Thesis for the Degree of Master of Science

Identification of sex-biased genes through gonadal transcriptome analysis in marine medaka

by

Tuluğ Gülce Ataman

Department of Fisheries Biology

The Graduate School

Pukyong National University

February 2019

Identification of sex-biased genes through gonadal transcriptome analysis in marine medaka

바다송사리(*Oryzias dancena*)로 부터 생식소 전사체 분석을 통한 암수 특이적 발현 유전자 탐색

Advisor: Prof. Yoon Kwon Nam

by

Tuluğ Gülce Ataman

A thesis submitted in partial fulfillment of the requirements for the degree of

Master of Science

in Department of Fisheries Biology, The Graduate School,

Pukyong National University

February 2019

Identification of sex-biased genes through gonadal transcriptome analysis in marine medaka

A dissertation

by

Tuluğ Gülce Ataman

Approved by:

(Chairman) Dr. Dong Soo Kim

(Member) Dr. Seung Pyo Gong

(Member) Dr. Yoon Kwon Nam

Table of Contents

List of Tables

List of Figures

Identification of sex-biased genes through gonadal transcriptome analysis in marine medaka

Tuluğ Gülce Ataman

Department of Fisheries Biology, The Graduate School,

Pukyong National University

ABSTRACT

Marine medaka; *Oryzias dancena* is an important species in coastal areas of Asian distribution which has a potential of being an ideal vertebrate model for biological development and comparative genomic studies. Therewithal, increasing genomic studies feasible with aquaculture, generates a need for further insight especially in sex-related mechanisms such as; gametogenesis, gonadal development and sex determination/differentiation process in these species. However, for this species, the recent studies about the molecular process related with the sex-biased mechanisms has not been yet extensively studied. By using Illumina RNA-Seq technology, high-quality reads from the cDNA libraries of ovary (58,412,985), testis (58,605,939) and muscle (67,583,838), were assembled into 369,054 unigenes with a successfully annotated sequence rate with 16,548 transcripts. Identification of the candidate

genes were chosen through this transcriptome database analysis by using published literature database, trimmed mean of M values matrix (TMM) and differentially expressed genes (DEG's) obtained from transcriptome analysis. Differential expression analysis generated 23,497 differentially expressed genes through the comparison of ovary and testis. Also muscle transcriptome against gonadal transcriptome (ovary-testis) analysis of differential expression was estimated as 29,556 genes in total, which gave us a wide range of potential sex-biased sequence database. Sex biased genes were analyzed and compared through categorizing them as sex-enriched genes, sex-specific genes, gonad-specific genes, TMM and differentially expressed genes in gonadal transcriptome. The aim of this thesis is to generate a comprehensive list of candidate sex-biased genes that are differentially expressed between gonads which will provide a better understanding of tissue-specific genes from both sexes by further investigations for functional studies.

바다송사리*(Oryzias dancena)*로 부터 생식소 전사체 분석을 통한 암수 특이적 발현 유전자 탐색

툴루 귤제 아타만

부경 대학교 대학원 수산 생물학과

요약

바다 송사리 (*Oryzias dancena*)는 아시아 분포의 해안 지역에서 서식하며, 생물학적 발달과 게놈 연구를위한 이상적 인 척추 동물로서의 잠재력을 가진 중요한 종이다. 그에 따라 양식과 관련한 게놈 연구가 증가하고 있으나 특히 성 관련 메커니즘 생식 발생 생식 샘 발달 및 성 결정과 분화 과정과 같은 추가적 인 이해가 필요하다. 그러나,이 종에 대해 생식선, 성 결정 및 배우자 형성의 발달 과정에 관여하는 성 편향적인 메커니즘과 관련된 분자 과정에 리) 관한 최근의 연구로는 부족하다. *O. dancena* 생식선에서 유라 된 DNA 이브러리의. 난소 (58,412,985)7 의 cDNA 라이브러리 Illumina RNA-Seq 기출을 사용하여 고환율 (58,605,939) 및 근육 (67,583,838)을 성공적으로 측정 하였다. 난소와 고환의 비교를 통해 23,497 개의 차별적으로 발현 하는 유전자를 선 별했다. 또한 근육 전사체와 미성숙 전사체 (난소-고환) 발현 차이는

총 29,556 개의 유전자로 잠재적으로 성별에 편차가 있는.서열 데이터베이스가 광범위하다는 것을 알 수 있다. 본 논문의 결과는 배우자 형성, 생식기의 발달 및 성 분화에 관여하는 성 편향적인 후보 유전자의 포괄적인 목록을 제공함으로써 향후 추가적인 기능적 연구를 위한 유용한 기초 자료를 제공할 수 있다고 기대된다.

1.INTRODUCTION

Marine medaka, *Oryzias dancena*, is an euryhaline teleost with an increasing value in aquaculture and molecular studies according to their physiological characteristics which can be explained with the interval between generations, daily spawners and spawning possibilities just 60 days after hatching (Inoue & Takei, 2003). Also they are considered to be an important model organism for transgenic studies because of their embryo transparency which makes them favorable for tracing gene expression using Green fluorescent protein (Cho et al., 2011). Studies in recent years made by *Oryzias* species have come into prominence with transgenic studies (Tanaka et al., 2001; Cho & Nam, 2016).

Recent researches using tissue-specific targeting of the transgene expression in zebrafish by using fluorescent proteins certified that the expression of recombinant proteins in muscle tissue from a fish doesn't have a negative effect of the endogenous mylz2 mRNA which is considered as a favorable feature for bioreactor studies (Gong et al., 2003).

Gene expression of temporal and spatial and variation carries information about genes function (Bassett et al., 1999). The tissue-specific genes are defined as the genes whose function and expression are limited to a cell type or a specific tissue. Tissue specificity definition extended by tissue selectivity in which the gene expression is enriched in one or more cell types or tissues (Xiao et al., 2010).

Latest experiments related with sex-specific- and tissue-specific promoter methylation in *O. latipes* (Japanese medaka) demonstrated that changes in

methylation may affect the regulation of normal gene expression and the transmission of these to offspring changes is possible. This indicates that tissue specific expressions in gonadal tissues are crucial and can be lead to maternal transmission of different changes occurred in tissues (Contractor et al., 2004).

In *O. dancena*, the sex is genetically determined at the time of fertilization. The males are heterogametic with XY chromosome composition, while females are of XX chromosome composition (Shibata et al., 2010).

According to the Darwinian evolution; divergence between each species is crucial and one of the doctrines that is leading evolutionary biology occurs through sexual reproduction which causes variation between two sexes within differences amongst morphological, behavioral and physiological traits (Grath & Parsch, 2016). In theory, evolution of both sexes should be subject to the same forces that causes the natural selection, sexual selection and genetic drift (Parsch & Ellegren, 2013).

Some of the causes of sex-biased expression can be explained, with sexual antagonism, gene duplication, dosage compensation and expression of sex-limited chromosomes. Recent studies have shown that, in addition to gene-specific processes such as regulatory element evolution and gene duplication, chromosome-wide processes such as dosage compensation have an important role in shaping sex-biased gene expression (Parsch & Ellegren, 2013).

Latest studies show that distinctness between individuals of the same sex has lesser differences comparing to the male and female individuals. Regarding to these components; sex-biased gene expression can be explained through morphological differences between two sexes which caused by the differentially expressed genes that are present in both male and female in wide variety of tissues of each individuals. Differentially expressed genes between both sexes causes most of the sexually

dimorphic characteristics and these genes with sexually dimorphic expressions are defined as sex-biased genes. These differentially expressed genes can be expressed in both sexes while the expression level can be higher than the other sex which is referred as sex-enriched genes. Also genes may be expressed in only one sex can be referred as sex-specific genes. These genes are also can be divided in two definitions as male-biased and female-biased. The genes that are expressed equally in both sexes are defined as sex-biased genes (Ellegren & Parsch, 2007). However, a gene's sex bias is not fixed and it can vary among tissues or change within different developmental stages (Parsch & Ellegren, 2013).

Recent advances in genomics, such as RNA sequencing (RNA-seq), have revealed the nature and extent of sex-biased gene expression in diverse species (Parsch & Ellegren, 2013). RNA sequencing (RNA-Seq) ensures the comparison of the expression levels within thousands of genes between samples (in multiple tissues within an organism) which reveals that sex had greater influence on the divergence on the gene expression more than genotypical and age related components (Jin et al., 2001).

Increasing amount of studies from various species on sex related genes through large scale of gonadal transcriptome analysis includes remarkable information that can be useful for further functional and developmental researches on several disciplines.

The first whole genome sequenced by an early next-generation sequencing technology was the GS 20 (Margulies et al., 2005). Next-generation sequencing' likely to sanger sequencing, has constant steps of nucleotide incorporation, followed by a detection step where the output, either light or pH indifference is being detected and afterwards it comes into elutriation phase where blocking terminators are removed. However, these steps are performed in parallel on millions of DNA fragments. What made possible this

parallelization of the process was the invention of flow cells, where the different DNA fragments are spatially separated.

Prior to loading the DNA fragments onto the sequencing machine, the DNA is converted into a library of DNA fragments. During this process the DNA is fragmented and a set of platform specific adaptors are ligated on to the DNA fragments. Once the library is loaded, the fragments attach to the complementary adaptors that are on the flow cell. Then the fragments are amplified in situ on the flow cell. This amplification step is needed to provide sufficient signal during each DNA reaction step. Currently, four next-generation sequencing platforms are available: Roche 454, Illumina, SOLiD and Ion Torrent. Which of these platforms differ in their, sequencing approaches, amplification step and output (Pillai et al., 2017). Among all these platforms Illumina is the most widely used platform.

Recent studies made in two commercially important species; Atlantic halibut (*H. hippoglossus*) and Guppy fish (*P. reticulate*) for the purpose of enlightening mechanism behind sex-biased genes states that gonads between two sexes have significant difference regarding their expression levels on different genes and miRNA expressions (Bizuayehu et al., 2012; Qian et. al, 2014).

There are studies for the purpose of enlightening sex-biased mechanisms in *O. dancena* using gonadal tissues pointing out Sox3 gene as the male-determining factor which highly expressed in male gonads (Takehana et. al, 2014) and a novel Choriogenin H Isoform (*odChgH*) which was also highly expressed in ovary tissues (Lee et. al, 2012). One of the latest expression profiling in *Oryzias* species (*O. melastigma*) through Illumina RNA sequencing (RNA-Seq) for miRNA transcriptome analysis in brain, liver, and

gonads from sexually mature male and female suggests different expression patterns between gonads of each sexes (Lau et al., 2014). However, there is a need to identify a wide range of database for sex-biased genes in *O. dancena*.

By using Illumina RNA-Seq technology, high-quality reads from the cDNA libraries of ovary (58,412,985), testis (58,605,939) and muscle (67,583,838), were assembled into 369,054 unigenes and muscle transcriptome analysis was used as a control group to be able to define gonadal transcriptome analysis.

In brief, on the basis of these findings indicates that, tissue-specific patterns of gene expression are fundamental to establishing and preserving tissue identity and function. Study of sex-biased gene expression through RNA-Seq analysis should provide a wide range of data that, in conjunction with genetic and epigenetic data, will help elucidate the regulatory mechanisms controlling sex-biased expression. This study will provide a high range of data of sex-biased genes in *O. dancena* by the information of their differential expression and tissue-specifity, later to be used for developing tissue specific promoters for recombinant protein production.

