P-ISSN 2586-9000
E-ISSN 2586-9027
Homepage: https://tci-thaijo.org/index.php/SciTechAsia
Science & Technology Asia
Vol.28 No.2 April - June 2023
Page: [204-215]
Original research article

Investigation of Transcriptome Profiling in *Oryzias uwai* (Teleostei: Adrianichthyidae)

Arin Ngamniyom[1,*], Babi Kyi Soe[2], Weerawich Wongroj[3], Ruthairat Siriwattanarat[4]

[1]*Major in Environment, Faculty of Environmental Culture and Eco-tourism, Srinakharinwirot University, Bangkok 10110, Thailand*
[2]*Department of Livestock Breeding and Veterinary, University of Veterinary Science, Nay Pyi Taw 05282, Myanmar*
[3]*Prasarnmit Elementary Demonstration School, Srinakharinwirot University, Bangkok 10110, Thailand.*
[4]*Faculty of Science and Technology, Suan Sunandha Rajabhat University, Bangkok 10300, Thailand.*

Received 16 November 2022; Received in revised form 16 January 2023
Accepted 16 May 2023; Available online 14 June 2023

ABSTRACT

Oryzias is a genus of teleost fish that contains more than 30 recognized species. One species in this genus, *O. latipes*, is widely used as a model animal in various fields of biological research. However, the transcriptomic profiles are unclear for some species in genus *Oryzias*. Here, we report transcriptome profiling of freshwater ricefish (*Oryzias uwai*), wherein RNA sequences of adult *O. uwai* were analyzed using a BGISEQ-500 platform with gene-annotation analysis. Transcriptomic data showed 58,483 unigenes with a total length, average length, N50, and GC content of 57,148,837, 977, and 1725 bp, and 48.40%, respectively. Unigenes were annotated by alignment using six functional databases: 40,335 from the National Center for Biotechnology Information (NCBI) Protein database, 46,456 from the NCBI Nucleotide database, 34,385 from Swiss-Prot, 29,919 from Eukaryotic Orthologous Groups (KOG), 34,240 from the Kyoto Encyclopedia of Genes and Genomes (KEGG), and 10,050 from Gene Ontology (GO). The numbers of transcriptome coding sequences, transcription-factor coding sequences, and simple sequence repeats (SSRs) were 30,132, 6,184, and 8,306, respectively. To the present body of knowledge, we contribute data from the RNA transcriptome assembly of *O. uwai*, and our results are of importance for furthering the understanding of gene function in freshwater fish.

Keywords: Medaka; Ricefish; RNA sequences; Transcriptome

***Corresponding author:** arin@g.swu.ac.th **doi:** 10.14456/scitechasia.2023.39

1. Introduction

Ricefish, or medaka, are small teleost fish that belong to the *Oryzias* genus, distributed in freshwater, brackish, and marine environments [1, 2]. More than 25 species in *Oryzias* have been recorded in East Asia, South Asia, and Southeast Asia [3, 4]. The Japanese medaka *O. latipes* is frequently utilized in many biological experiments in fields such as molecular genetics, evolution, endocrinology, and embryology [5-8]. For two decades, the basic knowledge derived from Japanese medaka, and zebrafish (*Danio rerio*), a famous model organism, has furthered researchers' understanding of biological processes [9-12]. Among species related to Japanese medaka, Java medaka (*O. Javaicus*) and marine medaka (*O. melastigma*) are also used as model vertebrates, especially in the study of molecular physiology and toxicology [13-15]. Many studies have constructed transcriptomes that comprise the full range of messenger RNA (mRNA), in order to unravel and elucidate gene functions, expression profiling, and molecular biomarkers in several vertebrates, including bony fish [16-19]. Transcriptomic data have already been compiled and evaluated in some *Oryzias* species, including *O. latipes*, *O. javanicus*, *O. melastigma*, and *O. minutillus* [15, 20, 21]. Despite the above studies on the transcriptomics of many ricefish species, there is little knowledge on *O. uwai*.

Oryzias uwai is a small fish that mainly inhabits natural freshwater in small ponds and shallow canals in Myanmar [22]. A molecular report on *O. uwai* published only the nucleotide sequences of cytochrome c oxidase subunit I [23]. Therefore, to increase the important knowledge of RNA transcripts in teleost fish, this study aimed to perform a *de novo* transcriptome assembly of *O. uwai*.

2. Materials and Methods

Oryzias uwai fish (originating from Yangon, Myanmar), characterized by dominant melanated lines on their pelvic fins, were kept in tanks containing freshwater without chlorine. Fish feeding was performed at the laboratory of Srinakharinwirot University, Ongkharak campus, Nakhon Nayok province. Oxygen was dissolved in the water using an air pump. The environment for *O. uwai* was set up with the following aquatic conditions: pH, 6.9-7.5; salinity, 0.01-0.05 ppt; temperature, 27-29 °C; and dissolved oxygen, 5.8-6.4 mg/L. The photoperiod was 12 h of light and 12 h of dark. The freshwater was changed every three days. Fish were fed ad libitum with Kyorin Hikari Food for medaka (Fukuoka, Japan) once a day. In this study, *O. uwai* were moved to new aquaria with the above conditions. Adult males and females, with a standard length of 12-18 mm, were distinguished by the secondary sex characteristics of the anal and dorsal fins. Males and females were divided into separate aquariums. These fish were used for RNA sequencing analysis. All animal experiments were conducted under the National and Institutional Guidelines for the Animal Care and Use for Vertebrates by the Institute for Animals for Scientific Purpose Development (IAD), National Research Council of Thailand (NRCT). The license was provided by the Animal Care and Use Committee of Srinakharinwirot University (SWU-A-001.1_2563).

