Research article

Molecular analysis of population and *De Novo* transcriptome sequencing of Thai medaka, *Oryzias minutillus* (Teleostei: Adrianichthyidae)

Arin Ngamniyom [a,*], Thayat Sriyapai [a], Pichapack Sriyapai [b]

[a] Major in Environment, Faculty of Environmental Culture and Eco-tourism, Srinakharinwirot University, Bangkok, 10110, Thailand
[b] Department of Microbiology, Faculty of Sciences, Srinakharinwirot University, Bangkok, 10110, Thailand

ARTICLE INFO

ABSTRACT

Thai medaka (*Oryzias minutillus*) are alternatively known as Thai rice-fish or dwarf medaka, and they widely inhabit natural freshwater environments in all regions of Thailand. In this study, we aimed to investigate the molecular genetics of the Thai medaka population in Thailand inferred from the mitochondrial control region (D-loop) and the cytochrome c oxidase subunit 1 (coxI) sequences. Furthermore, we examined RNA sequencing (RNA-seq) of adult males and females was performed with next-generation sequencing. Together, the combination of the D-loop and coxI sequences clearly distinguished the Thai medaka populations into 2 groups, such as group 1, which generally included samples from the central, northern, western, and eastern regions of the northeastern region. In this group, the fish populations seem to be a little monophyly in which the first subpopulation comprised the main samples from the northern and central regions. The second subpopulation commonly contained fish from the eastern region and specimens from the southern part of the central region near the Gulf of Thailand. Although these subgroups related to geographical distribution, bootstrap values were low in branch considered significant for both subgroups. Group 2 consisted of almost all samples from the southern population and those from the central and southern part of the northeastern region. Group 2 was found that it was made of samples from the northeastern region and samples from the southern population. A total of 73551 unigenes were identified after gene annotation. Signal transduction was the predominant protein classification among the Thai medaka orthologous groups. A differentially expressed gene (DEG) analysis identified 6 subclusters between both sexes that were composed of 257, 131, 364, 386, 114 and 108 genes. Phototransduction was the most enriched pathway and was highly expressed in males, while viral carcinogenesis, oocyte genesis, and the complement and coagulation cascades were highly expressed in females. Further details of these DEGs are discussed below. These results suggest that Thai medaka may genetically exhibit independent populations in the geographic habitats of Thailand. Moreover, these fish also reveal the genes that are conserved in other organisms and those that may be specific to this species.

1. Introduction

Teleost fish, the fish of the genus *Oryzias* that belongs to Adrianichthyidae, are recognized to consist of more than 30 species that are distributed in eastern and southern Asia; in addition, they are native to fresh and/or brackish environments (Parenti, 2008; Parenti and Hadiaty, 2010; Parenti et al., 2013; Mokodongan et al., 2014). Over the past two decades, medaka fish (rice-fish), especially Japanese medaka (*O. latipes*), have been well known as vertebrate animal models in several fields, including development biology, endocrinology, toxicology, evolution, and molecular genetics (Ishikawa, 2000; Naruse et al., 2000, 2011;

Wittbrodt et al., 2002; Kinoshita et al., 2012; Nishimura et al., 2018; Horie et al., 2018). In addition, two tropical Asian medaka, Java medaka (*O. javanicus*) and Indian rice-fish (*O. dancena*), have also been used for biological experiments as marine and/or brackish test fish (Roberts, 1998; Koyama et al., 2008; Yusof et al., 2012, 2014; Kim et al., 2014). As the smallest species of *Oryzias*, Thai medaka (Thai rice-fish or dwarf medaka, *O. minutillus*) are widespread and native to the natural freshwater of Southeast Asia, certain provinces of southern China and all regions of Thailand (Magtoon and Uwa, 1985; Uwa et al., 1988; Termvidchakron and Magtoon, 2008). They mainly inhabit paddy fields and shallow ponds [20]. In Thailand, Thai medaka populations were

genetically different and dependent on their geographical distribution by using an allozyme and karyotype analysis (Magtoon and Uwa, 1985; Uwa et al., 1988; Takata et al., 1993). Nagai et al. (2008) identified the sex determination mechanism in *O. minutillus* as the XX/XY sex determination system. Furthermore, Takehana et al. (2005) completed the phylogenetic evolution of the genus *Oryzias*, including Thai medaka, using the mitochondrial 12S and 16S rRNA and nuclear tyrosinase genes. Recently, Smitthikunanon et al. (2009) originally determined the molecular phylogeny of Thai medaka from twelve localities that were based on the displacement loop (D-loop) region of mitochondrial DNA (mtDNA), suggesting that this marker was potentially able to be used for the Thai medaka population. However, molecular data are needed to evaluate for Thai medaka, and studies of the Thai medaka populations have been remained to examine (Ngamniyom et al., 2009; Ngamniyom and Panyarachun, 2012).

The D-loop is a control region and a non-coding gene region in mtDNA that is found in many vertebrates (Saccone et al., 1987; Sumida et al., 2000). In teleost fish, the D-loop is an intraspecific genetic diversity molecular marker of fish (Xie et al., 2006). In molecular phylogenetics, this genetic differentiation marker has been utilized for fish populations, such as anchovies (*Coilia ectenes taihuensis*) in Taihu Lake, China (Zhang et al., 2017); brown chromis (*Chromis multilineata*) in the Western Atlantic Ocean of Brazil (Cruz et al., 2014); loaches (*Lefua* spp.) in Japan (Sakai et al., 2003); skipjack tuna (*Katsuwonus pelamis*) on the Indian coast (Menezes et al., 2012); bagrid catfish (*Chrysichthys nigrodigitatus*) from the Nigerian rivers (Nwafili and Gao, 2016); and African catfish (*Clarias gariepinus*) in Kenya (Nyunja et al., 2017). Similar to mitochondrial genes, cytochrome c oxidase subunit I (cox1 or co1) is a mitochondrial gene that is frequently used for population-level phylogenetic analyses or DNA barcodes in various animals (Razo-Mendivil et al., 2010; Hebert et al., 2003).

RNA sequencing (RNA-Seq) is a tool that allows the analysis of the transcriptome to provide information on the gene expression, gene prediction, differentiation, and functional transcription of RNA as well as the understanding of biological processes (Wang et al., 2009.

Kukurba and Montgomery, 2015). RNA-Seq also allows for transcriptomic approaches to study teleost fishes (Qian et al., 2014), for instance, the transcriptome characterization of adult zebrafish (*Danio rerio*) hair cells (Cody et al., 2018), transcriptome response to ectoparasitic infection in mangrove rivulus (*Kryptolebias marmoratus*) (Pawluk et al., 2018), and transcriptional response to stress in rainbow trout (*Oncorhynchus mykiss*) livers (Liu et al., 2014). In *Oryzias*, transcriptome profiles were already reported in *O. melastigma* and *O. latipes* (Lai et al., 2015; Kim et al., 2015; Wang et al., 2009).

As the aims of this study are from a molecular point of view, here, we filled the phylogenetic data gap of Thai medaka living in natural habitats in Thailand based on the mtDNA D-loop and cox1 sequences. The transcriptome profile is provided for both male and female fish. The transcriptome data may be evidence to support the knowledge through a deeper understanding of functional genes in an egg-laying fish.