2.MATERIALS AND METHODS

2.1 Ethics Statement, Experimental species and sample collection

Oryzias dancena was maintained in recirculating aquarium tanks equipped with sand filtered salt water with 10 per mill (ppt) Salinity at 25 ± 1 C° with continuous aeration and natural photoperiod at the Institute of Marine Living Modified Organism, Pukyong National University, Busan, Korea. *O. dancena* were fed with a commercial diet 3 times a day. After the sexual maturation occurred testis, ovary and muscle tissues were dissected out from medaka males and medaka females that were euthanized with an overdose of tricaine methanesulfonate (MS-222) and unfertilized eggs were collected from five females, of which the abdomens were pressed gently to discharge the eggs into glass dish. The collected sample were immediately frozen and stored at -80 prior to RNA extraction. All animal research procedures were approved by the Animal Ethics Committee of Pukyong National University and performed according to the guidelines for the care and use of laboratory animals.

2.2 RNA Isolation, Library Construction and Illumina Sequencing

Total RNA was extracted by using RNeasy Plus Mini Kit (Qiagen, Hilden, Germany) following the manufacturer's instructions. The concentration and quality of each RNA sample were examined using a NanoDrop-2000 spectrophotometer (Thermo Scientific, Waltham, MA, USA), and the integrity of RNA was checked by ethidium bromide staining of 28S and 18S ribosomal bands on a 1% MOPS formaldehyde agarose gel. Also the quality and quantity of RNA-seq library was confirmed by using a Bioanalyzer 2100 (Agilent Technologies) (Figure 1).

Sequencing and construction of the library was carried out by DNA link Inc. cooperation in Korea. (http://www.dnalink.com/korean/index.html). The

cDNA library construction performed by using muscle, testis and ovary tissues of *Oryzias dancena* was sequenced by using Illumina HiSeq 2500 platform and paired-end reads were generated. Equal amounts of high-quality RNA samples of each tissue were used to synthesize a cDNA library. Briefly, mRNA was purified from total RNA and used as templates to synthesize the first-strand and the second-strand cDNA, according to the protocol of TruSeq RNA sample preparation kit v.2 (Illumina, San Diego, USA) and cDNA was cut into short fragments following the TruSeq RNA sample preparation guide. After the process of end repair and the addition of poly(A), the short fragments were ligated with sequencing adapters and enriched by PCR amplification to construct the cDNA library templates. Finally, the library was inserted into the channels of the Illumina HiSeq™ 2500 for sequencing (Figure 1).

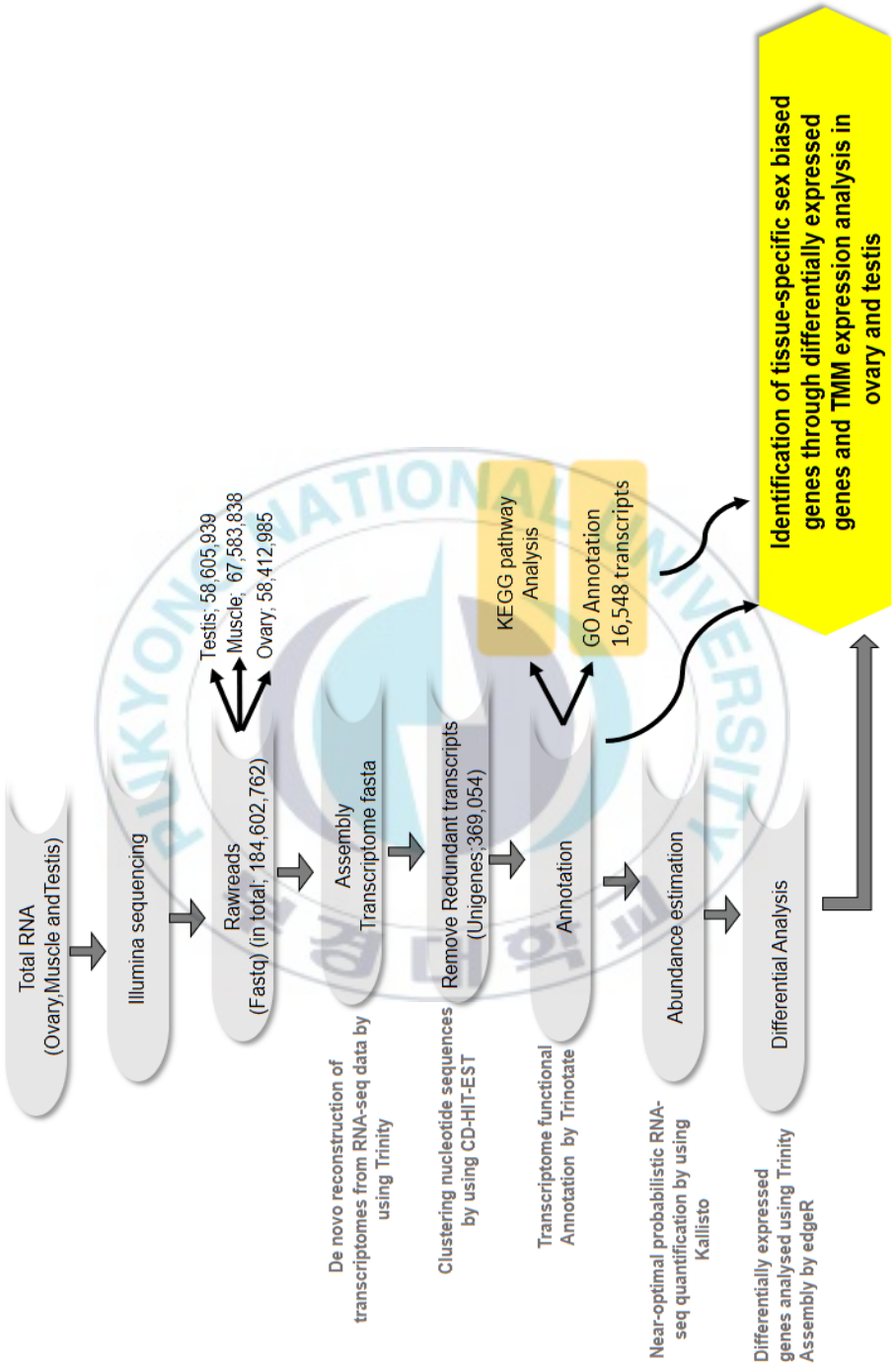

Figure 1: Experimental flowchart from RNAseq analysis of *O.dancena* ovary, muscle and testis transcriptome analysis.

2.3 Illumina Read processing and De novo assembly

The raw reads obtained by Illumina sequencing were processed to be able to obtain clean reads by removing; low quality bases in both end of the reads (N ratio > 5 %), unknown nucleotides and indexing adaptors by using Trimmomatic v0.32 tool. After the trimming process, FastQC v0.10.1 program was used by running Perl program quality of the remaining raw reads were verified to be able to find out any impurities in the raw reads to be able to obtain clean reads which was later stored in FASTQ format.

De novo transcriptome assembly of clean reads of all tissue libraries was aggregated to be able to constitute a reference genome by using the Trinity Assembler v2.2.0 which is defined as a software series that assemblies the transcriptomes by using short reads. Following the assembling process, the transcripts from ovary and testis libraries were merged, clustered and afterwards, the duplicates were removed by using CD-HITest (v4.6.1).

Bench-marking universal single-copy orthologues (BUSCO) analysis was used to obtain quantitative assessment of the annotation.

2.4 Functional assignment, Ontology analysis and annotation of the transcripts

Determining the functional annotation of de novo assembled transcriptomes were analyzed by Trinotate software through Trinity suite (Grabherr et al., 2011). Trinotate includes multiple analysis through Blastp/Blastx against reference sequence databases (e-value < 10^{-6}). PFAM to find out the protein domains by using HMMER v3.101 (Finn et al. 2013). Annotations were made through Uniprot and eggNOG/GO Pathways database (Apweiler et al., 2004). Nonredundant transcripts aligned against NCBI database later to be processed by Blast2GO analysis to be able to determine related GO annotations of unigenes for analyzing in three different ontologies; biological process (BP), cellular component (CC) and molecular function (MF) (GO; http://www.geneontology.org/).

Comparison of the annotated sequences based on their sequence similarity was carried out through GO analysis. Functional classification and metabolic pathway analysis was performed using online annotation server KEGG (Kyoto Encyclopedia of Genes and Genomes) (https://www.genome.jp/kegg/kaas/). Prediction of ribosomal RNA was analyzed through RNAmmer. The prediction of open reading frames (ORFs) and Peptide coding regions was analyzed using TransDecoder scripts.

2.5 De novo assembly and Transcriptome sequencing

Construction of cDNA libraries derived from *Oryzias dancena* by using ovary, testis and muscle, which was carried out by using Illumina HiSeq2500. The total reads were estimated as 184,602,762 from 3 different cDNA libraries of ovary, testis and muscle. Total reads count in each stranded cDNA libraries

which are for muscle; 67,583,838, for ovary; 58,412,985 and for testis; 58,605,939.

Through raw reads, the total clean reads were obtained, amount of 645,13 transcript contigs (average) in total 254,962,138 assembled bases which were assembled from the clean reads through Trinity assembler (Grabherr et. al, 2011). These reads may be either paired-end or single-end but paired-end sequence data are preferred since they are able to guide more distant connections between regions of transcript isoforms during assembly (Haas et al., 2013).

Stats based on ALL transcript contigs	Contig N10	4,007
	Contig N20	2,792
	Contig N30	2,061
	Contig N40	1,499
	Contig N50	1,045
	Median contig length	341
	Average contig	645.13
	Total assembled bases	254,962,138
Total reads (Library type; 101x2 stranded)	Ovary	58,412,985
	Testis	58,412,985
	Muscle	67,583,838
Counts of transcripts	Total trinity genes	301,778
	Total trinity transcripts	395,212
	Percent GC	45,38
Remove redundant transcripts	Redundant genes	20,119
	Redundant transcripts	26,158
	Remain genes	281,659
	Remain transcripts	369,054

Table 1: Stats based on all transcript contigs

Sequencing (Library type; 101x2 stranded)	Ovary	Testis	Muscle
Total reads	58,412,985	58,605,939	67,583,838
Mapping rate	85%	91%	90%
Isoform Read count > 0	251,642	287,609	263,888
Isoform FPKM > 1	56,430	69,191	61,776

Table 2: Denovo transcriptome assembly and sequencing demographics.

(*Reference: Denovo transcriptome assembly results of 369,054 isoforms of O. dancena)

After the removal of 26,158 redundant transcripts and 20,119 redundant genes by using CD-HIT-EST tool, 369,054 non-redundant transcripts as 281,659 genes were obtained as a representative of total trinity genes in the number of 301,778 and GC ratio of 45.38 %.

To be able to avoid false-positive results caused by computational methods for estimating transcript abundance from RNA-seq data which are systematic errors stem largely from a failure to model fragment GC content deviation, we used Kallisto for quantifying abundances of transcripts from RNA-Seq data, or more generally of target sequences using high-throughput sequencing reads (Teng et al., 2016) (https://pachterlab.github.io/kallisto/).

Authenticity of the non-redundant transcripts were determined by using BUSCO v3.0.2 software by the comparing through eukaryote dataset as a reference to be able to obtain a reliable de novo assembles.