For transcriptomic analysis, total RNA was extracted from 30 male and 30 female individuals using the RNeasy Mini Kit (Qiagen, Germany) according to the manufacturer's instructions. Total RNA was treated with DNase I from the RNase-Free DNase Set (Qiagen, Germany) during RNA purification. The quality and quantity of nucleic acids were assessed by electrophoresis on 1% agarose gels, and the concentrations were measured on NanoDrop 2000/2000c spectrophotometers (Thermo Fisher Scientific, MA, USA) and confirmed on an Agilent 2100 Bioanalyzer (Agilent RNA 6000 Nano Kit).

For library construction, mRNAs were enriched from total RNA using oligo(dT)25 beads (200 ng of RNA per sample) or random hexamers and ribosomal RNA (rRNA) depletion, which entailed rRNA fragmentation into 250 bp fragments. RNA was reverse-transcribed to double-strand complementary DNA (cDNA) with reverse transcription using N6 random primers for nonpolyadenylated RNA and oligo(dT) for polyadenylated RNA. Second-strand cDNA synthesis was performed in buffer supplemented with deoxynucleoside triphosphates (dNTPs), RNase H, and DNA polymerase I; cDNAs were purified by AMPure XP beads. cDNA libraries were constructed using the NEBNext kit (BioLabs, MA, USA). cDNA library quantification was measured using a Qubit 2.0 fluorometer (Life Technologies), and insert size was verified with an Agilent 2100 Bioanalyzer. Sequencing was conducted on a BGISEQ-500 platform (BGI-Shenzhen, China) with a paired-end sequencing length of 100 bp.

Raw reads, including low-quality, noisy, and adaptor-polluted reads, and reads with a high content of unknown bases, were filtered to obtain clean reads using the SOAPnuke v1.5.2 software [24]. Assembly of *de novo* with clean reads was performed with Trinity v2.0.6 (containing Inchworm, Chrysalis, and Butterfly), and Tgicl v2.0.6 was used to cluster transcripts [25].

For gene annotation, Nt (NCBI Nucleotide database sequence) identification was performed using the NCBI Basic Local Alignment Search Tool (BLAST) v2.2.23. Diamond software v0.8.31 was used for Nr (NCBI nonredundant protein sequences), Swiss-Prot (Universal Protein Resource Knowledgebase (UniProtKB/Swiss-Prot), and KOG (Eukaryotic Orthologous Groups), while the Kyoto Encyclopedia of Genes and Genomes (KEGG) Automatic Annotation Server (KAAS) r140224 was used for KEGG. HMMER 3.0 package hmmscan was employed for Pfam and protein prediction. Blast2GO v2.5.0 was used for Pfam. Gene Ontology (GO) annotation of the Nr results was carried out. KEGG enrichment was performed using GOSeq 1.10.0, whereas topGO 2.10.0 was used for GO, and KOBAS v2.0.12 was used for KEGG. The candidate coding area was identified using TransDecoder v3.0.1. General unigenes that were an important role for gonadal development of male and female were recorded and manually counted from results. Transcription-factor (TF) prediction was mapped to the Animal Transcription Factor Database (AnimalTFDB2.0 database), and open reading frames (ORFs) were identified using getorf EMBOSS: 6.5.7.0 [26]. TF domains were identified using hmmseach v3.0b [27].

3. Results and Discussion

After sequencing reads were filtered to obtain clean reads, *de novo* assembly of *O. uwai* produced 80,543 transcripts in total; the read length of transcripts was 66,370,692, the average length of transcripts was 824, and the GC content for all transcripts was 48.08%. After quality control, unigene metrics decreased to a total number of 58,483 unigenes, with a read length of 57,148,837, an average length of 977, and a GC content of 48.40% in all unigenes. In the distribution of annotated Nr, protein sequences displayed a 76.73% match to *O. latipes*, a 2.23% match to *Acanthochromis polyacanthus*, a 2.09% match to *Lates calcarifer*, a 1.75% match to *Stegastes partitus*, and a 17.21% match to other species (Fig. 1A). For KOG function, 25 classifications involved signal transduction mechanisms, and "general function prediction only" was attributed to the highest number of genes. In contrast, "coenzyme transport and metabolism" was attributed to the lowest number of genes in this KOG (Fig. 1B).

