://doi.org/10.1534/g3.119.400725

Ullah, I., Zahid, M., Jawad, M., & Arsh, A. (2021). Assessment of DNA damage and oxidative stress among traffic conductors and coal miners. *Pakistan Journal of Medical Sciences*, *37*(2). https://doi.org/10.12669/pjms.37.2.2848

Upadhyay, A., & Joshi, V. (2024). The ubiquitin tale: Current strategies and future challenges. *ACS Pharmacology & Translational Science*, *7*(9), 2573-2587. https://doi.org/10.1021/acsptsci.4c00278

Usman, S., Razis, A. F. A., Shaari, K., Amal, M. N. A., Saad, M. Z., Isa, N. M., & Nazarudin, M. F. (2021). Polystyrene microplastics exposure: An insight into multiple organ histological alterations, oxidative stress and neurotoxicity in Javanese medaka fish (*Oryzias javanicus* Bleeker, 1854). *International Journal of Environmental Research and Public Health*, *18*(18), Article 9449. https://doi.org/10.3390/ijerph18189449

Viarengo, A., Lowe, D., Bolognesi, C., Fabbri, E., & Koehler, A. (2007). The use of biomarkers in biomonitoring: A 2-tier approach assessing the level of pollutant-induced stress syndrome in sentinel organisms. *Comparative Biochemistry and Physiology Part C Toxicology & Pharmacology, 146*(3), 281-300. https://doi.org/10.1016/j.cbpc.2007.04.011

Wang, J., Zhou, H., Hickford, J. G. H., Luo, Y., Gong, H., Hu, J., Liu, X., Li, S., Song, Y., Ke, N., Qiao, L., & Wang, J. (2020). Identification of the ovine keratin-associated protein 2-1 gene and its sequence variation in four Chinese sheep breeds. *Genes*, *11*(6), Article 604. https://doi.org/10.3390/genes11060604

Weir, B. S., & Cockerham, C. C. (1984). Estimating F-statistics for the analysis of population structure. *Evolution*, *38*(6), 1358-1370. https://doi.org/10.1111/j.1558-5646.1984.tb05657.x

Wigginton, J. E., Cutler, D. J., & Abecasis, G. R. (2005). A note on exact tests of Hardy-Weinberg equilibrium. *American Journal of Human Genetics*, *76*(5), 887-893. https://doi.org/10.1086/429864

Wilfinger, W. W., Mackey, K., & Chomczynski, P. (1997). Effect of pH and ionic strength on the spectrophotometric assessment of nucleic acid purity. *BioTechniques*, *22*(3), 474-481. https://doi.org/10.2144/97223st01

Woo, S., Yum, S., Jung, J. H., Shim, W. J., Lee, C., & Lee, T. (2006). Heavy metal-induced differential gene expression of metallothionein in Javanese medaka, *Oryzias javanicus*. *Marine Biotechnology*, *8*(6), 654-662. https://doi.org/10.1007/s10126-006-6046-0

Woo, S., Yum, S., Park, H.-S., Lee, T.-K., & Ryu, J.-C. (2009). Effects of heavy metals on antioxidants and stress-responsive gene expression in Javanese medaka (*Oryzias javanicus*). *Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology*, *149*(3), 289-299. https://doi.org/10.1016/j.cbpc.2008.08.002

Wright, S. (1965). The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution*, *19*(3), 395-420. https://doi.org/10.2307/2406450

Yang, Y., Li, J., Yang, S., Li, X., Fang, L., Zhong, C., Duke, N. C., Zhou, R., & Shi, S. (2017). Effects of Pleistocene sea-level fluctuations on mangrove population dynamics: A lesson from *Sonneratia alba*. *BMC Evolutionary Biology*, *17*(1), Article 22. https://doi.org/10.1186/s12862-016-0849-z

Yusof, S., Ismail, A., Alias, M. S., & Khoo, K. H. (2012a). Tolerance of Javanese medaka (*Oryzias javanicus*) embryos to salinity stress. *Sains Malaysiana*, *41*(9), 1093-1098.

Yusof, S., Ismail, A., Koito, T., Kinoshita, M., & Inoue, K. (2012b). Occurrence of two closely related ricefishes, Javanese medaka (*Oryzias javanicus*) and Indian medaka (*O. dancena*) at sites with different salinity in Peninsular Malaysia. *Environmental Biology of Fishes*, *93*(1), 43-49. https://doi.org/10.1007/s10641-011-9888-x

Yusof, S., Ismail, A., Alias, M. S., & Khoo, K. H. (2013). Effect of salinity on the survival and morphological development of Javanese medaka (*Oryzias javanicus*) larvae. *Pertanika Journal of Tropical Agricultural Science*, *36*(2), 205-217.