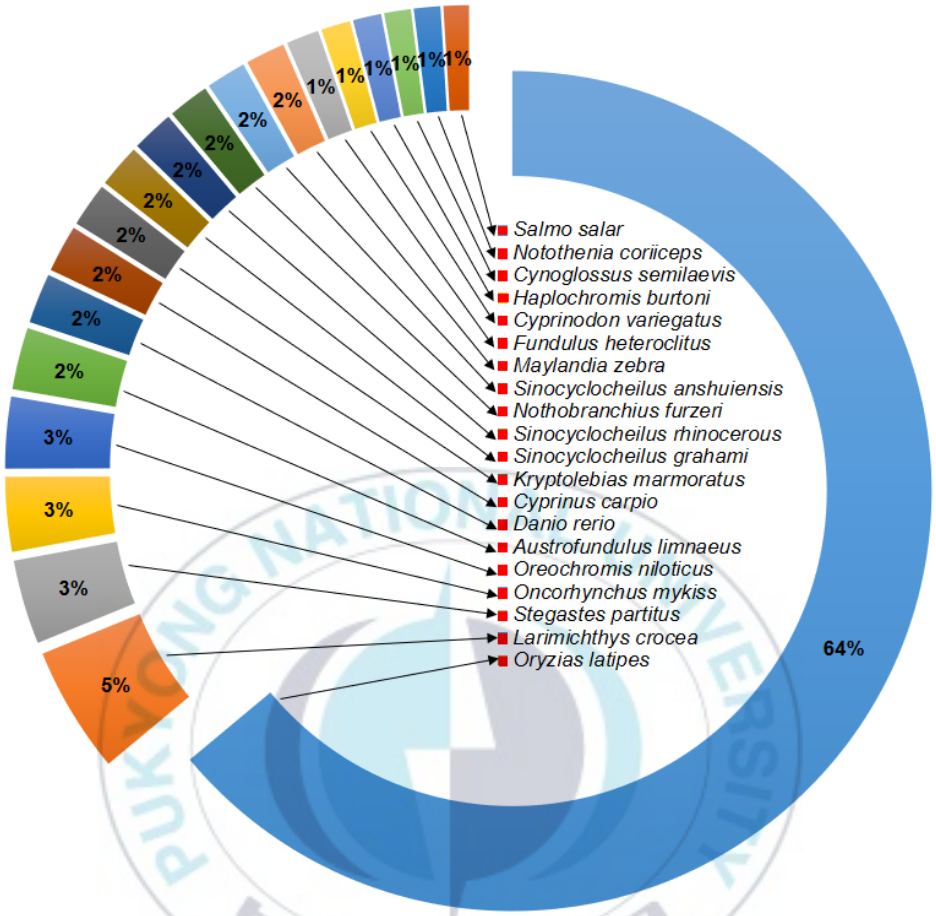

Figure 2: Species distribution of mostly aligned species according to the Blast results.

According to the BLAST results most homologies between aligned sequences from each species were; *Oryzias latipes* (64%), *Larimichthys crocea* (%5) and *Stegastes partitus* (%3).

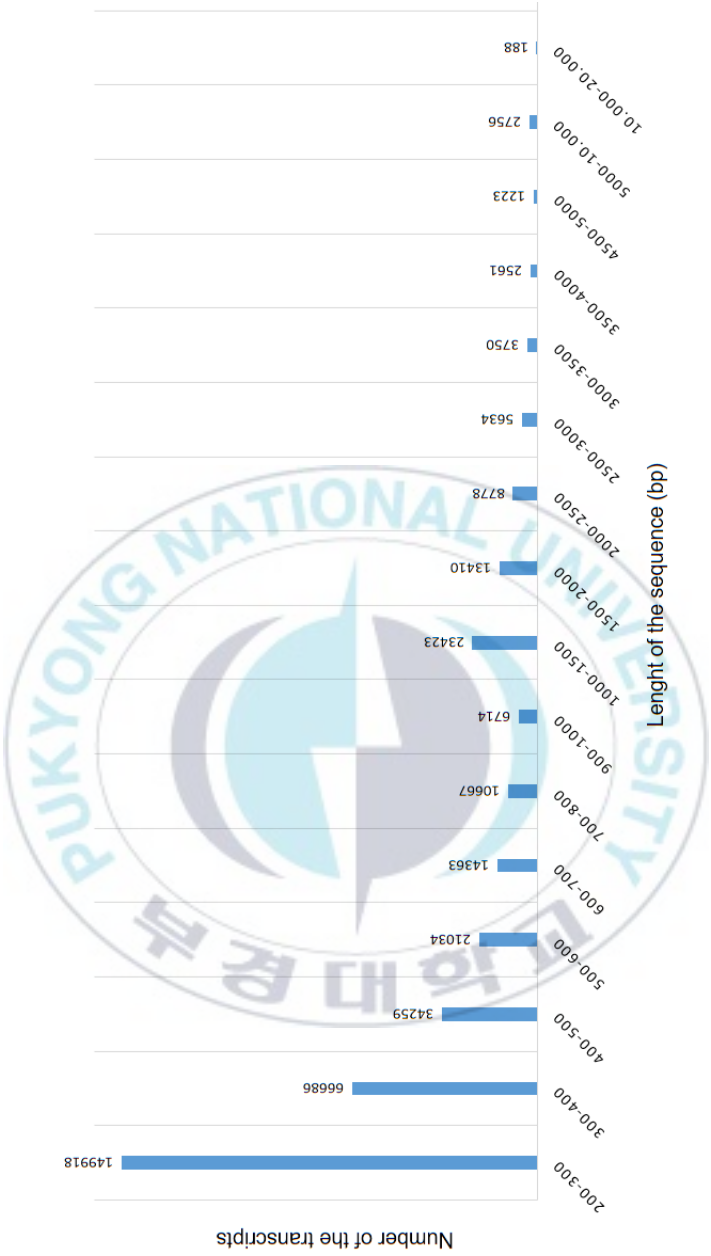

Figure 3: Number of the transcripts and distribution of bp length in transcriptome analysis.

*The length distribution of the transcriptome assembly from *O. dancena* was shown and number of transcripts indicates the proportion of sequences with matches in Unigene database is greater among the shorter assembled sequences (200 bp to 300 bp). Specifically, match efficiency was increased for sequences longer than 1,000 bp, whereas the match efficiency decrease to about for those ranging from 500 to 1,000 bp and for sequences between 1500 to 2000 bp

2.6 Trinotate annotation

All functional annotations were performed by using sequence comparison with public databases through comparison of all unigenes with the NCBI non-redundant protein database (NR http://www.ncbi.nlm.nih.gov/), the SwissProt database (http://www.expasy.ch/sprot), the NCBI non-redundant nucleic acid database (NT) and the Clusters of Orthologous Groups database (http://www.ncbi.nlm.nih.gov/COG/) using BLAST with an E-value less than 1e-6.

De novo assembly of RNA-seq data enables researchers to study transcriptomes without the need for a genome sequence; this approach can be usefully applied, for instance, in research on 'non-model organisms' of ecological and evolutionary importance, cancer samples or the microbiome (Haas et. al, 2013).

Amount of transcripts annotated blast results of De novo transcriptome assembly of *Oryzias dancena* is 369,054 while the sequences without sequence alignments estimated as 165,765 transcripts.

Databases	Annotated transcripts
SwissProt/ BlastX	90,644
SwissProt/ BlastP	57,123
Kegg	70,356
Eggnog	70,619
pFAM	25,920
Successfully Annotated genes	16,548

Table 3: Functional annotations of unigenes derived from ovary, testis and muscle cDNA libraries of *Oryzias dancena*

2.6.1 Gene ontology analysis

Gene Ontology (GO) terms were assigned to 16,548 annotated transcripts to estimate unigenes functions which was analyzed through The Blast2GO tool.

All 16,548 annotated transcripts were categorized by three major GO functional Domains: biological process, molecular functions and the metabolic process. According to these predictions; 1062 (GO) terms are involved in biological process, 410 (GO) terms are cellular components and (GO) terms 224 have molecular functions.

In the biological process category, the cellular process (GO:0009987) with 9450 sequences, metabolic process (GO:0008152) with 9008 sequences and single-organism process (GO:0044699) with 7892 sequences in level 2 terms were the most abundant terms.

In the molecular function category, binding (GO:0005488) with 8352 sequences, catalytic activity (GO:0003824) with 6843 sequences and transporter activity (GO: 0005215) with 1174 sequences were the most abundant, while, in the cellular component category, cell (GO:0005623) with 5607, membrane (GO:0016020) with 4095 sequences and organelle (GO:0043226) with 3564 sequences were the most abundant level 2 terms.

Level 2 terms were used since some genes were classified into more than one subcategory within each of the three major categories so the sum of genes in the subcategories could exceed 100%.

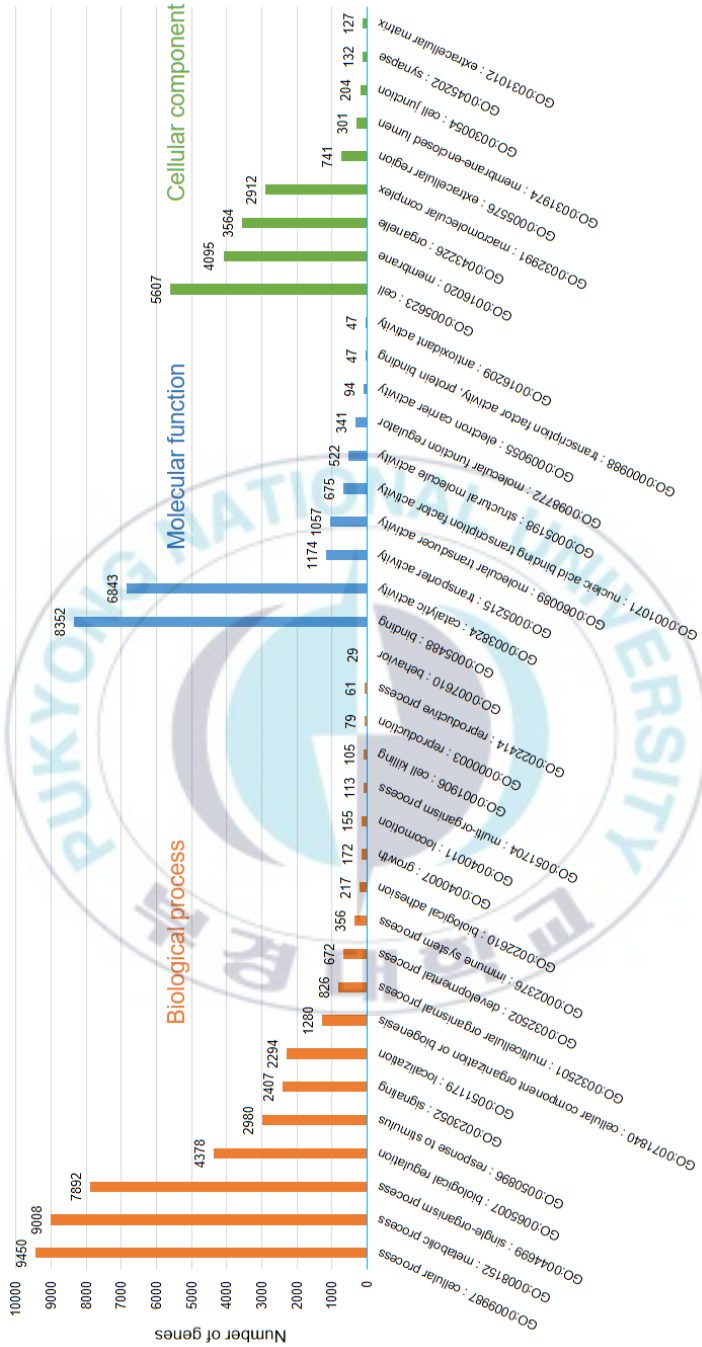

Figure 4:Gene ontology classification.

*Level 2 terms were used since some genes were classified into more than one subcategory within each of the tree major categories so the sum of genes in the subcategories could exceed 100 %

2.6.2 Functional classification based on KEEG Pathway Analysis

Functional classification and pathway assignments were based on Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathway database records the networks of molecular interactions in the cells and variants specific to particular organisms. This annotation served as a basis for analyzing not only the role of individual transcripts, but also the interaction with other genes.