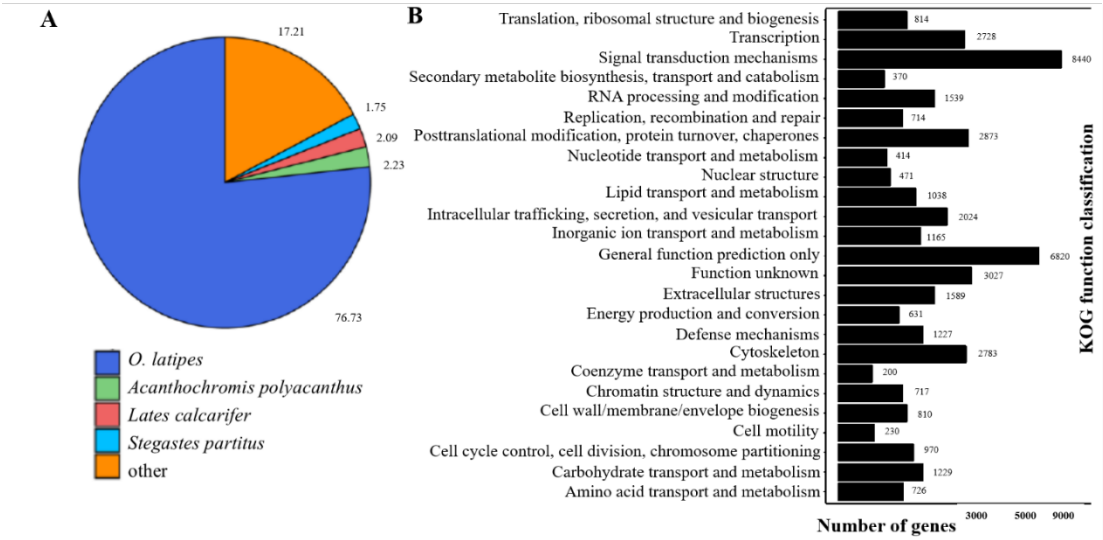

Fig. 1. Distribution of nonredundant database-annotated species (%) of *Oryzias uwai* in fish (A). Functional distribution of Eukaryotic Orthologous Groups (KOG) annotation (B).

GO annotations consisted of 26 biological processes, 18 cellular components, and 12 molecular functions. "Cellular process" was the predominant biological process. In contrast, "detoxification" and "cell killing" were attributed to relatively few genes. In "cellular component", "cell" and "cell part" were attributed to a large number of genes. Conversely, only one gene was found for "other organism" and "other organism part". In " molecular function" , "binding" was attributed to the highest number of genes, whereas "protein tag" was attributed to the lowest (Fig. 2). Among KEGG annotations, there were four included in "cellular processes", three included in "environmental information processing" , four included in "genetic information processing", 11 included in "human diseases", 12 included in "metabolism", and 10 included in "organismal systems". "Transport and catabolism" and "cellular community eukaryotes" were abundant in gene number for cellular processes. "Signal transduction" was the highest in "environmental information processing", and "folding, sorting, and degradation" was the most abundant in "genetic information processing". "Cancer: overview", "global and overview maps", and "immune system" were the predominant terms in " human diseases" , " metabolism" , and " organismal systems" , respectively (Fig. 3). For Swiss-Prot annotation, the length of the aligned genes in *O. uwai* was 295, and close to that in zebrafish (*Danio rerio*; sp|Q6DG32|S2536_DANRE) from the UniProt Knowledgebase and a bit score of 524.2. In addition, 132 aligned genes matched the Norway rat (*Rattus norvegicus*; sp|Q6P773|TAF1C_RAT), with an expectation value of 1.0×10^{-8} and a bit score of 61.6. For genes in both sexes, unigenes of males were higher than those of females in the annotations of Nr, Nt, and KEGG, but not in UniProtKB/Swiss-Prot. In males, the numbers of annotations of Nr, Nt, UniProtKB/Swiss-Prot, and KEGG were 253, 194, 283, and 407, respectively. In females, the numbers of annotations of Nr, Nt, UniProtKB/Swiss-Prot, and KEGG were 197, 161, 325, and 395, respectively.

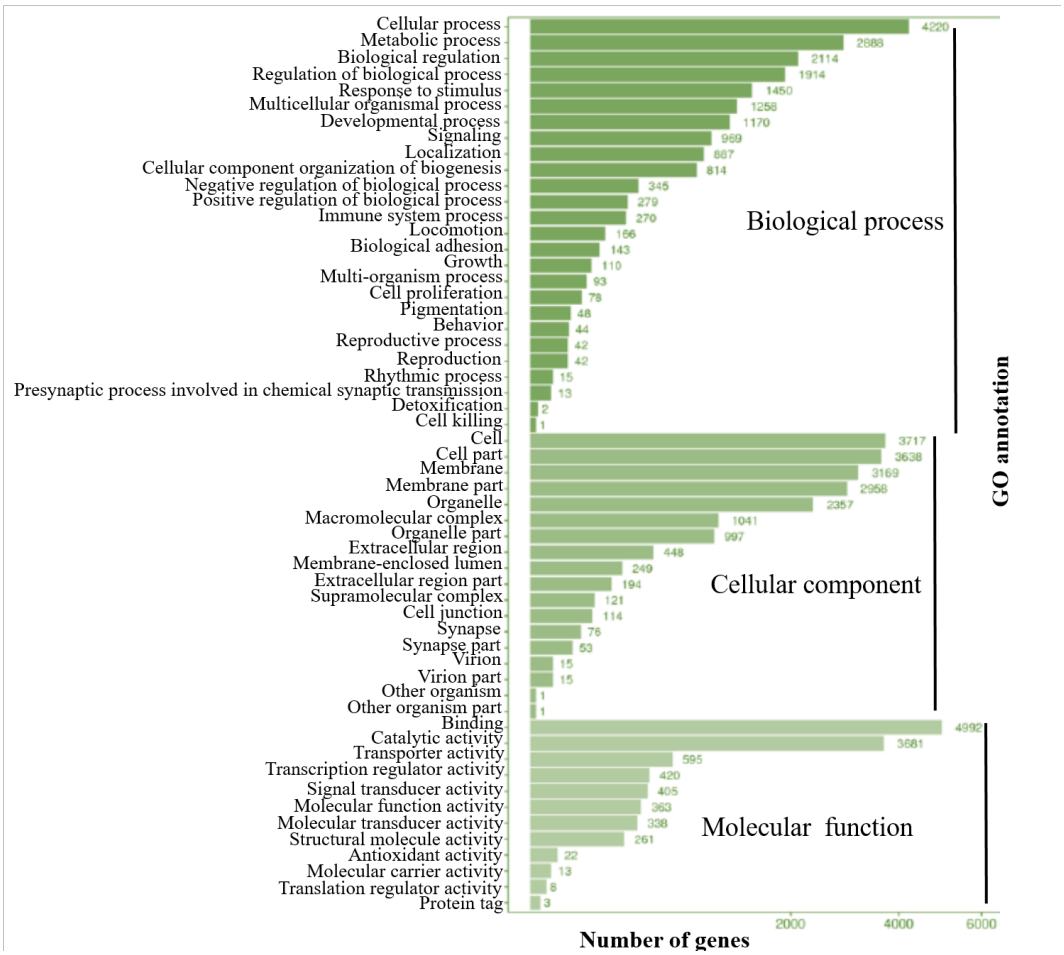

Fig. 2. Functional distribution of Gene Ontology (GO) functional categories representing biological processes, cellular components, and molecular functions in *O. uwai*.