Zhang, L., & Li, W. (2004). Mammalian housekeeping genes evolve more slowly than tissue-specific genes. *Molecular Biology and Evolution*, *21*(2), 236-239. https://doi.org/10.1093/molbev/msh010

Zhou, H., Visnovska, T., Gong, H., Schmeier, S., Hickford, J., & Ganley, A. R. (2018). Contrasting patterns of coding and flanking region evolution in mammalian keratin associated protein-1 genes. *Molecular Phylogenetics and Evolution*, *133*, 352-361. https://doi.org/10.1016/j.ympev.2018.12.031

Zuin, A., Isasa, M., & Crosas, B. (2014). Ubiquitin signaling: Extreme conservation as a source of diversity. *Cells*, *3*(3), 690-701. https://doi.org/10.3390/cells3030690

Zulkifli, S. Z., Mohamat-Yusuff, F., Arai, T., Ismail, A., & Miyazaki, N. (2012). An assessment of the toxicokinetics of toxic metals in *Oryzias javanicus* as a surrogate species for tropical marine fish. *Journal of Environmental Science and Health, Part A*, *47*(12), 1909-1916. https://doi.org/10.1080/10934529.2012.680329

SUPPLEMENTARY DATA

Supplementary Table 1
General water quality parameters of the eight sampling localities

Locality	DO (mg/L)	BOD (mg/L)	COD (mg/L)	EC (mg/L)	pH	WQI class
Kuala Perlis (Perlis)	5.6	6.82	21.23	9,400	7.8	IV
Sungai Kilim (Langkawi, Kedah)	6.47	3	26	35,000	8.2	III
Sungai Merbok (Kedah)	10.8	4.5	92	30,000	7.4	III
Kuala Sepetang (Perak)	10.74	3.79	50	30,000	8.0	III
Manjung (Perak)	5.94	4.03	45	9,795	8.4	III
Bagan Lalang (Selangor)	10.74	7	47	98,466	7.8	IV
Seberang Perai (Penang)	10.74	7.14	50	98,466	8.3	IV
Tanjung Piai (Johor)	5.76	3	35	35,000	9.37	II

Note. DO, dissolved oxygen; BOD, biochemical oxygen demand; COD, chemical oxygen demand; EC, electrical conductivity; WQI, Water Quality Index

Supplementary Table 2

Ubiquitin (UB) gene sequences (region between the primers) of 25 Oryzias javanicus individuals from eight populations (Kuala Perlis of Perlis, Sungai Kilim of Langkawi, Sungai Merbok of Kedah, Kuala Sepetang of Perak, Manjung of Perak, Bagan Lalang of Selangor, Seberang Perai of Penang, and Tanjung Piai of Johor). GenBank accession numbers: SK22-SK36, PX890231-PX890233; SMK14-SMK39, PX890234-PX890236; KSP8-KSP48, PX890237-PX890239; MP25-MP45, PX890240-PX890242; BL7-BL22, PX890243-PX890245; SPP32-SPP42, PX890246-PX890248; TP7-TP16, PX890249-PX890251; KP24-KP44, PX869318-PX869321. IUPAC codes, heterozygous sites; N, unresolved base

Individual	Sequence 5'→ 3'
KP24	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN NNNNNNNNNNNNAGGCTGATCTATCTGTTGTCTCTCCACCA GAATCAACCCTTCACTTGGTTCTCCGTCTGCGTGGTGGCATCA TTGAGCCTTCCCTCAGACAGCTGGCTCA
KP33	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTCWAACTGCTAAATGTTTGTTTTGTNNNNAGTGCT AGTGTATTCTGAGGCTGATCTATCTGTTGTCTCTCCACCAGAA TCAACCCTTCACTTGGTTCTCCGTCTGCGTGGTGGCATCATTG AGCCTTCCCTCAGACAGCTGGCTCA
KP38	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTCWAACTGCTAAATGTTNNTTTTGTTAAAAGTGCTA GTGTATTCTGAGGCTGATCTATCTGTTGTCTCTCCACCAGAAT CAACCCTTCACTTGGTTCTCCGTCTGCGTGGTGGCATCATTGA GCCTTCCCTCAGACANCTGGCTCA
KP44	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTCAAACTGCTAAATGTTCTTTTTGTYAAAAGTGCTA GTGTATTCTGAGRCTGATCTATCTGTTGTCTCTCCACCAGAAT NNACCCTTCACTTGGTTCTCCGTCTGCGTGGTGGCATCATTGA GCCTTCCCTCAGACAGCTGGCTCA
SK22	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTCWAACTGCTAAATGTTTGTTTTGTYAAAAGTGCTA GTGTATTCTGAGGCTGATCTATCTGTTGTCTCTCCACCAGAAT CAACCCTTCACTTGGTTCTCCGTCTGCGTGGTGGCATCATTGA GCCTTCCCTCAGACAGCTGGCTCA
SK32	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTCWAACTGCTAAATGTTYKTTTTGTCAAAAGTGCT AGTGTATTCTGAGGCTGATCTATCTGTTGTCTCTCCACCAGAA TCAACCCTTCACTTGGTTCTCCGTCTGCGTGGTGGCATCATTG AGCCTTCCCTCAGACAGCTGGCTCA
SK36	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTCAAACTGCNNNNNNNNNNNNNNNNNNAAGTGC TAGTGTATTCTNAGGCTGATCTATCTGTTGTCTCTCCACCAGAA TCAACCCTTCACTTGGTTCTCCGTCTGCGTGGTGGCATCATTG AGCCTTCCCTCAGACAGCTGGCTCA