Pathway-based analysis helps to further understand the biological functions and interactions of genes. Using KEGG, we annotated 70,356 transcripts and this annotation served as a basis for analyzing not only the role of individual transcripts, but also the interaction with other genes. Among the 70,356 annotated sequences, 3885 that were annotated with an enzyme code "ec" and a pathway code "map" for each number which were mapped to 123 KEGG pathways.

The top five KEGG pathways were Purine metabolism, Biosynthesis of antibiotics, Pyrimidine metabolism, Glycolysis / Gluconeogenesis and Glutathione metabolism. (Kanehisa et. al, 2004). All these pathways categorized Organismal Systems. Which were analyzed under the title of 11 sub-groups; Immune System, Translation, Signal Transduction, Lipid Metabolism, Energy Metabolism, Carbohydrate Metabolism, Amino acid Metabolism, Global and Overview Maps, Nucleotide Metabolism, Metabolism of Cofactors and Vitamins, Xenobiotic Biodegradation and Metabolism. Also these sub-groups represented by 4 larger groups; Genetic Information Processing, Environmental Information Processing, Metabolism and Organismal Systems.

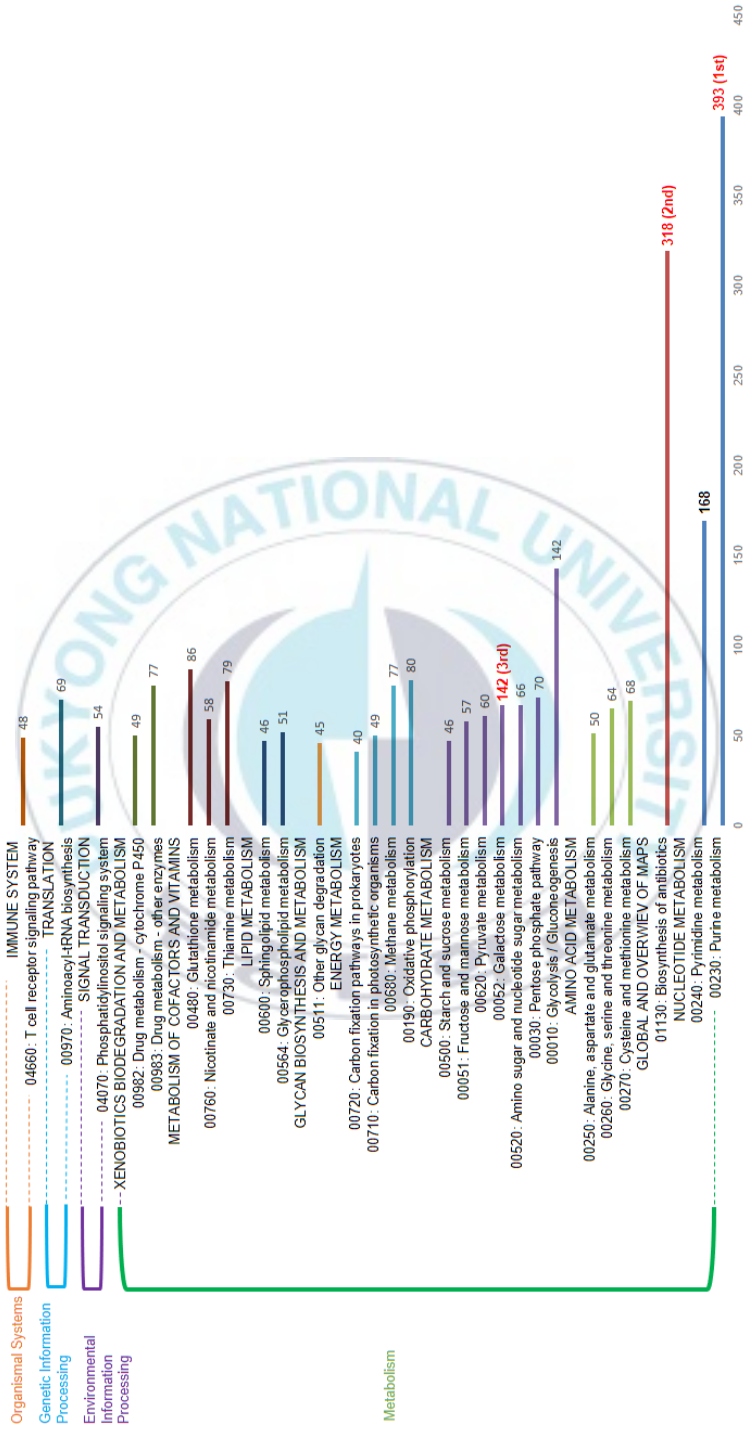

Figure 5: Kegg pathway (Functional classification and pathway assignments were based on Kyoto Encyclopedia of Genes and Genomes).

2.7 Identification of Sex-biased genes in *Oryzias dancena by* using DEG and TMM matrix expressions

Differentially expressed genes were found through comparison amount between the read counts which belongs to ovary, testis and muscle tissues of *O. dancena*. Differentially expressed genes (DEG's) were analyzed through edgeR v3.5.8 which enables screening the differential expression of replicated count data by using Bioconductor online server (https://www.bioconductor.org/) (Robinson et. al, 2010). While the P-value was determined by the false-discovery rate (FDR) (P-value>0.05) and the transcripts with an log2 fold change log2FC>1 and FDR<0.05 were regarded as differentially expressed genes. Expression differences between Ovary and Testis transcripts were explained through Volcano and MA Plot. Gene ontology (GO) annotation was performed to classify sex-biased genes.

TMM is explained as "Trimmed Mean of M-values" which is a normalization method used in edgeR. According to this method samples or observations that have the closest average expressions to mean of all samples is considered as reference samples while all others are test samples. For each test sample, the scaling factor is calculated based on weighted mean (weighted by estimated asymptotic variance) of log ratios between the test and reference, from a gene set removing most or lowest expressed genes and genes with highest or lowest log ratios.

Genes with sexually dimorphic expressions are often referred to as sex-biased genes although considering that in this sense genes themselves are not biased but rather their expression is considered as biased.

These genes include those that are expressed exclusively in one sex (sex-specific expression), as well as those that are expressed in both sexes but at a higher level in one sex (sex-enriched expression). The sex-biased genes can be further separated into male-biased and female-biased genes, depending on which sex shows higher expression. Genes with equal expression in the two sexes are referred to as unbiased genes (Ellegren & Parsch, 2007).

While identifying sex-biased genes, muscle transcriptome used as a control group. Sex-biased genes were investigated by 4 different terms; sex enriched genes that are expressed exclusively in one sex were analyzed as; female-enriched and male enriched genes. Female biased genes which are derived from ovary transcriptome were analyzed with differentially expressed genes, TMM expression of DEG's and female specific genes. Male biased genes which are derived from testis transcriptome were analyzed with differentially expressed genes, TMM expression of DEG's and male specific genes. Each analysis was investigated with a representative top 20 group of genes which shows the most explicit features (Figure 1).

All identifications of each gene were made through NCBI Blastx analysis and homologies were evaluated by e-value which is considered widely and accurate explanation of homologies between genes (Pearson, 2013).

It is indicated that e-values and bit-scores are much more useful for inferring homology. 30% identity threshold for homology underestimates the number of homologs detected by sequence similarity between humans and yeast by 33% (this is a minimum estimate; even more homologs can be detected by more sensitive comparison methods) (Pearson, 2013).

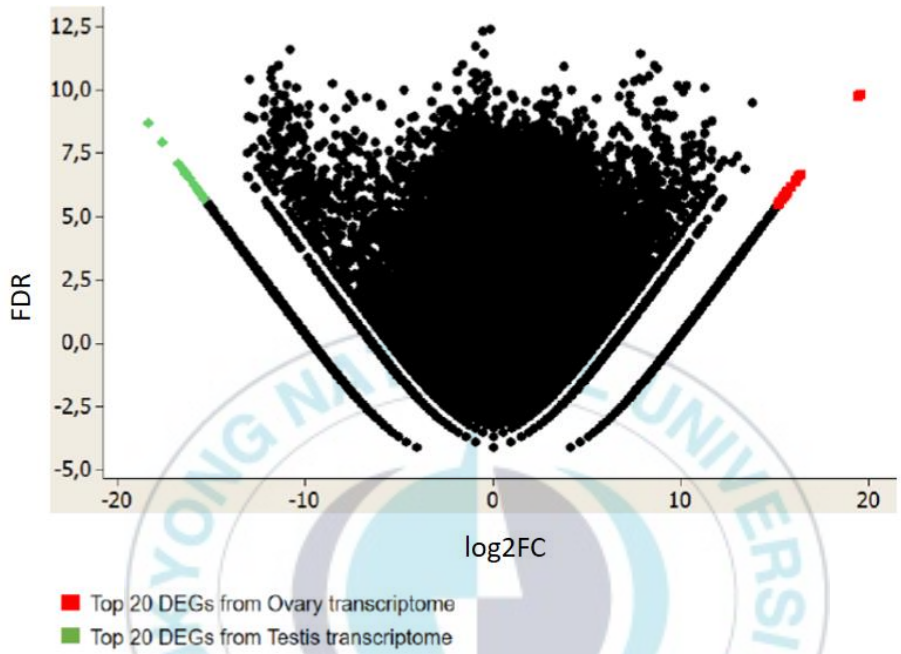

Figure 6: Top 20 DEG's between ovary vs testis expression analysis.

*(FDR < 0.05 & log2FC > 1). (false discovery rate (FDR) and log2 Fold change log2FC). **Compaired tissues from:** Ovary vs. Testis. **Compared transcripts:** 23,497. **Up-regulated genes:** 15,299. **Downregulated genes:** 8,198

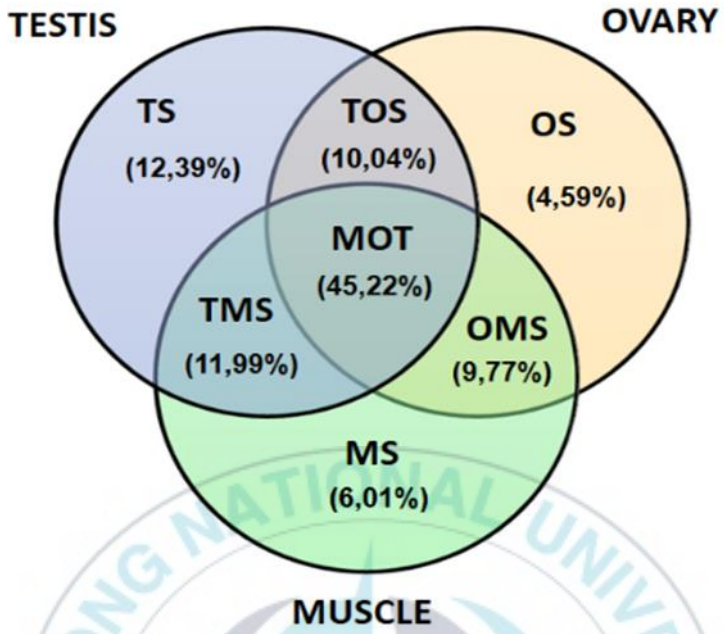

Figure 7: Distribution percentage of expressed transcripts in different tissues.