Examples of predominant genes in males included testis-expressed protein 2, male-specific lethal 1 homolog isoform, and dual-specificity testis-specific protein kinase 1 for Nr and Nt, testis-expressed protein 2 and spermatogenesis- associated protein 1 for UniProtKB/Swiss-Prot, and sperm acrosome membrane-associated protein for KEGG. In females, oocyte zinc finger protein, choriogenin H, and zona pellucida sperm-binding protein dominated the genes for Nr and Nt. Vitellogenin and zona pellucida sperm- binding protein were predominant in females for UniProtKB/Swiss-Prot. Oocyte zinc finger protein and zona pellucida sperm-binding protein were detected as predominant genes in females for the KEGG annotation. Moreover, there were 70 TF family classifications for unigenes. The "zinc finger C2H2 transcription factors" (zf-C2H2) group was attributed to the highest number of genes in this family classification. In contrast, only one gene was found in the "nuclear transcription factor Y subunit beta" (NF-YB) and "lysosomal transcription factor" (NCU-G1) groups (Fig. 4).

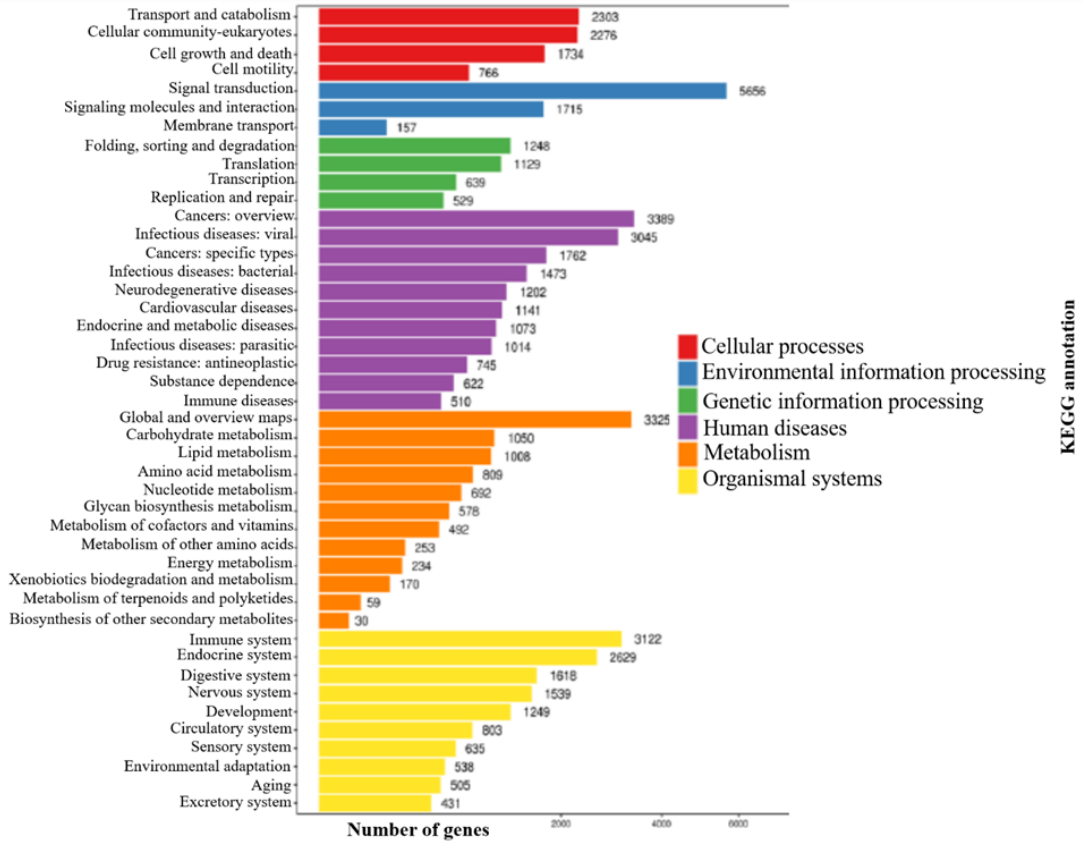

Fig. 3. Annotation of unigene distribution in *O. uwai* using Kyoto Encyclopedia of Genes and Genomes (KEGG).

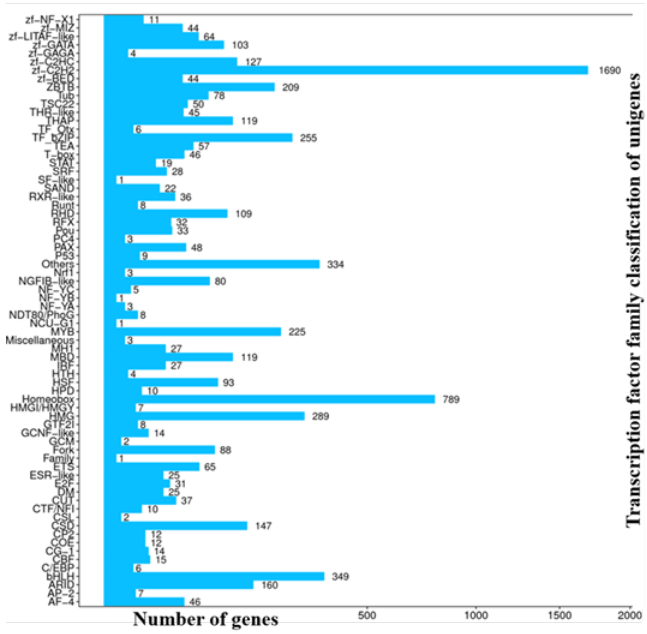

Fig. 4. Transcription-factor family classification of unigenes in *O. uwai*.