2. Materials and methods

2.1. Fish capturing

O. minutillus were collected by using a hand net in natural freshwater (swallow ponds, paddy fields or small canals) from 70 localities of 70 provinces in Thailand from March 2016 to January 2018. *O. mekongensis* were captured from the northeastern region near the Mekong River, and *O. dancena* were captured from brackish water near the estuary of the Chao Phraya River and the Gulf of Thailand. *O. javanicus* were caught in the brackish water from the mangrove area of southern Thailand. *O. woworae* and *O.* cf. *songkhramensis* were purchased from a commercial pet shop in Bangkok, Thailand.

2.2. Molecular genetic analysis

The fish were anaesthetized in a tricaine methane-sulfonate solution ($\approx$300 mg/L). The caudal fins were cut and stored in absolute ethanol at –20 °C prior to DNA extraction. The genomic DNA of the fish fin was extracted by using the DNeasy Blood & Tissue Kit (Qiagen, Germany) according to the manufacturer's instructions. The two pairs of primers that were used were 5′–ggRggWATACCWgTTgAACACCC–3′ and 5′–ggTTTTRACCCgCAAAAgCCggg–3′ for the D-loop and 5′–CCYCA gggCTgRTAAgAAgAggA–3′ and 5′–ggTACgTgggAgTTWAACCCAC AA–3′ for Cox1 with Ex *Taq* polymerase (Takara, Japan) for DNA target amplification. The thermal cycler conditions for PCR consisted of an initial denaturation at 95 °C for 4 min; 35 cycles of denaturation for 30 s at 94 °C, annealing for 30 s at 56 °C, and extension for 2 min 30 s at 72 °C; and a final extension for 5 min at 72 °C. The PCR products were viewed under the UV transillumination of 1% agarose gels, were stained with SYBR® Safe DNA Gel Stain (Thermo Fisher Scientific, US) and were extracted from those gels by using the QIAquick Gel Extraction Kit (Qiagen, Germany). DNA sequencing was performed using an automated DNA analyzer ABI 3730xl system (Applied Biosystems, US). The DNA sequences were deposited in GenBank with the following accession numbers: MH510101-MH510174 and MK520907 for D-loop and MK156204-MK156277 and MK520908 for COX1. The nucleotide sequences of all samples were aligned and trimmed by MUltiple Sequence Comparison by Log-Expectation (MUSCLE) software (Edgar, 2004) and curation using Gblocks (Talavera and Castresana, 2007). Molecular tree with maximum likelihood was created and rendered using PhyML with 1000 bootstraps and TreeDyn (Dereeper et al., 2008). Bootstrap support was >70%.

2.3. RNA-Seq analysis

The adult males and females of *O. minutillus* were distinguished by identifying the sexual characteristics of the dorsal and anal fins with the standard length (>16 mm). Males and females were separately maintained in aquaria containing freshwater with conditions that consisted of dissolved oxygen (5–7 mg L^{-1}), pH 7.0–7.4, 28–29 °C and a 12:12 h (light:dark) light period cycle. Fish were acclimated for 2 weeks to these conditions and were supplied *ad libitum* with a Hikari medaka dance baby (Kyorin, Japan) 2 times per day. RNeasy Mini Kit (Qiagen, Germany) was used for the total RNA extractions according to the manufacturer's protocol, and the RNA was treated with DNase I from an RNase-free DNase set (Qiagen, Germany). The whole-body total RNA of 20 males and 20 females were separately pooled into Eppendorf tubes and were quantified with a Nanodrop Spectrophotometer. Then, equal concentrations of both samples were kept at –80 °C.

The mRNAs were enriched from the total RNA using Oligo(dT)$_{25}$ beads (200 ng RNA per sample), and fragments were randomly placed in RNA fragmentation buffer (Illumina). The cDNA was synthesized using a TruSeq mRNA kit (Illumina) with the SuperScript III First-Strand Synthesis System (Invitrogen) and random hexamers for first-strand synthesis. Second-strand cDNA synthesis was performed in a buffer (mRNAseq Illumina) supplemented with dNTPs, RNase H and *E. coli* polymerase I, and the cDNAs were purified by AMPure XP beads. The cDNA library quantification was measured by using a Qubit 2.0 fluorometer (Life Technologies), and the insert size was verified with an Agilent 2100 Bioanalyzer. The sequencing was conducted on a HiSeq 2500 Sequencing System (Illumina) according to Dillies et al. (2013). Transcriptome data was deposited in a public functional genomics data of Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/) with an accession of GSE133888 and GPL26891. Assembly data was analysed by the trinity software (r20140413p1).

For gene functional annotation, Nt (NCBI nucleotide sequences) was performed by NCBI blast 2.2.28+ with e-value threshold (1e-5). Diamond software version 0.8.22 was used for Nr (NCBI non-redundant protein sequences), SwissProt with e-value threshold (1e-5), KOG

(euKaryotic Orthologous Groups) with e-value parameter (1e-3) and KAAS r140224 with e-value (1e-10) for KEGG. HMMER 3.0 package of hmmscan was used for Pfam and the prediction of protein structure domain with the 0.01 e-value threshold. Blast2GO software