*TMM (Trimmed Mean of M values) Expression Matrix counts were used to describe the mutuality and specifity of the transcripts obtained from 3 different tissues; Ovary, Testis and Muscle. Transcripts that have "0" counts in each tissue at the same time were eliminated. Designations with initials explained as: **(A) TS**: Testis-Specific; transcripts that shows expression only in testis. Amount of transcripts: 44.611 (12,39%), **(B) OS:** Ovary-Specific; transcripts that shows expression only in ovary tissues. Amount of transcripts: 16.510 (4,59%) (**C)** MS: Muscle-Specific; transcripts that shows expression only in muscle tissues. Amount of transcripts: 21.630 (6,01%), **(D) TOS:** Ovary-Testis Specific: Transcripts that shows expression mutually only in ovary and testis. Amount of transcripts: 36.143 (10,04%), **(E) MOT:** Muscle-Ovary-Testis; Transcripts that shows expressions mutually in all tissues. Amount of transcripts: 16.2847 (45,22%), **(F) TMS:** Testis-Muscle Specific: Transcripts that shows expressions mutually only between testis and muscle tissues. Amount of transcripts: 43.158 (11,99%), **(G) OMS:** Ovary-Testis Specific: Transcripts that shows expressions mutually between muscle and ovary tissues. Amount of transcripts: 35.186 (9,77%).

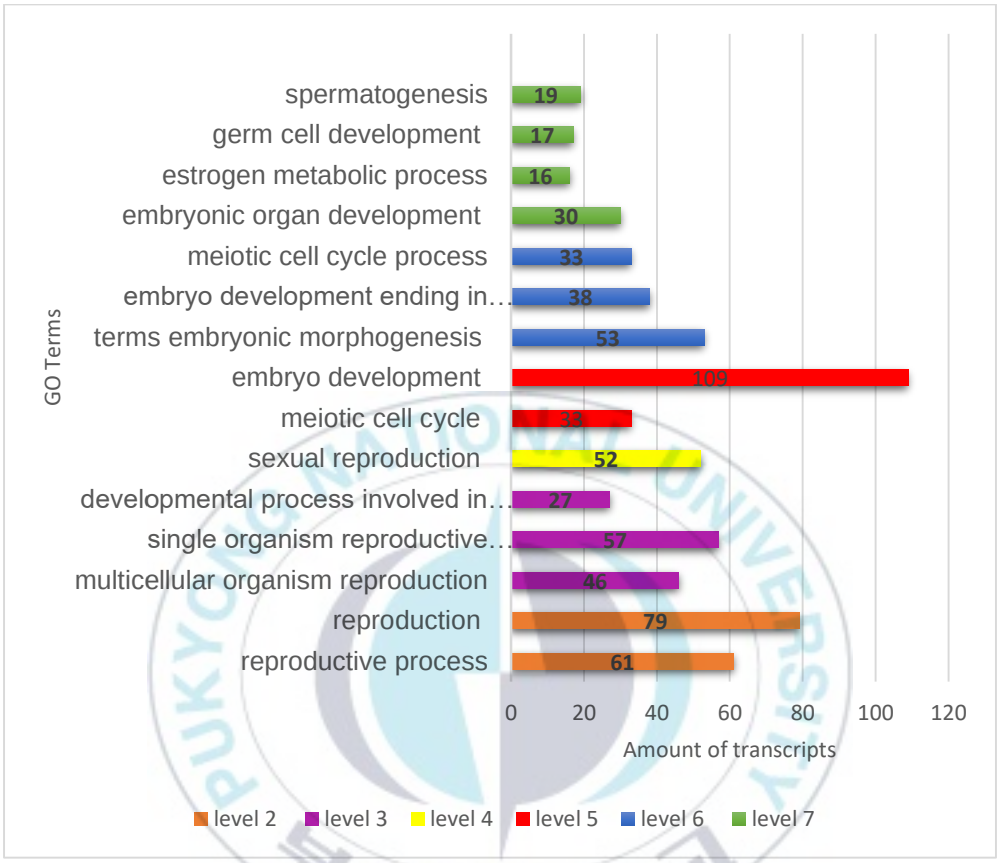

Figure 8: Gonad-biased GO Annotation terms:

*Within Biological process category among 1062 categories 15 subcategories from Level 2 to Level 7 were estimated directly related with sex-biased interactions mostly pertinent with reproduction which were assigned through GO Annotation. Gonad-biased Go Annotation explains some of the Biological Processes in ovary-biased and testis-biased genes

3.RESULTS

3.1 Sex-enriched genes from transcriptome analysis

Genes that are expressed predominantly in one sex compared to the other sex are identified as sex-enriched genes. These genes may be characteristic for one sex and be differentially expressed while it may not be a characteristic gene for one sex but still shows some expression (Figure 7D). Also genes from testis and muscle transcriptome with "0" TMM expression levels in genes from ovary transcriptome were trimmed to be able to find female-specific genes.

3.1.1 Female-enriched genes

In this study; female-enriched genes were referred as the genes from ovary and testis tissue transcriptome which shows more expression in ovary tissues than testis tissues. Female enriched genes were identified through elimination of muscle transcriptome, ovary-specific genes and testis specific genes (Figure 7d). Top 20 genes which shows highest TMM expressions were chosen as a representative group for female-enriched genes (Table 4).

According to the BlastX results derived from Trinotate annotation sprot and Blast2go; Female enriched genes have the abundancy of polysialoglycoproteins, ZP domain genes and hornerin genes compared to the male-enriched genes.

Transcript ID	Blast Hit	logFC	FDR	Ovary TMM Count	Testis TMM Count	e-value	Species Blast
TRINITY_DN91464_c0_g1_i2	**polysialoglycoprotein-like isoform X2**	-11,358744	1,11E-24	9946,451	4,156	1,00E-11	*O. melastigma (XP_024115070.1)*
TRINITY_DN110075_c1_g15_i1	**polysialoglycoprotein-like isoform X2**	-12,304285	1,99E-25	6393,401	1,296	8,00E-29	*O. melastigma (XP_024115070.1)*
TRINITY_DN103098_c0_g2_i1	**polysialoglycoprotein-like isoform X2**	-11,124304	8,47E-24	5524,497	2,765	3,00E-21	*O. melastigma (XP_024115070.1)*
TRINITY_DN103643_c0_g1_i2	**uncharacterized protein LOC112162118**	-12,996735	1,23E-28	2372,87	0,327	3,00E-174	*O. melastigma (XP_024153545.1)*
TRINITY_DN103643_c0_g1_i4	**uncharacterized protein LOC112162118**	-10,48843	2,49E-23	1772,585	1,379	3,00E-174	*O. melastigma (XP_024153545.1)*
TRINITY_DN165962_c0_g1_i1	**uncharacterized protein LOC101175394**	-10,242415	1,13E-20	715,979	0,594	1,00E-24	*O. latipes (XP_004069411.2)*
TRINITY_DN33152_c0_g1_i1	**zona pellucida sperm-binding protein 4-like**	-10,255883	2,25E-19	700,342	0,501	4,00E-46	*O. melastigma(XP_024129004.1)*
TRINITY_DN120831_c6_g2_i7	**uncharacterized protein LOC112161774**	-8,8286062	5,16E-18	631,484	1,482	5,00E-44	*O. melastigma (XP_024152986.1)*
TRINITY_DN114781_c3_g9_i1	**protein SSXT-like isoform X1**	-9,6754775	8,35E-18	622,103	0,89	1,00E-18	*O. melastigma (XP_024136794.1)*
TRINITY_DN213862_c0_g1_i1	**uncharacterized protein LOC112157708**	-10,649937	4,78E-21	597,572	0,397	5,00E-69	*O. melastigma (XP_024146380.1)*
TRINITY_DN103098_c0_g4_i1	**polysialoglycoprotein-like isoform X1**	-10,420867	6,88E-19	539,111	0,402	1,00E-15	*O. melastigma (XP_024115069.1)*
TRINITY_DN34108_c0_g1_i1	**protein SSXT-like**	-10,479728	5,77E-20	517,394	0,398	0,42	*O. melastigma (XP_024136799.1)*
TRINITY_DN114781_c3_g5_i1	**annexin A7-like**	-11,340309	1,39E-19	469,837	0,178	2,00E-11	*O. melastigma (XP_024136801.1)*
TRINITY_DN113113_c0_g1_i2	**ribosyldihydronicotinamide dehydrogenase [quinone]-like**	-10,234074	4,31E-21	360,179	0,315	2,00E-91	*O. melastigma (XP_024141800.1)*
TRINITY_DN185053_c0_g1_i1	**adenomatous polyposis coli protein-like isoform X1**	-10,35957	6,07E-17	328,743	0,241	7,00E-33	*O. melastigma (XP_024153950.1)*
TRINITY_DN53597_c0_g1_i1	**uncharacterized protein LOC112158501**	-11,030856	9,39E-19	326,183	0,151	8,00E-54	*O. melastigma (XP_024147696.1)*
TRINITY_DN133204_c0_g1_i1	**protein Z-dependent protease inhibitor-like**	-8,441536	4,71E-15	315,469	1,065	4,00E-41	*O. melastigma (XP_024124201.1)*
TRINITY_DN96332_c1_g2_i2	**vegetative cell wall protein gp1-like**	-6,1644005	7,37E-12	312,788	4,902	5,00E-137	*O. melastigma (XP_024135475.1)*
TRINITY_DN214418_c0_g1_i1	**zona pellucida sperm-binding protein 3-like**	-6,0211265	3,2E-11	280,001	4,69	6,00E-57	*O. melastigma (XP_024129291.1)*
TRINITY_DN101518_c0_g1_i7	**hornerin**	-9,6686726	8,7E-18	276,563	0,347	3,00E-32	*O. latipes (XP_004085507.1)*

Table 4 Top 20 female-enriched genes from gonadal transcriptome

*TMM expression values of male and female specific genes were compared and highest values of testis TMM expression levels were chosen

3.1.2 Male-enriched genes

Male-enriched genes were identified as the genes from ovary and testis tissue transcriptome which shows more expression in testis tissues than ovary.

Male enriched genes were explained through elimination of muscle transcriptome, ovary-specific genes and testis specific genes are shown in Figure 7 and top 20 genes which is with the highest TMM expressions were chosen as a representative group for male-enriched genes (Table 5).

Male enriched genes are mostly represented by trypsin, elastase-1 and E-3 ubiquitin proteins. Cilia and flagella associated proteins and Kelch-like proteins are also predominant genes found through analysis.