For the prediction of unigene coding DNA sequences (CDS), the number of CDSs, total length, maximal and minimal length, and GC content were 30,132, 29,319,330, 11,370, 297, and 54.36%, respectively (Fig. 5A). For the detection of simple sequence repeats (SSRs) in unigenes, 20 types of repeated nucleotides were found (Fig. 5B). For single-nucleotide polymorphism (SNP) detection, the numbers of A/G and C/T variants were 34,595 and 34,181, respectively, with 68,776 A/G and C/T variants in total. The number of A/C, A/T, C/G, and G/T variants were 11,362, 11,766, 8070, and 11,711, respectively. The total number of A/C, A/T, C/G, and G/T variants combined was 42,909.

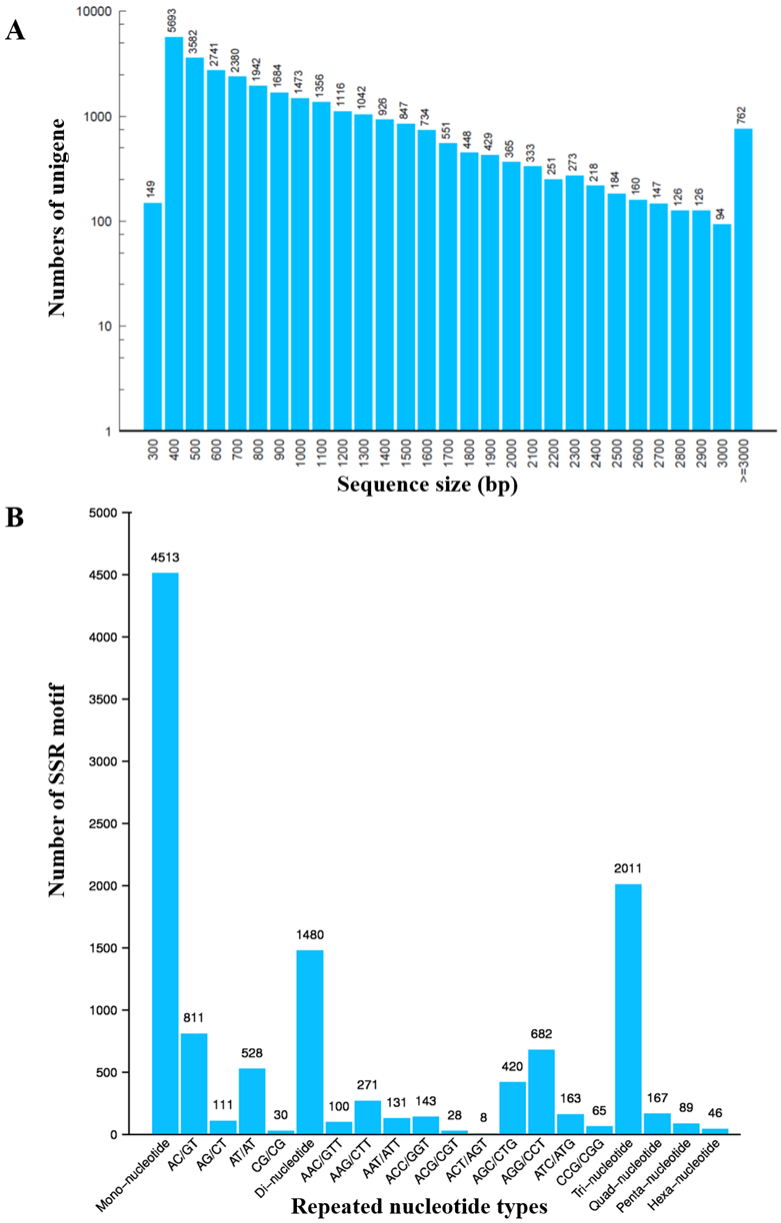

Fig. 5. Coding-sequence length distribution of unigene number with sequence size (A) and simple sequence repeat (SSR) size distribution of *O. uwai* (B).