b2g4pipe_v2.5 (Götz et al., 2008) and novogene script were preferred for Pfam which the e-value was 1e-6. Gene ontology (GO) annotation was based on the protein annotation results of NR. KEGG (Kyoto Encyclopedia of Genes and Genome) Automatic Annotation Server was employed
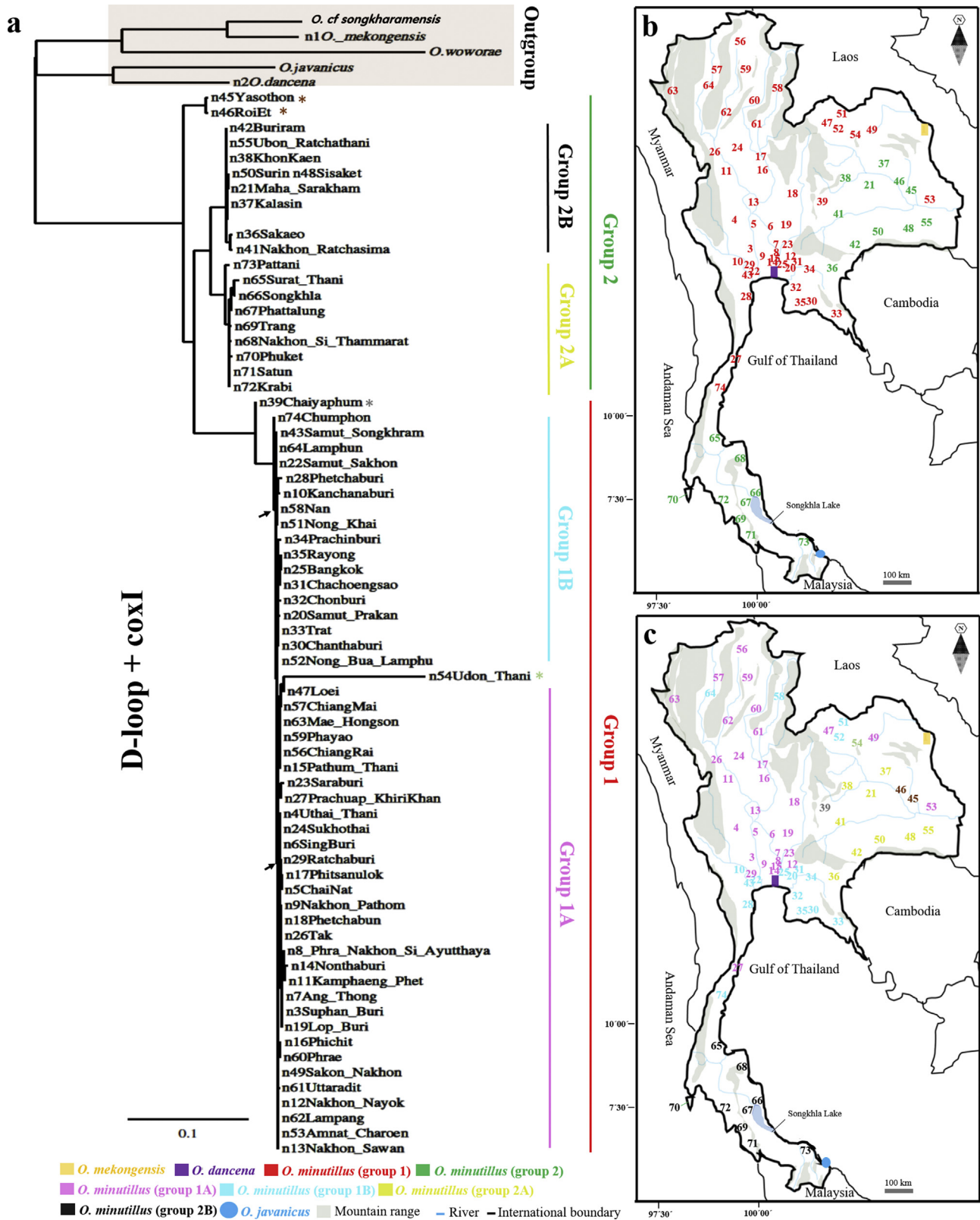

Figure 1. Molecular analysis of the maximum likelihood for concatenating sequences of D-loop and cox1 within the Thai medaka population and 5 species of *Oryzias* as outgroup (a). A map of the sample sites and the locations of groups 1 and 2 are shown. (b). A map of the fish distribution of groups 1A, 1B, 2A and 2B is shown (c). Asterisks indicate no group for classification. Arrows indicate node of 1A and 1B.

for KEGG that the threshold was1e-10. Go and KEGG enrichment were processed by GOSeq 1.10.0, topGO 2.10.0 for GO and KOBAS v2.0.12 for KEGG with corrected p-value <0.05. Hierarchiacal clustering was provided by the corset v1.05. In differential expression analysis, there were examined by using DEGSeq 1.12.0, DESeq 1.10.1 and edgeR 3.0.8 for a replicating, nonreplicating sample and a specific condition, respectively which e-value was 1e-10.

All animal experiments were conducted under the National and Institutional Guidelines for the Animal Care and Use for vertebrates by Institute for Animals for Scientific Purpose Development (IAD) National Research Council of Thailand (NRCT) with a permission license (U1-04855-2559). In addition, the Canadian Council on Animal Care guidelines on the care and use of fish in research, teaching and testing, 2005 (https://www.ccac.ca/Documents/Standards/Guidelines/Fish.pdf) was also followed.

3. Results

We aimed to infer the molecular population of 70 Thai medaka individuals with the D-loop and cox1 sequences; there were 2 groups, and group 1 consisted of localities 3 to 22, 23 to 35, 39, 43, 47, 49, 51 to 64 and 74. Group 2 consisted of localities 21, 35 to 38, 41, 42, 46, 48, 50, 55, 65 to 73 (bootstrap values >70%). From both groups, Thai medaka were isolated from *O. dancena, O. javanicus, O. mekongensis, O.* cf. *songkhramensis* and *O. woworae* (bootstrap values >90%) (Figure 1a and b). The population of group 1 was monophyletically subdivided into 2 subgroups as follows: group 1A contained fish samples from localities 3 to 9, 11 to 19, 23, 24, 26, 27, 29, 47, 49, 53, 56, 57 and 59 to 63; and group 1B contained samples from localities 10, 20, 22, 25, 27, 28, 30 to 35, 43, 51, 52, 58 and 64. However, the sample of locality 39 was distant from the clade of groups 1A and 1B and was polyphyletic to groups 1A and 1B (Figure 1a and c). However, bootstrap values were low in branch that considered no significant (bootstrap values <70%). In the population of group 2, groups 2A and 2B were monophyletic. Group 2A contained medaka samples from localities 21, 36, 37, 38, 41, 42, 48, 50 and 55 (bootstrap values >70). The samples of group 2B were fish from localities 66 to 73 (bootstrap values >70). In contrast, fish from localities 45 and 46 were polyphyletic to groups 1A and 1B. In the *Oryzias* genus of this study, *O. minutillus, O. dancena* and *O. javanicus* were phylogenetically grouped as a close relationship that was separate from *O. mekongensis, O. woworae* and *O.* cf. *songkhramensis* (Figure 1a and c).

RNA sequencing and gene annotation were performed, and a total of 73,551 unigenes were identified in this species; these unigenes were characterized into the following databases: 51,845 were from Nr, 58,175 were from Nt, 26,320 were from KOG, 45,542 were from SwissProt, 23,995 were from KEGG, 40,574 were from GO, 40,513 were from PFAM, 14,177 were found in all databases, and 62,043 were found in at least one database. The length distribution of unigenes was found to be 300–2,000 nt, and the 2,000 nt length was the most frequent (Figure 2).

The genomic RNA of Thai medaka has a similar distribution to that of *O. latipes* and to other fish species, such as *Stegastes partitus, Larimichthys crocea, Oreochromis niloticus* and *Fundulus heteroclitus* at 69.1, 14.5, 6.2, 4.8, 2.7 and 2.5%, respectively (Figure 3a). For KOG annotation, the total annotated genes were divided into 26 groups. Signal transduction and the general function prediction only of Thai medaka were the predominant groups among the KOG classifications (>15%). Transcription, posttranslational modification, protein turnover and chaperones, function unknown, intercellular trafficking, secretion, and vesicular transport and cytoskeleton were between 5% and 10%. Nineteen KOG groups were less than 5% (Figure 3b).