Individual	Sequence 5'→ 3'
SMK14	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTCWAACTGCTAAATGTTYKTTTTGTYAAAAGTGCT AGTGTATTCTGAGGCTGATCTATCTGTTGTCTCTCCACCAGAA TCAACCCTTCACTTGGTTCTCCGTCTGCGTGGTGGCATCATTG AGCCTTCCCTCAGACAGCTGGCTCA
SMK26	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTCTAACTGCTAAATGTTYKTTTTGTYAAAAGTGCTA GTGTATTCTGAGGCTGATCTATCTGTTGTCTCTCCACCAGAAT CAACCCTTCACTTGGTTCTCCGTCTGCGTGGTGGCATCATTGA GCCTTCCCTCAGACAGCTGGCTCA
SMK39	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTCTAACTGCTAAATGTTTGTTTTGTYAAAAGTGCTA GTGTATTCTGAGGCTGATCTATCTGTTGTCTCTCCACCAGAAT CAACCCTTCACTTGGTTCTCCGTCTGCGTGGTGGCATCATTGA GCCTTCCCTCAGACAGCTGGCTCA
KSP8	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGYGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTCWAACTGCTAAAYGTTTGTTTTGTCAAAAGTGCT AGTGTATTCTGAGGCTGATCTATCTGTTGTCTCTCCACCAGAA TCAACCCTTCACTTGGTTCTCCGTCTGCGTGGTGGCATCATTG AGCCTTCCCTCAGACAGCTGGCTCA
KSP23	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTCTAACTGCTAAATGTTTGTTTTGTCAAAAGTGCTA GTGTATTCTGAGGCTGATCTATCTGTTGTCTCTCCACCAGAAT CAACCCTTCACTTGGTTCTCCGTCTGCGTGGTGGCATCATTGA GCCTTCCCTCAGACAGCTGGCTCA
KSP48	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTCWAACTGCTAAATGTTYKTTTTGTYAAAAGTGCT AGTGTATTCTGAGGCTGATCTATCTGTTGTCTCTCCACCAGAA TCAACCCTTCACTTGGTTCTNCGTCTGCGTGGTGGCATCATTG AGCCTTCCCTCAGACAGCTGGCTCA
MP25	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTCAAACTGCTAAATGTTCTTTTTGTCAAAAGTGCTA GTGTATTCTGAGGCTGATCTATCTGTTGTCTCTCCACCAGAAT CAACCCTTCACTTGGTTCTCCGTCTGCGTGGTGGCATCATTGA GCCTTCCCTCAGACAGCTGGCTCA
MP41	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTNTTAAGCCCAATATTCAT TTAGAAGTCAAACTGCTAAATGTTCTYTTTGTTAAAAGTGCTA GTGTATTCTGAGGCTGATCTATCTGTTGTCTCTCCACCAGAAT CAACCCTTCACTTGGTTCTNCGTCTGCGTGGTGGCATCATTGA GCCTTCCCTCAGACAGCTGGCTCA