Within the *O. woworae* species, Mokodongan et al. [28] reported the succession of phylotranscriptomic analysis from the eyeballs of wild *O. asinua*, *O. wolasi*, and *O. woworae* associated with body shape and geographical variations. *O. uwai* is distributed in various regions of natural freshwater in Myanmar, with apparent geographical differences [1, 22]. It was suggested that transcriptomic data are necessary for further examination, and a potential tool for creating a molecular tree on the basis of RNA sequencing of wild *O. uwai* populations from diverse geographical regions. The Nr distribution of *O. uwai* annotated to Japanese medaka is almost 77%. Recently, Ngamiyom et al. [21] reported that the Nr distribution of *O. minulluus* (Thai medaka) is 69.1% according to the database of Japanese medaka. In *O. uwai*, protein prediction strongly supports the existence of conserved proteins within *Oryzias*, and these results fill a gap in the data of molecular studies on medaka. Many gene functions of *O. uwai* are consistent with those of Thai medaka, as reported in a previous study by Ngamiyom et al. [21]. According to the NCBI Nucleotide database, the mitochondrial cytochrome c oxidase subunit I (COI) of *O. uwai* shares the closed identity with the COI of Thai medaka compared with that of other *Oryzias* spp. [2, 23, 29]. The morphology of *O. uwai* is quite similar to that of Thai medaka; the main morphological difference is that *O. uwai* has melanated lines on the pelvic fin, whereas Thai medaka lacks these markings [1]. In addition, the environmental habitat of *O. uwai* is not accessible to Thai medaka [1, 22]. These results suggest that the molecular biological variation is congruent with morphological and geographical variation among *Oryzias* species. On the basis of annotated genes, the general transcriptome of *O. uwai* is consistent with that of other *Oryzias* species and many others reported for freshwater fish, including zebrafish, *Cyprinus carpio* (common carp), and *Poecilia reticulata* (guppy) [15, 20, 21, 28, 30-32]. In Japanese medaka, the DM-domain gene on the Y chromosome (DMY) is the Y-specific DM domain of a sex-determining gene that plays a crucial role in male development [33]. DMY is absent in other *Oryzias* spp. [34]. In this study, DMY was not found in *O. uwai* in any of the functional databases. In the biological-process category, pigmentation unigenes were fewer in *O. uwai* than in the Java medaka species group, as reported in the excellent study by Takehana et al. [15]. Java medaka has a predominant orange- yellow band on the caudal fin [1]. However, *O. uwai* is not colorful and lacks this orange- yellow band on the caudal fin [22]. The number of unigenes may also correspond to morphological color. In males, dual-specificity testis-specific protein kinase 1, spermatogenesis-associated protein 1, and sperm acrosome membrane-associated protein play a role in testis or sperm development [35-37]. In females, choriogenin and vitellogenin are known to be important genes for oogenesis [38], and zona pellucida sperm-binding protein is necessary for the egg surface's interaction with sperm [39]. Therefore, the genes of *O. uwai* might correspond to gonad-specific organs of males and females. In TF family classification, zf-C2H2 was found to be associated with a large number of unigenes in *O. uwai*. zf-C2H2 plays an important role in the regulation of gene expression [40]. Schep and Adryan [41] reported that zf-C2H2 is found in various metazoans, including zebrafish. The zinc finger proteins are widely binding to regulatory regions for gene expression that play crucial roles in biological functions of cells [42-44]. Therefore, this result confirmed that zf-C2H2 might be an important protein involved in gene regulation of this fish species. However, at present, the precise function of zf-C2H2 in adult *O. uwai* remains unknown. This zinc finger may be the main transcription factor for regulating gene expression in adult *O. uwai*.

In addition, research by Yamahira et al. [45] clarified the genome of ten *Oryzias* spp. including *O. uwai,* which was utilized to understand the geobiology of Adrianichthyidae via molecular genetics. In this study, therefore, the present data might fill the gap of knowledge on the molecular biology in *O. uwai*. RNA sequencing data of *O. uwai* may be crucial for advancing transcriptomic analysis in *Oryzias* fish.

4. Conclusion

To the present body of knowledge, we contribute data from the RNA transcriptome assembly of adult *O. uwai*. These data may provide insights that increase our understanding of evolutionary gene function in freshwater fish.

Acknowledgments

This research was funded by the research grant of Srinakharinwirot University no. 011/2563.

References

[1] Parenti LR. A phylogenetic analysis and taxonomic revision of ricefishes, *Oryzias* and relatives (Beloniformes, Adrianichthyidae). Zool J Linn Soc 2008;154: 494-610.

[2] Setiamarga DH, Miya M, Yamanoue Y, Azuma Y, Inoue JG, Ishiguro NB, Mabuchi K, Nishida M. Divergence time of the two regional medaka populations in Japan as a new time scale for comparative genomics of vertebrates. Biol Lett 2009;5(6): 812-6.

[3] Parenti LR, Hadiaty RK, Lumbantobing D, Herder F. Two New Ricefishes of the Genus *Oryzias* (Atherinomorpha: Beloniformes: Adrianichthyidae) Augment the Endemic Freshwater Fish Fauna of Southeastern Sulawesi, Indonesia. Copeia 2013;3: 403-14.

[4] Hilgers L, Schwarzer J. The untapped potential of medaka and its wild relatives. Elife 2019; p.e46994.

[5] Nagahama Y. Molecular mechanisms of sex determination and gonadal sex differentiation in fish. Fish Physiol Biochem 2005;31(2-3): 105-9.

[6] Takeda H, Shimada A. The art of medaka genetics and genomics: what makes them so unique? Annu Rev Genet 2010;44: 217-41.

[7] Matsuda M, Sakaizumi M. Evolution of the sex-determining gene in the teleostean genus *Oryzias*. Gen Comp Endocrinol 2016;239:80-8.

[8] Ishibashi H, Uchida M, Temma Y, Hirano M, Tominaga N, Arizono K. Choriogenin transcription in medaka embryos and larvae as an alternative model for screening estrogenic endocrine-disrupting chemicals. Ecotoxicol Environ Saf 2020;193: 110324.

[9] Shima A, Mitani H. Medaka as a research organism: past, present and future. Mech Dev 2004;121(7-8): 599-604.

[10] Lin CY, Chiang CY, Tsai HJ. Zebrafish and Medaka: new model organisms for modern biomedical research. J Biomed Sci 2016;23:19.

[11] Hayasaka O, Takeuchi Y, Shiozaki K, Anraku K, Kotani T. Green light irradiation during sex differentiation induces female-to-male sex reversal in the medaka *Oryzias latipes*. Sci Rep 2019;9: 2383.