According to the KEGG database pathway for classification, the number of annotated genes was as follows: cellular processes: 1,351 for transport and catabolism, 1,713 for cellular community, 625 for cell motility and 675 for cell growth and death; environmental information processing: 1,463 for signal molecule and interaction, 4,270 for signal transduction and 116 for membrane transport; and genetic information
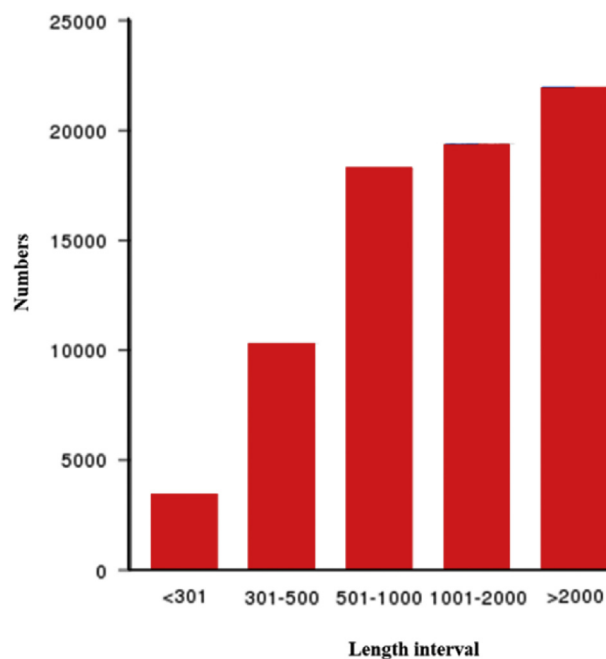

Figure 2. Length distribution with number of unigenes in Thai medaka.

processing: 734 for translation, 289 for transcription, 258 for replication and repair and 783 for folding, sorting and degradation. In addition, metabolism covered 124 for xenobiotic biodegradation, 333 for overview, 462 for nucleotide metabolism, 33 for terpenoids and polyketides, 158 for other amino acids, 308 for cofactors and vitamins, 677 for lipid, 466 for glycan biosynthesis, 231 for energy, 689 for carbohydrate, 27 for other secondary metabolites and 498 for amino acids. Organismal systems covered 488 for sensory, 1527 for nervous, 1627 for immune, 381 for excretory, 524 for environmental adaptation, 2051 for endocrine, 1101 for digestion, 878 for development and 920 for circulatory (Figure 3c).

The differentially expressed genes (DEGs) were compared, and 52,943 genes were similar in males and females. There were 9,360 and 6,203 individual genes from the males and females, respectively (Figure 3d). DEGs revealed 6 subclusters (H-clustering) between the males and females of Thai medaka that consisted of 257, 131, 364, 386, 114 and 108 genes. The subclusters 1, 4, 5 and 6 were higher in males than they were in females, whereas subclusters 2 and 3 were lower in males than they were in females (Figure 3e).

Of the DEGs that were similar between the sexes, phototransduction was the most enriched pathway and was high in males. The enriched pathways of viral carcinogenesis, ribosome biogenesis, oocyte genesis, complement and coagulation cascades and cell cycle were lower in males than they were in females (Figure 4).

In GO analysis, cellular, metabolic and single-organism processes were predominant in the biological process classification. Cell and cell part were highest in the cellular component classification. Binging functions and catalytic activities predominated among the molecular function classification (Figure 5a). In female annotated from GO database, lipid transporter activity and starch biding were highly expressed. However, in male, single organismal cell-cell adhesion and single organism cell adhesion showed the high expression.

The heatmap that represents the protein coding gene from the SwissProt and NR description (supplemental data) indicates that there were differences between males and females (Figure 5b and supplement data 1), and the following were predominantly expressed in males: parvalbumin, myosin, crystallin, rhodopsin, insulin-like growth factor-binding protein 1, serine protease, ferritin, DNA damage-inducible transcript 4 protein, phosphoenolpyruvate carboxykinase, pyruvate dehydrogenase,
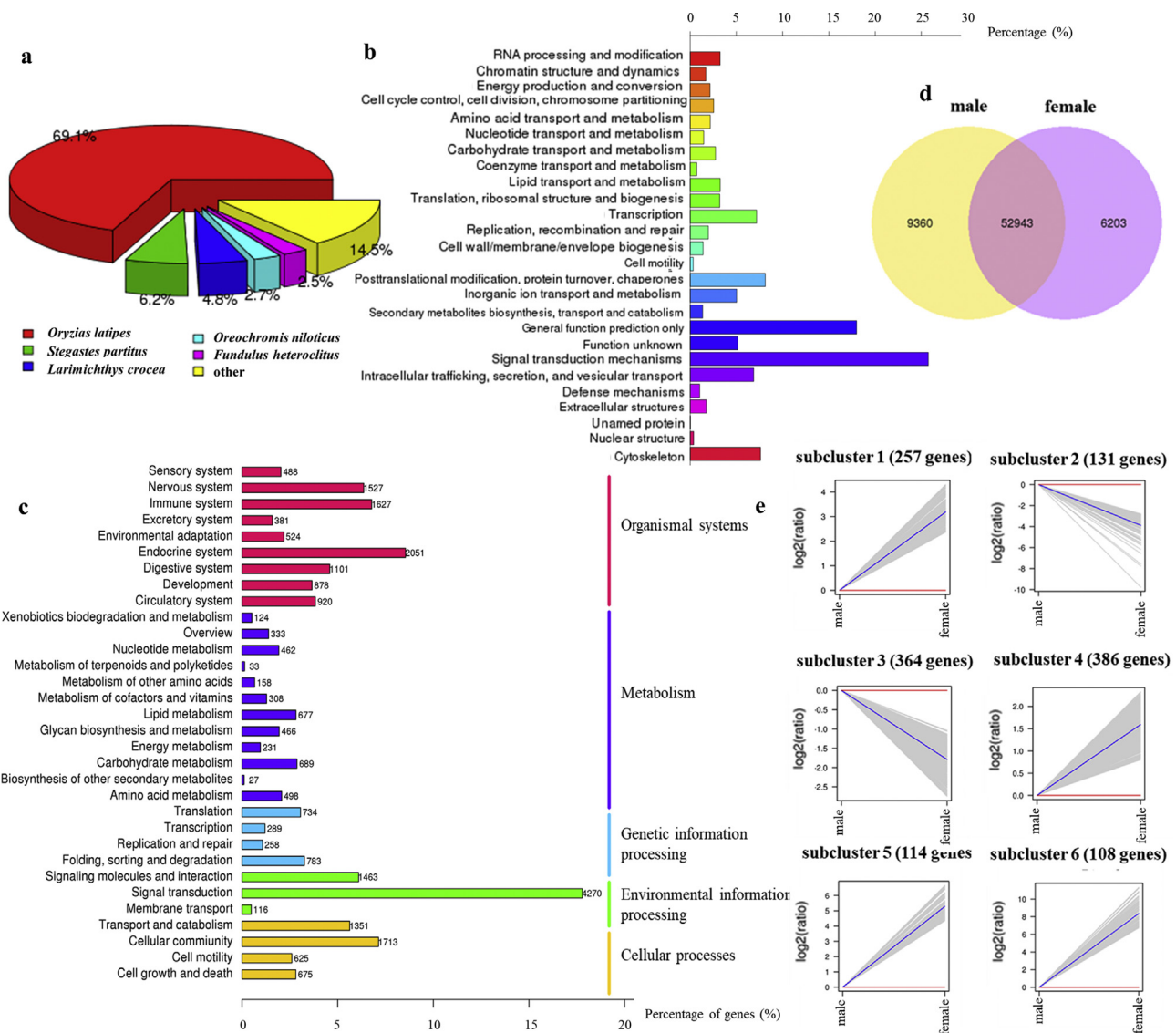
Figure 3. Species classification (a). The gene classifications are shown according to the KOG annotation (b). The KEGG classification results are shown (c). A Venn diagram of the genes expressed between males and females is shown (d). The H-cluster of males vs. females is shown (e).