Transcrip ID	Blast hit	logFC	FDR	TMM testis	TMM Ovary	e-value	Species Blast
TRINITY_DN110158_c0_g3_i1	**trypsin-2-like**	11,23106	6,04E-25	4276,965	9,931	0.0	*O. melastigma (XP_024119829.1)*
TRINITY_DN110158_c0_g3_i2	**trypsin-2-like**	19,53853	3,26E-33	918,824	0,257	0.0	*O. melastigma (XP_024119829.1)*
TRINITY_DN121297_c2_g1_i1	**DMT family transporter**	19,4451	3,26E-33	789,604	0,296	0.72	*R. tolerans (WP_074785497.1)*
TRINITY_DN110158_c0_g3_i4	**trypsin-2-like**	13,81363	3,23E-29	618,01	0,362	0.0	*O. melastigma (XP_024119829.1)*
TRINITY_DN182029_c0_g1_i1	**polyadenylate-binding protein 1-like**	12,99523	2,18E-25	522,234	16,358	7e-42	*O. mykiss (XP_021417737.1)*
TRINITY_DN26142_c0_g1_i1	**hypothetical protein**	11,66696	3,2E-25	443,833	1,172	6.8	*H. saccharovorum (WP_004046549.1)*
TRINITY_DN110128_c1_g1_i2	**elastase-1-like**	13,40689	6,87E-25	417,204	3,58	8e-150	*O. melastigma (XP_024125919.1)*
TRINITY_DN114049_c0_g1_i1	**uncharacterized protein LOC112156151**	13,39921	7,09E-25	377,316	0,935	8e-65	*O. melastigma (XP_024144109.1)*
TRINITY_DN86235_c0_g1_i2	**embryonic polyadenylate-binding protein 2 isoform X1**	12,73664	7,11E-25	328,86	60,552	3e-42	*O. melastigma (XP_024150579.1)*
TRINITY_DN73243_c0_g1_i1	**No hit**	16,16105	9,64E-25	241,783	0,14	-	-
TRINITY_DN123028_c2_g1_i1	**protein kintoun isoform X1**	12,28345	1,24E-24	225,891	0,377	3e-82	*O. melastigma (XP_024131672.1)*
TRINITY_DN100664_c0_g1_i2	**outer dynein arm protein 1-like**	11,31168	1,34E-24	222,585	0,297	0.0	*O. melastigma (XP_024137129.1)*
TRINITY_DN97106_c1_g1_i2	**putative E3 ubiquitin-protein ligase UBR7**	12,16355	2,37E-24	196,997	0,055	0.0	*O. melastigma (XP_024123651.1)*
TRINITY_DN119893_c1_g1_i2	**centrosomal protein of 55 kDa-like isoform X3**	11,66971	2,4E-24	183,45	0,114	0.0	*O. melastigma(XP_024132448.1)*
TRINITY_DN120526_c0_g2_i3	**kelch-like protein 10**	11,27969	2,47E-24	162,276	0,068	2e-80	*O. melastigma (XP_024136685.1)*
TRINITY_DN45618_c0_g2_i1	**cilia- and flagella-associated protein 161**	11,24481	2,97E-24	149,965	1,004	1e-24	*O. melastigma (XP_024136579.1)*
TRINITY_DN211665_c0_g1_i1	**protein brambleberry-like**	11,97752	6,46E-24	146,379	28,994	9e-28	*O. melastigma (XP_024134239.1)*
TRINITY_DN114202_c4_g1_i1	**uncharacterized protein LOC112149136**	11,59982	6,53E-24	141,736	0,103	3e-135	*O. melastigma (XP_024132393.1)*
TRINITY_DN127640_c5_g1_i2	**stress response protein NST1-like**	11,09454	9,99E-24	137,353	0,6	5e-17	*O. melastigma (XP_024116739.1)*
TRINITY_DN30778_c0_g1_i1	**solute carrier family 35 member E4**	11,09671	1,25E-23	136,222	0,195	1e-52	*O. melastigma(XP_024154615.1)*

Table 5: Top 20 male-enriched genes from gonadal transcriptome.

*TMM expression values of male and female specific genes were compared and highest values of testis TMM expression levels were chosen

3.2 Female-biased genes

Genes that are differentially expressed in female tissues are defined as female-biased genes. In this study, female-biased genes were identified by using the comparison of differentially expressed genes against combined data of differentially expressed genes in Trimmed Mean of M-values (TMM expression Matrix) to be able to find a significant result. Also genes from testis and muscle transcriptome with "0" TMM expression levels in genes from ovary transcriptome were trimmed to be able to find female-specific genes (Figure 7B,7D,7E and 7G).

3.2.1 Differentially expressed genes and TMM expression levels of Female biased genes

Compared results obtained by using TMM/DEG combined expression table and DEG expression table gives the conclusion that different genes are expressed and different genes are predominant in each table of the top 500 genes according to the annotation results.

According to that; the results which were analyzed as in DEG (Differentially expressed genes) in ovary transcriptome choriogenin H-related proteins were abundant compared to the TMM/DEG combined expression data. Nevertheless, TMM/DEG combined expression data shows abundancy of immunoglobulin, hepcidin and zygote arrest protein compared to DEG data only.

Transcript ID	Blast Hit	logFC	FDR	TMM Count	e-value	Species Blast
TRINITY_DN110075_c1_g11_i1	**polysialoglycoprotein-like isoform X2**	-10,97175	1,93743E-24	26086,3	2E-21	*O. melastigma (XP_024115070.1)*
TRINITY_DN103098_c0_g6_i2	**polysialoglycoprotein-like isoform X2**	-10,41865	5,72343E-23	14059,01	5E-19	*O. melastigma (XP_024115070.1)*
TRINITY_DN91464_c0_g1_i2	**polysialoglycoprotein-like isoform X2**	-11,35874	1,10655E-24	9946,451	1E-11	*O. melastigma (XP_024115070.1)*
TRINITY_DN110075_c1_g15_i1	**polysialoglycoprotein-like isoform X2**	-12,30429	1,99156E-25	6393,401	8E-29	*O. melastigma (XP_024115070.1)*
TRINITY_DN103098_c0_g2_i1	**polysialoglycoprotein-like isoform X2**	-11,1243	8,47249E-24	5524,497	3E-21	*O. melastigma (XP_024115070.1)*
TRINITY_DN70523_c0_g1_i1	**gamma-gliadin-like isoform X6**	-11,18505	1,33805E-24	5154,741	1E-20	*O. melastigma (XP_024115784.1)*
TRINITY_DN106694_c0_g3_i3	**YEATS domain-containing protein 2**	-10,87193	4,70761E-24	4932,711	2E-48	*O. latipes (XP_023805023.1)*
TRINITY_DN106694_c0_g3_i2	**YEATS domain-containing protein 2**	-11,33245	9,03083E-25	4130,57	9E-49	*O. latipes (XP_023805023.1)*
TRINITY_DN102545_c0_g1_i4	**uncharacterized protein LOC101175394**	-10,24058	7,84555E-22	4025,734	2E-17	*O. latipes (XP_004069411.2)*
TRINITY_DN65935_c0_g1_i1	**serine protease inhibitor A3N-like isoform X1**	-9,236261	9,77381E-20	3787,899	5E-18	*O. melastigma (XP_024131974.1)*
TRINITY_DN110075_c1_g7_i1	**polysialoglycoprotein-like isoform X1**	-10,90873	1,45551E-23	3744,812	2E-27	*O. melastigma (XP_024115069.1)*
TRINITY_DN103643_c0_g3_i1	**uncharacterized protein LOC112147994**	-10,36198	1,83537E-22	3664,093	4E-16	*O. melastigma (XP_024130508.1)*
TRINITY_DN123624_c0_g12_i1	**uncharacterized protein LOC105355301**	-11,49019	2,67417E-24	3185,736	1E-53	*O. latipes (XP_023819300.1)*
TRINITY_DN110075_c1_g16_i1	**polysialoglycoprotein-like isoform X2**	-10,65742	2,30581E-22	3131,811	2E-26	*O. melastigma (XP_024115070.1)*
TRINITY_DN103643_c0_g1_i1	**uncharacterized protein LOC112162118**	-11,76842	3,89056E-26	3028,826	0.0	*O. melastigma (XP_024133545.1)*
TRINITY_DN114781_c3_g4_i1	**DUF1373 domain-containing protein**	-12,06995	3,97154E-24	2679,027	2E-16	*A. baumannii (WP_071217127.1)*
TRINITY_DN87325_c0_g2_i1	**polysialoglycoprotein-like isoform X2**	-9,717204	1,1162E-20	2448,508	7E-16	*O. melastigma (XP_024115070.1)*
TRINITY_DN87325_c0_g1_i1	**polysialoglycoprotein-like isoform X1**	-10,38525	8,81757E-22	2407,017	2E-12	*O. melastigma (XP_024115069.1)*
TRINITY_DN103643_c0_g1_i2	**uncharacterized protein LOC112162118**	-12,99674	1,22819E-28	2372,87	3E-174	*O. melastigma (XP_024133545.1)*
TRINITY_DN95607_c0_g2_i2	**polysialoglycoprotein-like isoform X1**	-9,292552	6,88877E-20	2205,206	7E-23	*O. melastigma (XP_024115069.1)*

Table 6: Top 20 DEG/TMM expression from ovary transcriptome:

*Combined data were lined up according to the highest TMM expression values from ovary transcriptome.

Transcript ID	Blast Hit	Log FC	FDR	TMM	e-value	*Species Blast*
TRINITY_DN114781_c3_g7_i5	**calcium-binding protein P-like**	-18,376	4E-30	2124,7	6E-59	*O. melastigma (XP_024136790.1)*
TRINITY_DN101518_c0_g1_i5	**hornerin**	-17,618	3E-28	1050,7	1E-33	*O. melastigma (XP_004085507.1)*
TRINITY_DN76787_c0_g2_i1	**calcium-binding protein P-like**	-16,802	4E-26	1371,3	8E-50	*O. melastigma (XP_024138693.1)*
TRINITY_DN84904_c0_g2_i1	**calcium-binding protein P-like**	-16,755	5E-26	477,33	1E-44	*O. melastigma (XP_024138693.1)*
TRINITY_DN120831_c6_g2_i5	**protein transport protein SEC31-like**	-16,471	2E-25	1007,5	1E-37	*O. melastigma (XP_023819472.1)*
TRINITY_DN74329_c0_g1_i1	**ribosyldihydronicotinamide dehydrogenase [quinone]-like**	-16,402	3E-25	1046	9E-60	*O. melastigma (XP_024141800.1)*
TRINITY_DN20780_c0_g2_i1	**ZPC1**	-16,369	3E-25	295,19	5E-71	*O. latipes (AAN31188.1)*
TRINITY_DN122460_c0_g2_i1	**zona pellucida sperm-binding protein 3-like**	-16,322	4E-25	386,61	3E-40	*A. ocellaris (XP_023141066.1)*
TRINITY_DN113113_c0_g1_i1	**ribosyldihydronicotinamide dehydrogenase [quinone]-like**	-16,274	6E-25	364,04	1E-74	*O. melastigma (XP_024141756.1)*
TRINITY_DN120831_c6_g2_i8	**protein transport protein SEC31-like**	-16,173	9E-25	795,93	7E-38	*O. melastigma (XP_023819472.1)*
TRINITY_DN122885_c0_g1_i4	**dihydropyrimidinase-related protein 2-like isoform X1**	-15,95	3E-24	32,275	0.0	*O. melastigma (XP_024120802.1)*
TRINITY_DN123436_c2_g3_i1	**hornerin-like isoform X3**	-15,893	4E-24	1091,7	8E-20	*O. latipes (XP_011472519.2)*
TRINITY_DN108387_c5_g3_i1	**immunoglobulin light chain**	-15,785	7E-24	99,291	3E-110	*A. schlegelii (ACH72079.1)*
TRINITY_DN113498_c0_g2_i2	**zona pellucida sperm-binding protein 3-like**	-15,711	1E-23	89,868	0.0	*O. melastigma (XP_024149062.1)*
TRINITY_DN110359_c1_g2_i1	**putative mediator of RNA polymerase II transcription subunit 12 isoform X1**	-15,703	1E-23	77,184	0.0	*O. melastigma (XP_024122939.1)*
TRINITY_DN125696_c0_g1_i1	**zonadhesin-like**	-15,599	2E-23	47,471	0.0	*O. melastigma (XP_024129599.1)*
TRINITY_DN106825_c0_g1_i1	**transcription factor IIIA**	-15,597	2E-23	87,313	0.0	*O. melastigma (XP_024136158.1)*
TRINITY_DN107918_c0_g1_i11	**uncharacterized protein LOC101163526 isoform X2**	-15,538	3E-23	132,12	3E-31	*O. latipes (XP_011486131.2)*
TRINITY_DN3850_c0_g1_i1	**zona pellucida sperm-binding protein 3-like isoform X5**	-15,496	4E-23	306,24	5E-66	*O. melastigma (XP_024127244.1)*
TRINITY_DN127304_c0_g1_i3	**alpha-2-macroglobulin-like protein 1 isoform X1**	-18,376	7E-23	14,959	0.0	*O. latipes (XP_023817529.1)*

Table 7: Top 20 differentially expressed genes from Ovary transcriptome.

*(false discovery rate ;**FDR** and log Fold change; **logFC**).