[12] Bajoghli B, Dick AM, Claasen A, Doll L, Aghaallaei N. Zebrafish and Medaka: Two Teleost Models of T-Cell and Thymic Development. Int J Mol Sci 2019;20(17): 4179.

[13] Koyama J, Kawamata M, Imai S, Fukunaga M, Uno, S, Kakuno A. Java medaka: a proposed new marine test fish for ecotoxicology. Environ. Toxicol 2008;23(4): 487-91.

[14] Kim BM, Kim J, Choi IY, Raisuddin S, Au DW, Leung KM, Wu RS, Rhee JS, Lee JS.

Omics of the marine medaka (*Oryzias melastigma*) and its relevance to marine environmental research. Mar Environ Res 2016;113: 141-52.

[15] 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. Genome Sequence of the Euryhaline Java fish Medaka, *Oryzias javanicus*: A Small Aquarium Fish Model for Studies on Adaptation to Salinity. G3 (Bethesda) 2020;10(3): 907-15.

[16] Irie N, Kuratani S. Comparative transcriptome analysis reveals vertebrate phylotypic period during organogenesis. Nat Commun 2011;2: 248.

[17] Gerstein MB, Rozowsky J, Yan KK, Wang D, Cheng C, Brown JB, Davis CA, Hillier L, Sisu C, Li JJ, Pei B, Harmanci AO, Duff MO, Djebali S, Alexander RP, Alver BH, Auerbach R, Bell K, Bickel PJ, Boeck ME, Boley NP, Booth BW, Cherbas L., Cherbas P, Di C, Dobin A, Drenkow J, Ewing B, Fang G, Fastuca M, Feingold EA, Frankish A, Gao G, Good PJ, Guigó R, Hammonds A, Harrow J, Hoskins RA, Howald C, Hu L, Huang H, Hubbard TJ, Huynh C, Jha S, Kasper D, Kato M, Kaufman TC, Kitchen RR, Ladewig E, Lagarde J, Lai E, Leng J, Lu Z, MacCoss M, May G, McWhirter R, Merrihew G, Miller DM, Mortazavi A, Murad R, Oliver B, Olson S, Park PJ, Pazin MJ, Perrimon N, Pervouchine D, Reinke V, Reymond A, Robinson G, Samsonova A, Saunders GI, Schlesinger F, Sethi A, Slack FJ, Spencer WC, Stoiber MH, Strasbourger P, Tanzer A, Thompson OA, Wan KH, Wang G, Wang H, Watkins KL, Wen J, Wen K, Xue C, Yang L, Yip K, Zaleski C, Zhang Y, Zheng H, Brenner SE, Graveley BR, Celniker SE, Gingeras TR, Waterston R. Comparative analysis of the transcriptome across distant species. Nature 2014;512(7515): 445-8.

[18] Qian X, Ba Y, Zhuang Q, Zhong G. RNA-Seq technology and its application in fish transcriptomics. OMICS 2014;18(2): 98-110.

[19] Sun Y, Huang Y, Li X, Baldwin CC, Zhou Z, Yan Z, Crandall KA, Zhang Y, Zhao X, Wang M, Wong A, Fang C, Zhang X, Huang H, Lopez JV, Kilfoyle K, Zhang Y, Ortí G, Venkatesh B, Shi Q. Fish- T1K (Transcriptomes of 1,000 Fishes) Project: large- scale transcriptome data for fish evolution studies. Gigascience 2016;5: 18.

[20] Lai KP, Li JW, Wang SY, Chiu JM, Tse A, Lau K, Lok S, Au DW, Tse WK, Wong CK, Chan TF, Kong RY, Wu RS. Tissue-specific transcriptome assemblies of the marine medaka *Oryzias melastigma* and comparative analysis with the freshwater medaka *Oryzias latipes*. BMC Genet. 2015;16(1): 135.

[21] Ngamniyom A, Sriyapai T. Sriyapai P. Molecular analysis of population and *de novo* transcriptome sequencing of Thai medaka, *Oryzias minutillus* (Teleostei: Adrianichthyidae). Heliyon 2020;6(1) : e03079.

[22] Roberts TR. Systematic observations on tropical Asian medakas or ricefishes of the genus *Oryzias*, with descriptions of four new species. Ichthyol Res 1998;45: 213-24.

[23] Kano Y, Musikasinthorn P, Iwata A, Tun S, Yun L, Win SS, Matsui S, Tabata R, Yamasaki T, Watanabe KA. Dataset of fishes in and around Inle Lake, an ancient lake of Myanmar, with DNA barcoding, photo images and CT/3D models. Biodivers Data J 2016; 4:e10539.

[24] Chen Y, Chen Y, Shi C, Huang Z, Zhang Y, Li S, Li Y, Ye J, Yu C, Li Z, Zhang X, Wang J, Yang H, Fang L, Chen Q. SOAPnuke: a MapReduce acceleration-supported software for integrated quality control and preprocessing of high-throughput sequencing data. Gigascience 2018;7(1): 1-6.

[25] Pertea G, Huang X, Liang F, Antonescu V, Sultana R, Karamycheva S, Lee Y, White J, Cheung F, Parvizi B, Tsai J, Quackenbush J. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. Bioinformatics 2003;19(5):651-2.

[26] Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet 2000;16(6): 276-7.

[27] Mistry J, Finn RD, Eddym SR, Bateman A, Punta M. Challenges in homology search: HMMER3 and convergent evolution of coiled-coil regions. Nucleic Acids Res 2013;41(12): e121.