synaptosomal-associated protein and 1,25-dihydroxyvitamin D(3) 24-hydroxylase. In contrast, zona pellucida sperm-binding protein, zonadhesin, vitellogenin, phospholipase, claudin-like protein, low-density lipoprotein receptor-related protein, protein B4, histone H2A type 1-A and zygote arrest proteins were highly expressed in females. In males, various genes were highly expressed, such as death-associated protein-like 1, beta-taxilin, guanylyl cyclase-activating protein 2, Krueppel-like factor 4, beta-crystallin B1, lens fibre membrane intrinsic protein, periostin, calsequestrin-1, s-arrestin, speriolin-like protein and testisin. In females, high gene expression was observed for 17-beta-hydroxysteroid dehydrogenase 14, D-beta-hydroxybutyrate dehydrogenase, bone morphogenetic protein 15 and very low-density lipoprotein receptor. In addition, this analysis also detected testis-expressed sequence 9 protein, sperm acrosome membrane-associated protein 4, sperm-specific antigen 2-like, testosterone 17-beta-dehydrogenase 3, spermatogenesis-associated protein 2, germ cell-specific gene 1-like protein, oocyte zinc finger protein, doublesex- and mab-3-related transcription factor 1, and SRY-box containing gene 13. However, those sample read counts were low. In general oncogene and related genes, there was no difference between male and female for expression. However, Ras-related protein

Rab-14 was higher in male than female. In contrast, proto-oncogene serine/threonine-protein kinase was highly expressed in female. The details of subset genes were provided in supplement data 1 (Figure S1) and supplement data 2 (Table S1).

4. Discussion

In this study, we created a molecular genetics of the Thai medaka population inferred from the D-loop and CoxI sequences, and we added the gene expression profiles of this species. In a previous report, Smitthikunanon et al. (2009) first examined the population genetics of Thai medaka using only the mitochondrial DNA D-loop region that showed 2 clades of twelve localities. Takata et al. (1993) analysed the genetic differentiation of Thai medaka by using an allozyme that represented three groups (the Mekong River, Chao Phraya River and Peninsular populations) from eleven sampling localities. Magtoon et al. (1992) found different chromosome characteristics in eighteen Thai medaka populations collected from Thailand that separated these fish populations from the central and northern populations to the northeastern and southern populations. These findings are consistent with our
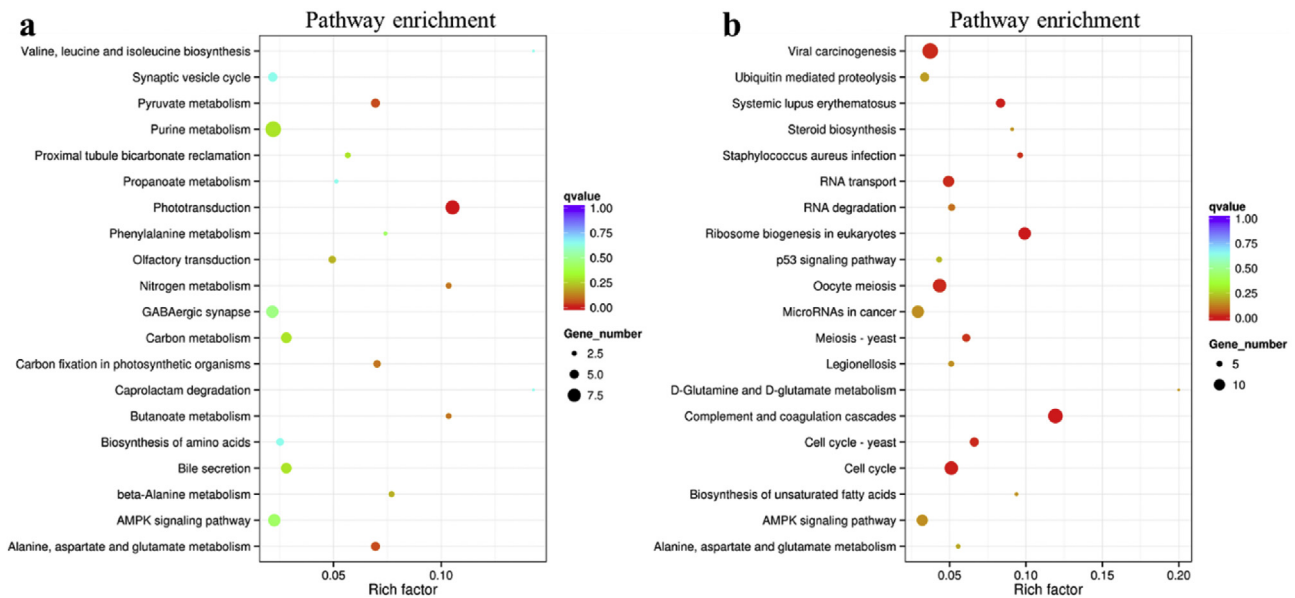
Figure 4. DEG enrichment analysis of the KEGG pathway. The genes with high expression are shown in males (a) and females (b).

phylogenetic results. These results suggest that the molecular data may be congruent with the cytogenetic data from the Thai medaka population in Thailand. The population distribution of this species may be a potential biological tool for understanding geographic evolution in Thailand. The Thai medaka populations of group 2A (from the central and southern areas of the northeastern region) were closely related to the populations of group 2B (the southern region). However, those populations were divided by the central region and the Gulf of Thailand.