3.2.2 Female-specific genes from transcriptome analysis

Female specific genes were identified by the elimination of the transcripts that have expression in all other tissues except ovary transcriptome by using TMM matrix data (Table 8) (Figure 7).

Top 500 genes from female-specific genes were analyzed according to the Trinoate annotation (sprot) results. Female specific genes show abundance of choriogenin H-related proteins, hornerin, cythchrome p450, catepsin and polysialoglycoproteins. These genes are also represented in top 20 genes in female-specific data set.

Transcript ID	Blast Hit	Log FC	FDR	TMM	e-value	Species Blast
TRINITY_DN84904_c0_g2_i1	**uncharacterized protein LOC112153723**	-16,75537852	5,044E-26	477,331	5E-104	*O. melastigma (XP_024139864.1)*
TRINITY_DN123436_c2_g7_i1	**hornerin-like isoform X1**	-14,19080596	1,16E-19	461,034	6E-20	*O. melastigma (XP_004085506.1)*
TRINITY_DN114781_c3_g2_i1	**protein SSXT-like isoform X3**	-14,27946307	6,696E-20	369,064	7E-21	*O. melastigma (XP_024136796.1)*
TRINITY_DN113113_c0_g1_i1	**ribosyldihydronicotinamide dehydrogenase [quinone]-like**	-16,27431574	5,787E-25	364,037	1E-74	*O. melastigma (XP_024141756.1)*
TRINITY_DN6788_c0_g1_i1	**Zona pellucida protein X**	-14,08574896	2,188E-19	255,947	2E-37	*S. maximus (AWP20165.1)*
TRINITY_DN103098_c0_g6_i1	**polysialoglycoprotein-like isoform X1**	-12,99605202	1,877E-16	217,117	1E-14	*O. melastigma (XP_024115069.1)*
TRINITY_DN123624_c0_g25_i1	**uncharacterized protein LOC112162118**	-13,92419919	5,928E-19	190,157	4E-21	*O. melastigma (XP_024153545.)*
TRINITY_DN102545_c0_g1_i6	**polysialoglycoprotein-like isoform X2**	-14,17277689	1,305E-19	189,078	4E-17	*O. melastigma (XP_024115070.1)*
TRINITY_DN110359_c1_g5_i1	**extensin-3-like isoform X2**	-12,47345089	4,927E-15	169,238		*O. melastigma (XP_024141736.1)*
TRINITY_DN193804_c0_g1_i1	**uncharacterized protein LOC112161484 isoform X2**	-12,95444499	2,438E-16	161,336	1E-48	*O. melastigma (XP_024152400.1)*
TRINITY_DN196649_c0_g1_i1	**uncharacterized protein LOC112139601**	-13,97171767	4,458E-19	154,915	9E-62	*O. melastigma (XP_024118187.1)*
TRINITY_DN60645_c0_g1_i1	**No hit**	-12,68297979	1,333E-15	137,683	-	-
TRINITY_DN107918_c0_g1_i11	**uncharacterized protein LOC101163526 isoform X2**	-15,53779952	3,021E-23	132,121	3E-31	*O. latipes (XP_011486131.2)*
TRINITY_DN123436_c2_g26_i1	**uncharacterized protein LOC101163526 isoform X2**	-12,10775739	4,723E-14	129,704	6E-51	*O. latipes (XP_011486131.2)*
TRINITY_DN106486_c2_g9_i1	**hornerin**	-12,48961289	4,454E-15	117,002	1E-11	*O. melastigma (XP_004085507.1)*
TRINITY_DN123624_c0_g8_i1	**uncharacterized protein LOC112162118**	-13,20012032	5,396E-17	115,106	2E-30	*O. melastigma (XP_024153545.1)*
TRINITY_DN173638_c0_g1_i1	**choriogenin H-related protein**	-12,44265032	5,958E-15	112,109	3E-39	*O. Latipes (44D40960.1)*
TRINITY_DN123436_c2_g31_i3	**hornerin**	-12,01030035	8,57E-14	109,718	7E-19	*O. latipes (XP_004085507.1)*
TRINITY_DN79841_c0_g1_i2	**choriogenin H-related protein**	-14,42365239	2,696E-20	106,379	4E-35	*O. latipes (44D40960.1)*
TRINITY_DN123624_c0_g11_i2	**uncharacterized protein LOC112144577**	-14,90507431	1,297E-21	106,116	2E-69	*O. melastigma (XP_024124922.1)*

Table 8: Top 20 Female-specific genes:

*Combined data were lined up according to the highest TMM expression values which expressed only in ovary transcriptome

3.3 Male Biased genes

Genes that are differentially expressed in male tissues are defined as male-biased genes. Identification of male-biased genes were made by using the comparison of differentially expressed genes against combined data of differentially expressed genes in Trimmed Mean of M-values (TMM expression Matrix) to be able to find a significant result. Also genes from ovary and muscle transcriptome with 0 TMM expression levels in genes from ovary transcriptome were trimmed to be able to find male-specific genes.

3.3.1 Differentially expressed genes and TMM expression levels of Male-biased genes

TMM/DEG combined expression table and DEG expression from testis transcriptome were analyzed and compared. In conclusion comparisons of both table gives the conclusion that different genes are expressed and different genes are predominant in each table of the top 500 genes according to the annotation results.

In accordance with these results in top 500 genes from each data which were analyzed as in DEG (Differentially expressed genes) in testis transcriptome; septin proteins and kelch-like proteins were abundant compared to the TMM/DEG combined expression data. Nevertheless, TMM/DEG combined expression data shows abundancy of elastase, pancreatic alpha amylase, chymotropsin, bile-salt activated-ligase, trypsin-2-like and kintaun compared to DEG data only. (Table: 9) (Table: 10) These genes were also represented in top 20 genes accordingly. However, 3 ubiquitin-protein ligase proteins, sperm-associated antigen, protein kintoun and la-related genes were equally abundant transcripts in each dataset.

Transcript ID	Blast hit	logFC	FDR	TMM	e-value	Species Blast
TRINITY_DN92991_c0_g1_i2	**No hit**	11.23106	6.04E-25	15773.516	-	-
TRINITY_DN92991_c0_g2_i2	**hypothetical protein**	19.53853	3.26E-33	7326.39	2.00e-07	***O. melastigma (XP_004085507.1)***
TRINITY_DN92991_c0_g1_i1	**hypothetical protein**	11.21837	1.77E-23	4415.008	1e-07	***E. faecium (WP_082194353.1)***
TRINITY_DN110158_c0_g3_i1	**trypsin-2-like**	8.5822	1.85E-18	4276.965	0.0	***O. melastigma (XP_024119829.1***
TRINITY_DN121297_c2_g1_i2	**No hit**	9.620992	4.95E-21	3778.473	-	-
TRINITY_DN110128_c1_g1_i3	**elastase-1-like**	8.758631	6.21E-19	3347.199	2e-169	***O. melastigma (XP_024125918.1)***
TRINITY_DN108808_c0_g1_i2	**elastase-1-like**	7.801752	2.37E-16	3077.594	3e-154	***O. melastigma (XP_024125919.1)***
TRINITY_DN94557_c2_g1_i1	**pancreatic alpha-amylase-like**	7.83898	1.73E-16	2977.791	0.0	***O. melastigma (XP_024118026.1)***
TRINITY_DN108373_c1_g5_i1	**trypsin-3-like**	8.320279	9.61E-18	2934.757	7e-156	***O. melastigma (XP_024119249.1)***
TRINITY_DN42764_c0_g1_i1	**chymotrypsin B-like**	7.988568	7.58E-17	2893.949	0.0	***O. melastigma XP_024116958.1)***
TRINITY_DN110158_c0_g3_i3	**trypsin-2-like**	7.483635	1.74E-15	2778.949	8e-95	***O. melastigma (NP_001098143.1)***
TRINITY_DN100443_c0_g1_i1	**nuclease-sensitive element-binding protein 1**	3.789827	6.16E-06	2507.244	0.0	***O.latipes (XP_024117542.1)***
TRINITY_DN114301_c0_g1_i9	**protein kintoun isoform X1**	11.09671	1.25E-23	2483.626	1.00e-93	***O. melastigma (XP_024131672.1)***
TRINITY_DN114049_c0_g1_i3	**uncharacterized protein LOC112156151**	9.861464	8.33E-22	2366.337	0.0	***O. melastigma (XP_024144109.1)***
TRINITY_DN106596_c0_g1_i1	**pancreatic alpha-amylase-like**	6.69365	2.57E-13	2175.87	2e-86	***O. melastigma(XP_024118026.1)***
TRINITY_DN114301_c0_g1_i5	**trypsin-2-like**	10.09385	2.36E-22	2120.691	5e-135	***O. melastigma (XP_024119829.1)***
TRINITY_DN53040_c0_g1_i1	**cystatin-like**	4.522659	1.1E-07	2073.743	2e-64	***O. melastigma (XP_024153315.1)***
TRINITY_DN114301_c0_g1_i7	**trypsin-2-like**	7.26116	7.18E-15	1923.878	4e-180	***O. melastigma (XP_024119829.1)***
TRINITY_DN106596_c0_g2_i2	**pancreatic alpha-amylase-like**	6.920017	5.59E-14	1789.255	0.0	***O. melastigma (XP_024118026.1)***
TRINITY_DN102556_c0_g1_i1	**insulin-like peptide INSL5**	8.6666	1.45E-18	1670.888	7e-68	***O. melastigma (XP_024124719.1)***

Table 9: Top 20 DEG/TMM expression from testis transcriptome.

*Combined data were lined up according to the highest TMM expression values from testis transcriptome.

Transcript ID	Blast Hit	Log FC	FDR	TMM	e-value	Species Blast
TRINITY_DN92991_c0_g2_i2	**No hit**	19,53853	3,26E-33	7326,39	6E-59	*O. melastigma (XP_024136790.1)*
TRINITY_DN106596_c0_g2_i3	**pancreatic alpha-amylase-like**	19,4451	3,26E-33	955,96	1E-33	*O. melastigma (XP_004085507.1)*
TRINITY_DN106623_c0_g1_i1	**cilia- and flagella-associated protein 99 isoform X1**	16,37055	3,44E-25	76,973	8E-50	*O. melastigma (XP_024138693.1)*
TRINITY_DN115142_c0_g1_i3	**solute carrier family 35 member E4**	16,25986	6,04E-25	54,772	1E-44	*O. melastigma (XP_024138693.1)*
TRINITY_DN73294_c0_g1_i1	**E3 ubiquitin-protein ligase ZNRF1-like**	16,16105	9,64E-25	645,391	1E-37	*O. melastigma (XP_023819472.1)*
TRINITY_DN106037_c0_g1_i2	**bile salt-activated lipase-like**	16,05531	1,63E-24	79,548	9E-60	*O. melastigma (XP_024141800.1)*
TRINITY_DN123304_c1_g7_i4	**S100P-binding protein isoform X3**	15,86269	4,75E-24	67,326	5E-71	*O. latipes (4AN31188.1)*
TRINITY_DN118889_c0_g1_i1	**protein tyrosine phosphatase domain-containing protein 1-like**	15,70284	1,15E-23	33,092	3E-40	*A. ocellaris (XP_023141066.1)*
TRINITY_DN110493_c1_g1_i3	**transcriptional activator Myb-like isoform X2**	15,64552	1,61E-23	42,794	1E-74	*O. melastigma (XP_024141756.1)*
TRINITY_DN109074_c7_g1_i4	**beta/gamma crystallin domain-containing protein 1-like isoform X2**	15,61138	1,96E-23	105,576	7E-38	*O. melastigma (XP_023819472.1)*
TRINITY_DN160098_c0_g1_i1	**Tubulin beta chain**	15,53135	3,13E-23	359,296	0.0	*O. melastigma (XP_024120802.1)*
TRINITY_DN123460_c0_g1_i1	**dynein intermediate chain 1, axonemal isoform X1**	15,46714	4,62E-23	25,439	8E-20	*O. latipes (XP_011472519.2)*
TRINITY_DN110462_c0_g1_i3	**putative E3 ubiquitin-protein ligase UNKL isoform X2**	15,40392	6,69E-23	143,189	3E-110	*A. schlegelii (ACH72079.1)*
TRINITY_DN106887_c1_g6_i4	**zinc finger cchc domain-containing protein 7-like**	15,39539	7E-23	85,762	0.0	*O. melastigma (XP_024149962.1)*
TRINITY_DN103020_c0_g2_i2	**uncharacterized protein C6orf118 homolog isoform X1**	15,37681	7,66E-23	58,31	0.0	*O. melastigma (XP_024122939.1)*
TRINITY_DN109767_c0_g1_i1	**la-related protein 6-like isoform X1**	15,25118	1,65E-22	43,036	0.0	*O. melastigma (XP_024129599.1)*
TRINITY_DN113024_c0_g1_i3	**rap1 GTPase-GDP dissociation stimulator 1 isoform X1**	15,21708	2E-22	19,395	0.0	*O. melastigma (XP_024136158.11)*
TRINITY_DN61114_c0_g1_i1	**protein phosphatase inhibitor 2**	15,18308	2,46E-22	334,976	3E-31	*O. latipes (XP_011486131.2)*
TRINITY_DN117564_c0_g1_i2	**armadillo repeat-containing protein 4**	15,17376	2,61E-22	22,737	5E-66	*O. melastigma (XP_024127244.1)*
TRINITY_DN100664_c0_g1_i1	**outer dynein arm protein 1-like**	15,16248	2,79E-22	44,387	0.0	*O. latipes (XP_023817529.1)*