[28] Mokodongan DF, Montenegro J, Mochida K, Fujimoto S, Ishikawa A, Kakioka R, Yong L, Mulis, Hadiaty RK, Mandagi IF, Masengi KWA, Wachi N, Hashiguchi Y, Kitano J, Yamahira K. Phylogenomics reveals habitat-associated body shape divergence in *Oryzias woworae* species group (Teleostei: Adrianichthyidae). Mol Phylogenet Evol 2018;118: 194-203.

[29] Sayers EW, Barrett T, Benson DA, Bolton E, Bryant SH, Canese K, Chetvernin V, Church DM, Dicuccio M, Federhen S, Feolo M, Fingerman IM, Geer LY, Helmberg W, Kapustin Y, Krasnov S, Landsman D, Lipman DJ, Lu Z, Madden TL, Madej T, Maglott DR, Marchler-Bauer A, Miller V, Karsch-Mizrachi I, Ostell J, Panchenko A, Phan L, Pruitt KD, Schuler GD, Sequeira E, Sherry ST, Shumway M, Sirotkin K, Slotta D, Souvorov A, Starchenko G, Tatusova TA, Wagner L, Wang Y, Wilbur WJ, Yaschenko E, Ye J. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2012; 40 (Database issue):D13-25.

[30] Fraser BA, Weadick CJ, Janowitz I, Rodd FH, Hughes KA. Sequencing and characterization of the guppy (*Poecilia reticulata*) transcriptome. BMC Genet 2011; 12:202.

[31] Krabbenhoft TJ, Turner TF. Comparative transcriptomics of cyprinid minnows and carp in a common wild setting: a resource for ecological genomics in freshwater communities. DNA Res 2018;25(1):11-23.

[32] Mehjabin R, Xiong L, Huang R, Yang C, Chen G, He L, Liao L, Zhu Z, Wang Y. Full-Length Transcriptome Sequencing and the Discovery of New Transcripts in the Unfertilized Eggs of Zebrafish (*Danio rerio*). G3 (Bethesda) 2019;9(6):1831-8.

[33] Otake H, Shinomiya A, Kawaguchi A, Hamaguchi S, Sakaizumi M. The medaka sex-determining gene DMY acquired a novel temporal expression pattern after duplication of DMRT1. Genesis 2008;46(12):719-23.

[34] Nagai T, Takehana Y, Hamaguchi S, Sakaizumi M. Identification of the sex-determining locus in the Thai medaka, *Oryzias minutillus*. Cytogenet Genome Res 2008;121(2):137-42.

[35] Toshima J, Ohashi K, Okano I, Nunoue K, Kishioka M, Kuma K, Miyata T, Hirai M, Baba T, Mizuno K. Identification and characterization of a novel protein kinase, TESK1, specifically expressed in testicular germ cells. Int J Biol Chem 1995;270(52):31331-7.

[36] Giesecke K, Hamann H, Stock KF, Woehlke A, Sieme H, Distl O. Evaluation of SPATA1-associated markers for stallion fertility. Anim Genet 2009;40(4):359-65.

[37] Hao Z, Wolkowicz MJ, Shetty J, Klotz K, Bolling L, Sen B, Westbrook VA, Coonrod S, Flickinger CJ, Herr JC. SAMP32, a testis- specific, isoantigenic sperm acrosomal membrane-associated protein. Biol Reprod 2002;66(3): 735-44.

[38] Ishibashi H, Uchida M, Koyanagi A, Kagami Y, Kusano T, Nakao A, Yamamoto R, Ichikawa N, Tominaga N, Ishibashi Y, Arizono K. Gene expression analyses of vitellogenin, choriogenin and

estrogen receptor subtypes in the livers of male medaka (*Oryzias latipes*) exposed to equine estrogens. J Appl Toxicol 2016; 36(11):1392-400.

[39] Wassarman PM. Zona pellucida glycoproteins. J Biol Chem 2008;283(36): 24285-9.

[40] Böhm S, Frishman D, Mewes HW. Variations of the C2H2 zinc finger motif in the yeast genome and classification of yeast zinc finger proteins. Nucleic Acids Res 1997;25(12):2464-69.

[41] Schep AN, Adryan BA. Comparative analysis of transcription factor expression during metazoan embryonic development. PLoS One 2013;8(6): e66826.

[42] Najafabadi HS, Mnaimneh S, Schmitges FW, Garton M, Lam KN, Yang A, Albu M, Weirauch MT, Radovani E, Kim PM, Greenblatt J, Frey BJ, Hughes TR. C2H2 zinc finger proteins greatly expand the human regulatory lexicon. Nat Biotechnol 2015; 33(5):555-62.

[43] Han BY, Foo CS, Wu S, Cyster JG. The C2H2- ZF transcription factor Zfp335 recognizes two consensus motifs using separate zinc finger arrays. Genes Dev 2016;30(13): 1509-14.

[44] Aizenshtein-Gazit S, Orenstein Y. DeepZF: improved DNA-binding prediction of C2H2-zinc-finger proteins by deep transfer learning. Bioinformatics. 2022;38: ii62-7.

[45] Yamahira K, Ansai S, Kakioka R, Yaguchi H, Kon T, Montenegro J, Kobayashi H, Fujimoto S, Kimura R, Takehana Y, Setiamarga DHE, Takami Y, Tanaka R, Maeda K, Tran HD, Koizumi N, Morioka S, Bounsong V, Watanabe K, Musikasinthorn P, Tun S, Yun LKC, Masengi KWA, Anoop VK, Raghavan R, Kitano J. Mesozoic origin and 'out-of India' radiation of ricefishes (Adrianichthyidae). Biol Lett 2021;17(8): 20210212.