Furthermore, the geographic patterns and environmental factors were different. Our molecular results disagreed with the geographical distribution that the fish population of group 2A should be related to that of group 1 rather than to that of group 2B. This suggests that the molecular data may be important to support the data of a geographical distance in the past. Among *Oryzias* spp., Thai medaka with *O. dancena* and *O. javanicus* in this study were polyphyletic to *O. mekongensis, O. woworae* and *O. cf. songkhramensis*; this corresponds to the report by Takehana
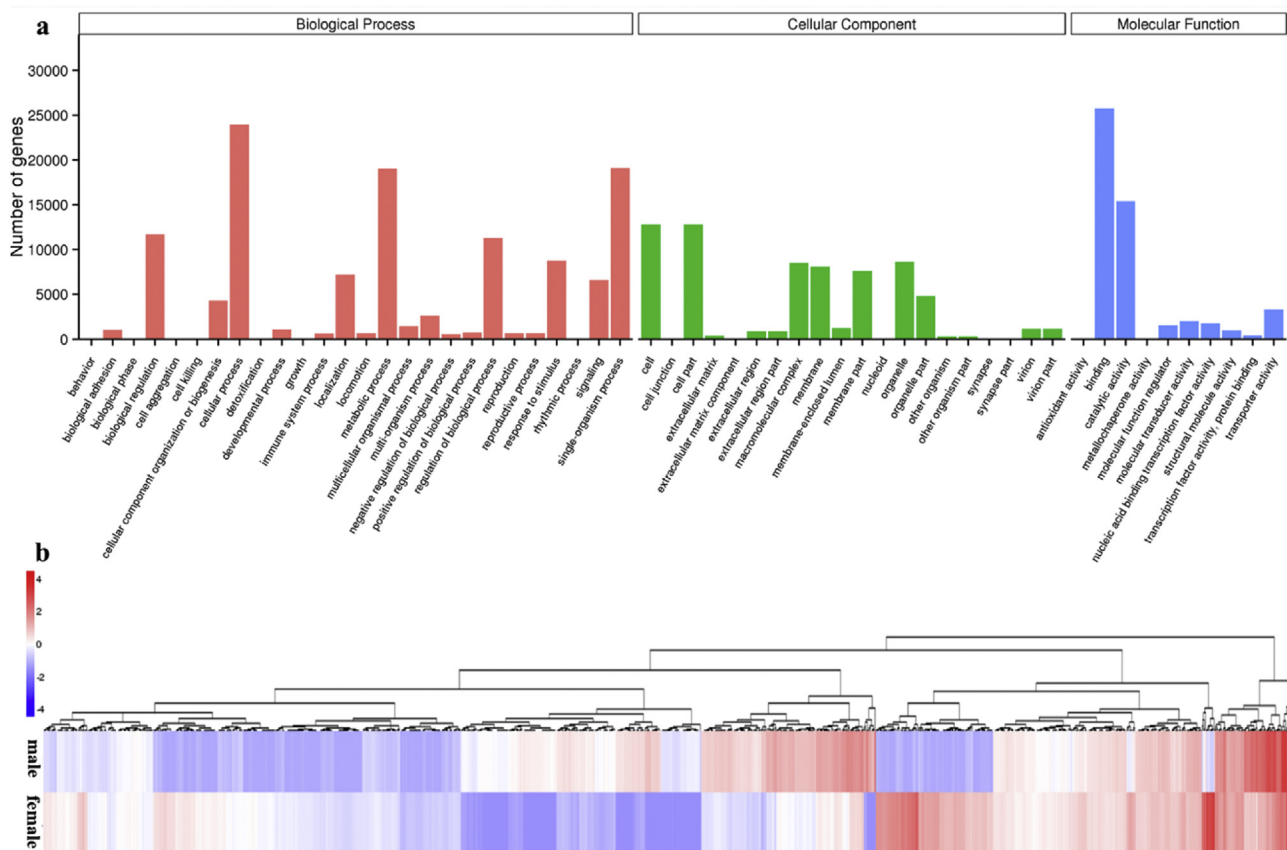

Figure 5. Gene ontology (GO) annotation (a) that grouped into biological process, cellular component and molecular function for *O. minutillus*. The heatmap that represents the protein-coding gene from the SwissProt and NR description (a detail in Figure S1 and Table S1).

et al. (2005), in which Thai medaka, *O. dancena* and *O. javanicus* were a part of the javanicus species group, and *O. mekongensis* was a part of the latipes species group based on 12S rDNA, 16S rDNA and tyrosinase sequences. In the present study, we filled the molecular gap of the *Oryzias* phylogeny with *O. cf. songkhramensis*, which may be related to the latipes species group.

Tian et al. (2019) analysed the male testes and female ovaries of Silver Sillago (*Sillago sihama*) and determined that the unigenes matched those of the marine fish *L. crocea* with a 30.07% similarity. Kim et al. (2015) reported that the transcriptomic data of the marine larvae of *O. melastigma* were similar to 67% of the data from Japanese medaka and to 8% of the data from *S. partitus*. In this study, the unigenes of Thai medaka matched to 69.1% of those in Japanese medaka and to 4.8% of those in *L. crocea*, which is related to the freshwater fish of the genus *Oryzias*. However, a large number of unknown unigenes was found in both male and female Thai medaka. In Thai medaka, phototransduction in males was more predominant than it was in females. The main genes identified were the lens fibre major intrinsic protein, rod cGMP-specific 3′,5′-cyclic phosphodiesterase subunit alpha, crystallin, rhodopsin kinase, S-arrestin and ES1 protein. These genes play an important role in phototransduction in the eyes of vertebrates (Golestaneh et al., 2004; Nishiwaki et al., 2008; Mahler et al., 2013; Osawa and Weiss, 2012; Masuda et al., 2016). Therefore, a sexual dimorphism in gene expression occurred in the eyes of Thai medaka. In addition, the high expression of muscle-specific genes was biased towards male fish. It is well known that androgens regulate masculine characteristics and muscle development via androgen receptors in fish (Martyniuk and Denslow, 2012). This suggests that the expression of muscle-specific genes in male Thai medaka corresponds to a basic concept of androgenic functions in male fish. In this transcriptomic scan, speriolin-like protein, testisin and nuclear autoantigenic sperm protein were identified. These protein coding genes are required for spermatogenesis in vertebrates (Goto et al., 2010; Tang et al., 2005; Alekseev et al., 2005). Kim et al. (2015) and Tian et al. (2019) found the expression of various genes in *S. sihama* and *O. melastigma*. These previous studies are consistent with our transcriptomic assembly, but the gene expression levels were low in Thai medaka.

It is clear that the protein-coding genes for the biological processes of the ovaries are predominantly expressed in Thai medaka females, and this is consistent with teleost fish and some mammalian species. Considering these results, the genes in the female gonads are conserved in the annotated databases of other species.

The results of this study show the geographic evolution within the relationships of Thai medaka populations in Thailand using two mitochondrial DNA sequences. A *de novo* unigene transcriptome assembly of *Oryzias* represents the genetic profiles between adult males and females; these profiles were based on the available data in the annotated gene databases.

Declarations

Author contribution statement

A. Ngamniyom: conceived and designed the experiments; analyzed and interpreted the data; wrote the paper.

T. Sriyapai: performed the experiments.

P. Sriyapai: contributed reagents, materials, analysis tools or data.

Competing interest statement

The authors declare no conflict of interest.

Additional information

Data associate with this study is available as supplementary material in Figure S1 and Table S1.

Supplementary content related to this article has been published online at https://doi.org/10.1016/j.heliyon.2019.e03079.

References

Alekseev, O.M., Widgren, E.E., Richardson, R.T., O' Rand, M.G., 2005. Association of NASP with HSP90 in mouse spermatogenic cells: stimulation of ATPase activity and transport of linker histones into nuclei. J. Biol. Chem. 4, 2904–2011.

Cody, L.B., Huizhan, L., Lei, C., Kimberlee, P.G., Yi, L., Kenneth, L.K., Kirk, W.B., David, Z.H., 2018. RNA-seq transcriptomic analysis of adult zebrafish inner ear hair cells. Sci. Data 5, 180005.

Cruz, D., Rodríguez, M., Kouri, V., Soto, Y., Zamora, L., Rodríguez, D., Barrera, M., Rehtanz, M., 2014. Genetic multipartitions based on D-loop sequences and chromosomal patterns in brown chromis, *Chromis multilineata* (Pomacentridae), in the Western Atlantic. BioMed Res. Int. 254698, 1–11.