Individual	Sequence 5'→ 3'
MP45	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTCAAACTGCTAAATGTTNTTTTTGTTAAAAGTGCTA GTGTATTCTGAGGCTGATCTATCTGTTGTCTCTCCACCAGAAT CAACCCTTCACTTGGTTCTNCGTCTGCGTGGTGGCATCATTGA GCCTTCCCTCAGACAGCTGGCTCA
BL7	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTCTAACTGCTAAATGTTTGTTTTGTCAAAAGTGCTA GTGTATTCTGAGGCTGATCTATCTGTTGTCTCTCCACCAGAAT CAACCCTTCACTTGGTTCTYCGTCTGCGTGGTGGCATCATTGA GCCTTCCCTCAGACAGCTGGCTCA
BL11	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTCAAACTGCTAAATGTTYKTTTTGTTAAAAGTGCTA GTGTATTCTGAGGCYGATCTATCTGTTGTCTCTCCACCAGAAT CAACCCTTCACTTGGTTCTCCGTCTGCGTGGTGGCATCATTGA GCCTTCCCTCAGACAGCTGGCTCA
BL22	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTCAAACTGCTAAATGTTYKTTTTGTTAAAAGTGCTA GTGTATTCTGAGGCTGATCTATCTGTTGTCTCTCCACCAGAAT CAACCCTTCACTTGGTTCTCCGTCTGCGTGGTGGCATCATTGA GCCTTCCCTCAGACAGCTGGCTCA
SPP32	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTCTAACTGCTAAATGTTTGTTTTGTCAAAAGTGCTA GTGTATTCTGAGGCTGATCTATCTGTTGTCTCTCCACCAGAAT CAACCCTTCACTTGGTTCTCCGTCTGCGTGGTGGCATCATTGA GCCTTCCCTCAGACAGCTGGCTCA
SPP35	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTCAAACTGCTAAATGTTCTTTTTGTTAAAAGTGCTA GTGTATTCTGAGGCTGATCTATCTGTTGTCTCTCCACCAGAAT CAACCCTTCACTTGGTTCTNCGTCTGCGTGGTGGCATCATTGA GCCTTCCCTCAGACANCTGGCTCA
SPP42	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTCAAACTGCTAAATGTTCTTTTTGTTAAAAGTGCTA GTGTATTCTGAGGCYGATCTATCTGTTGTCTCTCCACCAGAAT CAACCCTTCACTTGGTTCTCCGTCTGCGTGGTGGCATCATTGA GCCTTCCCTCAGACAGCTGGCTCA
TP7	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTAAGGGTTCTTTTCCCCAATGTTCATT TAGAAATCAAACTTTTAAATGTCCCTTTGTTGGAAGTGGGGTT AAAAGAGAGTGTATTCTGCGCACTAATCTAACAATATGTTGTC TCTCCACCAGAATCAACCCTTCACTTGGTTCTTCGCCTGCGTG GTGGCATCATTGAGCCCTCACTCAGACAGCTGGCTCA

Individual	Sequence 5'→ 3'
TP10	TTGCTGGAAAGCAGCTGGAGGATGGACGCACCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTNNTAAGCCCAATATTCAN NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN NNNNNNNNNNNNNNNNNNNNNNNNNNNNNNGTCTCTCCAC CAGAATCAACCCTTCACTTGGTTCTCCGTCTGCGTGGTGGCAT CATTGAGCCTTCCCTCANACAGCTGGCTCA
TP16	TTGCTGGAAAGCAGCTGGAGGATGGACGCCCCTTGTCAGACT ACAACATCCAGAAAGGTGAGAGCTCTTAAGCCCAATATTCAT TTAGAAGTCWAACTGCTAAATGTTTGTTTTGTCAAAAGTGCTA ATGTATTCTGAGGCTGATCTATCTGTTGTCTCTCCACCAGAATC AACCCTTCACTTGGTTCTCCGTCTGCGTGGTGGCATCATTGAG CCTTCCCTCAGACAGCTGGCTCA

Supplementary Table 3

BLASTn results of the 25 ubiquitin (UB) sequences against the NCBI core nucleotide database (core_nt). Values are ranges across the sequences

Species	Accession no.	Type	Query cover (%)	Identity (%)	E-value
Oryzias javanicus	DQ660326.1	mRNA	51-55	98.61-100.00	8.8e-25-1.7e-21
Oryzias melastigma	XM_024279027.2	mRNA	51-55	98.25-98.61	8.8e-25-1.7e-21
Oryzias latipes	XM_004067926.5	mRNA	51-55	96.49-98.25	2.4e-19-1.2e-16
Oryzias latipes	HF933210.1	Genomic (chr. 4)	100[a]	78.35-80.31[a]	6.8e-58-6.3e-52[a]
Nematolebias whitei	XM_037697970.1	mRNA	51-54	89.09-94.44	2.1e-20-1.5e-15
Leuresthes tenuis	XM_072389221.1	mRNA	51-54	91.67-92.73	2.9e-18-1.5e-15
Amphiprion melanopus	KY273720.1[b]	mRNA	54-55	91.23-92.98	1.2e-16-5.2e-15
Acanthochromis polyacanthus	KY273726.1[b]	mRNA	54-55	90.28-91.23	1.2e-16-5.2e-15
Pungitius pungitius	XM_037469426.2[b]	mRNA	54	90.28-91.67	1.0e-17-1.2e-16
Dunckerocampus dactyliophorus	XM_054789978.1[b]	mRNA	54	90.91-91.67	1.0e-17-1.5e-15
Echeneis naucrates	XM_029499496.1	mRNA	51-55	90.28-91.23	1.2e-16-5.6e-15

Note. Query cover, highest identity among aligned segments and lowest E-value, as reported by NCBI BLAST. Excluding TP7 (100%, 83.07%, 2.2e-64) and KP24, SK36 and TP10, whose unresolved intron bases reduced coverage or identity. [b]No hit for TP7 among the top 100 hits

PREPRINT