Dereeper, A., Guignon, V., Blanc, G., Audic, S., Buffet, S., Chevenet, F., Dufayard, J.F., Guindon, S., Lefort, V., Lescot, M., Claverie, J.M., Gascuel, O., 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res. 36, W465–W469.

Dillies, M.A., Rau, A., Auber, t J., Hennequet-Antier, C., Jeanmougin, M., Servant, N., Keime, C., Marot, G., Castel, D., Estelle, J., Guernec, G., Jagla, B., Jouneau, L., Laloë, D., Le Gall, C., Schaëffer, B., Le Crom, S., Guedj, M., Jaffrézic, F., 2013. A comprehensive evaluation of normalization methods for Illumina high-throughput RNA sequencing data analysis. Briefings Bioinf. 14, 671–683.

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32 (5), 1792–1797.

Golestaneh, N., Fan, J., Fariss, R.N., Lo, W.K., Zelenka, P.S., Chepelinsky, A.B., 2004. Lens major intrinsic protein (MIP)/aquaporin 0 expression in rat lens epithelia explants requires fibroblast growth factor-induced ERK and JNK signaling. J. Biol. Chem. 279, 31813–31822.

Goto, M., O'Brien, D.A., Eddy, E.M., 2010. Speriolin is a novel human and mouse sperm centrosome protein. Hum. Reprod. 25 (8), 1884–1894.

Götz, S., García-Gómez, J.M., Terol, J., Williams, T.D., Nagaraj, S.H., Nueda, M.J., Robles, M., Talón, M., Dopazo, J., Conesa, A., 2008. High-throughput functional annotation and data mining with the Blast2GO suite. Nucleic Acids Res. 36 (10), 3420–3435.

Hebert, P.D., Ratnasingham, S., de Waard, J.R., 2003. Barcoding animal life: cytochrome c oxidase subunit divergences among closely related species. Proc. Biol. Sci. 270, S96–S99.

Horie, Y., Yamagishi, T., Shintaku, Y., Iguchi, T., Tatarazako, N., 2018. Effects of tributyltin on early life-stage, reproduction, and gonadal sex differentiation in Japanese medaka (*Oryzias latipes*). Chemosphere 203, 418–425.

Ishikawa, Y., 2000. Medakafish as a model system for vertebrate developmental genetics. Bioessays 22, 487–495.

Kim, Y.K., Lee, S.Y., Kim, B.S., Kim, D.S., Nam, Y.K., 2014. Isolation and mRNA expression analysis of aquaporin isoforms in marine medaka *Oryzias dancena*, a euryhaline teleost. Comp. Biochem. Physiol. A. Mol. Integr. Physiol. 171, 1–8.

Kim, B.M., Choi, B.S., Kim, H.S., Rhee, J.S., Au, D.W., Wu, R.S., Choi, I.Y., Lee, J.S., 2015. Transcriptome profiling of larvae of the marine medaka *Oryzias melastigma* by Illumina RNA-seq. Mar. Genomics 3, 255–258.

Kinoshita, M., Murata, K., Naruse, K., Tanaka, M., 2012. Medaka: Biology, Management, and Experimental Protocols, first ed. John Wiley and Sons, US, pp. 1–444.

Koyama, J., Kawamata, M., Imai, S., Fukunaga, M., Uno, S., Kakuno, A., 2008. Java medaka: a proposed new marine test fish for ecotoxicology. Environ. Toxicol. 23, 487–491.

Kukurba, K.R., Montgomery, S.B., 2015. RNA sequencing and analysis. Cold Spring Harb. Protoc. 951–969, 2015.

Liu, S., Gao, G., Palti, Y., Cleveland, B.M., Weber, G.M., Rexroad, C.E., 2014. RNA-seq analysis of early hepatic response to handling and confinement stress in rainbow trout. PLoS One 18, e88492.

Magtoon, W., Uwa, H., 1985. Karyotype evolution and relationship of a small ricefish. *Oryzias minutillus*. from Thailand. Proc. Jpn. Acad. 61, 157–160.

Magtoon, W., Nadee, N., Higsdhitani, T., Takata, K., Uwa, H., 1992. Karyotype evolution and geographical distribution of the Thai-medaka, *Oryzias minutillus*, in Thailand. J. Fish Biol. 41, 489–497.

Mahler, B., Chen, Y., Ford, J., Thiel, C., Wistow, G., Wu, Z., 2013. Structure and dynamics of the fish eye lens protein, γM7-crystallin. Biochemistry 52 (20), 3579–3587.

Martyniuk, C.J., Denslow, N.D., 2012. Exploring androgen-regulated pathways in teleost fish using transcriptomics and proteomics. Integr. Comp. Biol. 52 (5), 695–704.

Masuda, T., Wada, Y., Kawamura, S., 2016. ES1 is a mitochondrial enlarging factor contributing to form mega-mitochondria in zebrafish cones. Sci. Rep. 6, 22360.

Menezes, M.R., Kumar, G., Kunal, S.P., 2012. Population genetic structure of skipjack tuna *Katsuwonus pelamis* from the Indian coast using sequence analysis of the mitochondrial DNA D-loop region. J. Fish Biol. 80, 2198–2212.

Mokodongan, D.F., Tanaka, R., Yamahira, K., 2014. A new ricefish of the genus *Oryzias* (beloniformes, Adrianichthyidae) from lake Tiu, central Sulawesi, Indonesia. Copeia 3, 561–567.

Nagai, T., Takehana, Y., Hamaguchi, S., Sakaizumi, M., 2008. Identification of the sex-determining locus in the Thai medaka, *Oryzias minutillus. Cytogenet.* Genome Res. 121, 137–142.

Naruse, K., Fukamachi, S., Mitani, H., Kondo, M., Matsuoka, T., Kondo, S., Hanamura, N., Morita, Y., Hasegawa, K., Nishigaki, R., Shimada, A., Wada, H., Kusakabe, T., Suzuki, N., Kinoshita, M., Kanamori, A., Terado, T., Kimura, H., Nonaka, M., Shima, A., 2000. A detailed linkage map of medaka, *Oryzias latipes*: comparative genomics and genome evolution. Genetics 154, 1773–1784. PMCID: PMC1461045.

Naruse, K., Tanaka, M., Takeda, H., 2011. Medaka: A Model for Organogenesis, Human Disease, and Evolution, first ed. Springer, Japan, pp. 1–387.

Ngamniyom, A., Panyarachun, B., 2012. Effects of the herbicide pendimethalin on hormone receptor expressions and dorsal fin biometrics in Thai Medaka, Oryzias Minutillus (Actinopterygii: beloniformes: Adrianichthyidae). Acta Ichthyol. Piscatoria 42, 239–246.

Ngamniyom, A., Magtoon, W., Nagahama, Y., Sasayama, Y., 2009. Expression levels of hormone receptors and bone morphogenic protein in fins of medaka. Zool. Sci. 26, 74–79.

Nishimura, T., Yamada, K., Fujimori, C., Kikuchi, M., Kawasaki, T., Siegfried, K.R., Sakai, N., Kinoshita, M., Tanaka, M., 2018. Germ cells in the teleost fish medaka have an inherent feminizing effect. PLoS Genet. 29, e1007259.

Nishiwaki, Y., Komori, A., Sagara, H., Suzuki, E., Manabe, T., Hosoya, T., Nojima, Y., Wada, H., Tanaka, H., Okamoto, H., Masai, I., 2008. Mutation of cGMP phosphodiesterase 6alpha'-subunit gene causes progressive degeneration of cone photoreceptors in zebrafish. Mech. Dev. 125, 932–946.

Nyunja, C., Maina, J., Amimo, J., Kibegwa, F., Harper, D., Jun, J., 2017. Stock structure delineation of the african catfish (*Clarius gariepinus*) in selected populations in Kenya using mitochondrial DNA (dloop). J. Aquac. Res. Development 8, 485.

Nwafili, S.A., Gao, T.X., 2016. Genetic Diversity in the mtDNA control region and population structure of *Chrysichthys nigrodigitatus* from selected Nigerian rivers: implications for conservation and aquaculture. Arch. Pol. Fish. 24, 85–97.

Osawa, S., Weiss, E.R., 2012. A tale of two kinases in rods and cones. Adv. Exp. Med. Biol. 723, 821–827.

Parenti, L.R., 2008. A phylogenetic analysis and taxonomic revision of ricefishes, *Oryzias* and relatives (Beloniformes, Adrianichthyidae). Zool. J. Linn. Soc. 154, 494–610.

Parenti, L.R., Hadiaty, R.K., 2010. A new, remarkably colorful, small ricefish of the genus *Oryzias* (beloniformes, Adrianichthyidae) from Sulawesi, Indonesia. Copeia 2, 268–273.

Parenti, L.R., Hadiaty, R.K., Lumbantobing, D., Herder, F., 2013. Two new ricefishes of the genus *Oryzias* (atherinomorpha: beloniformes: Adrianichthyidae) augment the endemic freshwater fish fauna of Southeastern Sulawesi, Indonesia. Copeia 3, 403–414.

Pawluk, R.J., Uren Webster, T.M., Cable, J., Garcia de Leaniz, C., Consuegra, S., 2018. Immune-related transcriptional responses to parasitic infection in a naturally inbred fish: roles of genotype and individual variation. Genome. Biol. Evol. 10, 319–327.

Qian, X., Ba, Y., Zhuang, Q., Zhong, G., 2014. RNA-Seq technology and its application in fish transcriptomics. OMICS 18, 98–110.

Razo-Mendivil, U., Vázquez-Domínguez, E., Rosas-Valdez, R., de León, G.P., Nadler, S.A., 2010. Phylogenetic analysis of nuclear and mitochondrial DNA reveals a complex of cryptic species in *Crassicutis cichlasomae* (Digenea: apocreadiidae), a parasite of Middle-American cichlids. Int. J. Parasitol. 40, 471–486.

Roberts, T.R., 1998. Systematic observations on tropical Asian medakas or ricefishes of the genus *Oryzias*, with descriptions of four new species. Ichthyol. Res. 45, 213–224.

Saccone, C., Attimonelli, M., Sbisà, E., 1987. Structural elements highly preserved during the evolution of the D-loop-containing region in vertebrate mitochondrial DNA. J. Mol. Evol. 26, 205.

Sakai, T., Mihara, M., Shitara, H., Yonekawa, H., Hosoya, K., Miyazaki, J., 2003. Phylogenetic relationships and intraspecific variations of loaches of the genus *Lefua* (Balitoridae, Cypriniformes). Zool. Sci. 20, 501–514.

Smitthikunanon, S., Rangsiruji, A., Magtoon, W., 2009. Studies of the mitochondrial control region in *Oryzias* and phylogenetic relationships of *O. minutillus* using nucleotide sequences of the partial mitochondrial control region. SWU Sci. J. 25, 102–117. http://www.thaiscience.info/Journals/Article/SSCJ/10894846.pdf.

Sumida, M., Kaneda, H., Kato, Y., Kanamori, Y., Yonekawa, H., Nishioka, M., 2000. Sequence variation and structural conservation in the D-loop region and flanking genes of mitochondrial DNA from Japanese pond frogs. Genes Genet. Syst. 75, 79–92.

Takata, K., Hoshino, M., Magtoon, W., Nadee, N., Uwa, H., 1993. Genetic differentiation of *Oryzias minutillus* in Thailand. Jpn. J. Ichthyol. 39, 319–327.

Takehana, Y., Naruse, K., Sakaizumi, M., 2005. Molecular phylogeny of the medaka fishes genus *Oryzias* (Beloniformes: Adrianichthyidae) based on nuclear and mitochondrial DNA sequences. Mol. Phylogenetics Evol. 36, 417–428.

Talavera, G., Castresana, J., 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. 56 (4), 564–577.

Tang, T., Kmet, M., Corral, L., Vartanian, S., Tobler, A., Papkoff, J., 2005. Testisin, a glycosyl-phosphatidylinositol-linked serine protease, promotes malignant transformation in vitro and in vivo. Cancer Res. 65 (3), 868–878.

Termvidchakron, A., Magtoon, W., 2008. Development of identification of ricefish, Oryzias in Thailand. Sci. Asia 34, 426–428.

Tian, C., Li, Z., Dong, Z., Huang, Y., Du, T., Chen, H., Jiang, D., Deng, S., Zhang, Y., Wanida, S., Shi, H., Wu, T., Zhu, C., Li, G., 2019. Transcriptome analysis of male and female mature gonads of Silver Sillago (*Sillago sihama*). Genes 10, 129.

Uwa, H., Wang, R.F., Chen, Y.R., 1988. Karyotypes and geographical distribution of ricefishes from Yunnan, Southwestern China. Jpn. J. Ichthyol. 35, 332–340.

Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10, 57–63.

Wittbrodt, J., Shima, A., Schartl, M., 2002. Medaka-a model organism from the far East. Nat. Rev. Genet. 3, 53–64.

Xie, Z.Y., Du, J.Z., Chen, X.Q., Wang, Y.X., Murray, B.W., 2006. The Significance of Mitochondria Control Region (D-Loop) in Intraspecific Genetic Differentiation of Fish, 28, pp. 362–368. Yi Chuan.

Yusof, S., Ismail, A., Koito, T., Kinoshita, M., Inoue, K., 2012. Occurrence of two closely related ricefishes, Javanese medaka (*Oryzias javanicus*) and Indian medaka (*O. dancena*) at sites with different salinity in Peninsular Malaysia. Environ. Biol. Fish. 93, 43.

Yusof, S., Ismail, A., Alias, M.S., 2014. Effect of glyphosate-based herbicide on early life stages of Java medaka (*Oryzias javanicus*): a potential tropical test fish. Mar. Pollut. Bull. 85, 494–498.

Zhang, J.J., Duan, J.R., Zhou, Y.F., Peng, J.Y., Fan, D.A., 2017. Genetic diversity of mitochondrial control region (D-Loop) polymorphisms in Coilia ectenes taihuensis inhabiting Taihu Lake, China. Genet. Mol. Res. 16, gmr16019457.