Table 10: Top 20 differentially expressed genes from Testis transcriptome.

*(false discovery rate ;FDR and log Fold change; logFC)

3.3.2 Male-specific genes from transcriptome analysis

Male specific genes were identified by the elimination of the transcripts that have expression in all other tissues except testis transcriptome by using TMM matrix data.

500 genes from male-specific genes were analyzed in accordance with annotation results; morn repeat-containing proteins and tata-box binding proteins were distinctively abundant compared to the other male-biased gene data. E3 ubiquitin-protein ligase proteins were abundant and these genes were also represented in top 20 genes accordingly (Table 11).

Transcript ID	Blast Hit	logFC	FDR	TMM	e-value	Species Blast
TRINITY_DN92991_c0_g2_i2	**No hit**	19,53853	3,26E-33	7326,39	-	-
TRINITY_DN114301_c0_g1_i2	**trypsin-1-like**	14,94104	1,05E-21	442,852	3,00E-55	*O. melastigma (XP_024123779.1)*
TRINITY_DN120574_c1_g7_i1	**fibronectin type III domain-containing protein 11-like isoform X2**	12,91753	3,07E-16	211,532	7,00E-41	*O. melastigma (XP_024125749.1)*
TRINITY_DN120864_c0_g4_i1	**ATPase family AAA domain-containing protein 2-like**	12,41237	7,25E-15	170,479	4,00E-35	*O. melastigma (XP_024154193.1)*
TRINITY_DN144328_c0_g1_i1	**E3 ubiquitin-protein ligase pellino homolog 1**	13,92783	5,79E-19	167,677	3,00E-67	*O. melastigma (XP_004077044.1)*
TRINITY_DN104667_c0_g2_i1	**E3 ubiquitin-protein ligase ZNRF1-like**	14,05452	2,65E-19	162,906	2,00E-37	*O. melastigma (XP_024137825.1)*
TRINITY_DN110462_c0_g1_i3	**putative E3 ubiquitin-protein ligase UNKL isoform X2**	15,40392	6,69E-23	143,189	7,00E-132	*O. melastigma (XP_024128185.1)*
TRINITY_DN68232_c0_g2_i1	**protein kintoun isoform X1**	13,79867	1,3E-18	127,722	5,00E-84	*O. melastigma (XP_024131672.1)*
TRINITY_DN41118_c0_g1_i1	**protein tyrosine phosphatase domain-containing protein 1-like**	12,39107	8,3E-15	120,113	2,00E-35	*O. melastigma (XP_024149236.1)*
TRINITY_DN101212_c0_g3_i1	**TATA-box-binding protein 1-like isoform X3**	12,035	7,46E-14	117,245	1,00E-24	*O. melastigma (XP_024113842.1)*
TRINITY_DN123669_c0_g1_i1	**histone-lysine N-methyltransferase SETD1B-A-like isoform X1**	12,02949	7,72E-14	116,718	1,00E-17	*O. melastigma (XP_024131428.1)*
TRINITY_DN114746_c0_g1_i2	**zinc finger MYND domain-containing protein 12**	14,23098	9,11E-20	106,781	3,00E-98	*O. melastigma (XP_024148695.1)*
TRINITY_DN87714_c0_g2_i2	**No hit**	12,84417	4,86E-16	104,849	-	-
TRINITY_DN189325_c0_g1_i1	**E3 ubiquitin-protein ligase RNF19B-like**	12,38245	8,74E-15	104,456	6,00E-09	*O. melastigma(XP_024144212.1)*
TRINITY_DN73723_c0_g1_i1	**DC-STAMP domain-containing protein 2 isoform X1**	12,62624	1,92E-15	86,586	3,00E-29	*O. melastigma (XP_020566182.1)*
TRINITY_DN106887_c1_g6_i4	**zinc finger CCHC domain-containing protein 7-like**	15,39539	7E-23	85,762	3,00E-54	*O. melastigma (XP_024145089.1)*
TRINITY_DN32576_c0_g1_i1	**coiled-coil domain-containing protein 38**	11,38272	4,05E-12	77,894	1,00E-06	*O. latipes (XP_020564554.1)*
TRINITY_DN202558_c0_g1_i1	**enkurin isoform X2**	11,95884	1,2E-13	76,587	3,00E-32	*O. melastigma (XP_024136276.1)*
TRINITY_DN49595_c0_g1_i1	**GTPase IMAP family member 8-like isoform X2**	11,13355	1,82E-11	76,201	6,00E-07	*O. melastigma (XP_024143009.1)*
TRINITY_DN110658_c0_g1_i3	**MORN repeat-containing protein 5 isoform X1**	12,89197	3,61E-16	75,33	3,00E-39	*O. melastigma(XP_023816414.1)*

Table 11: Top 20 Male-specific genes.

*Combined data were lined up according to the highest TMM expression values which expressed only in testis transcriptome.

4.DISCUSSION

One of the predominantly expressed genes in testis transcriptome is estimated as Trypsin which is known to be a key factor in the control of spermatogenesis. Furthermore, trypsin was detectable in the membranes of the spermatozoa and found to be associated with fertilization in fish (Miura et al., 2009).

According to a recent study elastase-1 also expressed in testis transcriptome (DEG analysis) in *Oryzias melastigma* (Fong et al., 2014). Elastase has been shown to disrupt tight junctions, cause damage to tissue complement system, and elastin metabolism was modulated by reproductive hormones (Chen et al., 2005).

Sperm-associated antigen 6 (*SPAG6*), which has been shown to be a critical protein in either the assembly or structural integrity of the sperm tail axoneme. It is shown that it also expressed in testis of 2-year-old adult yellow catfish (Lu et al., 2014).

Protein kintoun which shows abundance in testis transcriptome is a cytoplasmic protein which is required for dynein preassembly function of motile cilia and it is one of the predominant genes found in male-specific transcriptome analysis (Omran et al., 2008).

One of the common and abundant genes found in testis transcriptome was E3 ubiquitin-protein ligase proteins. While E3 ligase is related with nerve regeneration it is shown that ubiquitin protein ligases are functioning during spermatogenesis which is inducing histones when they must be degraded in

early elongating spermatids to permit chromatin condensation (Liu et al., 2005).

In DEG and TMM/DEG combined table pancreatic alpha-amylase-like protein was identified which is α-Amylase, a major pancreatic protein and starch hydrolase, also essential for energy acquisition (Date et al., 2015).

Top 20 female-enriched genes and top 20 DEG/TMM expressions indicates a significant protein; Polysialoglycoproteins (PSGP) which were predominantly expressed in female-biased genes and they are found to be a ubiquitous component of Salmonidae fish eggs which are a novel type of glycoprotein. (Kitajima et al., 1986)

While female-specific genes data shows zona pellucida genes predominantly, it's also rich with choriogenin H-related protein which was defined as a precursor protein of the inner layer subunits of egg envelope (chorion) of the teleost fish, *O. latipes* (Murata et al., 1997).

One of the abundant transcripts from top 20 differentially expressed genes from ovary was estimated from zona pellucida glycoprotein family (ZP1, ZP2 and ZP3) which are cell surface proteins that triggers fertilization. (Table 7) They are specialized extracellular matrix layer surrounding the developing oocyte within each follicle within the ovary and the layer is comprised of the secretions from the follicle granulosa cells and the oocyte. The zona pellucida has many different roles including in fertilization (Bleil & Wassarman, 1980) oocyte development (Epifano et al., 1995), protection during growth and transport (Murayama et. al, 2006), spermatozoa binding (Huang et al., 1981), preventing polyspermy (Burkart et al., 2012), blastocyst development (Barnes et al., 1995), and preventing premature implantation (Cohen et al.,1990). A search of GeneBank database revealed that the zona pellucida amino acid

sequences derived from *O. dancena* transcriptome analysis were homologous to that of the egg envelope glycoprotein ZP3 isolated from *Oryzias melastigma* and *Amphiprion ocellaris.*

Furthermore, zonadhesin, confers species-specificity to sperm-ZP adhesion which may be interpreted as in the ovary transcriptome analysis possibly indicates the close interactions between zona pellucida and zonadhesin during fertilization process. (Tardif et al., 2010) (Table 4-7-8)

Ovary transcriptome was rich with transcription factor IIIA that was expressed in several species ovaries such as *Xenopus laevis* oocytes. (Romaniuk, 1985) (Table 7).

Some genes like "sry-box containing gene partial" showing differential expressions in both gonad tissues while it is a known sox protein known for residing on the Y-chromosome (Whitfield et al., 1993)

Considering all testis transcriptome analysis there is an abundance of Cilia- and flagella-associated proteins, la-related protein and septin proteins which are involved in similar mechanisms (Peterson et al., 2007).

La-related protein 4B associates with poly(A)-binding protein and is required for male fertility and syncytial embryo development (Blagden et. al, 2009) It's isoforms also shown to be expressed in Amazon molly in its gonadal transcriptome (Schedina et al., 2018).

One of the salient results between ovary and testis analysis was explicit expressions of s100P family occurred in both DEG analysis. There were 5 abundant genes in ovary were estimated related with this protein family; 3 clones of calcium-binding protein P-like and hornerin-like isoform X3. Also in testis; s100P-binding protein isoform X3 was estimated to be one of the top

DEGs. This can be explicated with a possible similarity between both (ovary and testis) gonadal mechanisms in *O. dancena* may have similar functioning genes through gametogenesis, developmental process or immune response. Further experiments needed to be able to understand similar mechanisms behind mutual functions during fertilization or metabolic process in *O. dancena.*

One of the differentially expressed genes which were predominant in transcriptome analysis in ovary of *O. dancena* was estimated as hornerin (*HRNR*) which is also an s100 family protein. s100 is a part of a super family of Ca2± binding proteins which are known to be involved in several normal and pathological cell functions including inflammatory and immune responses, Ca2± homeostasis, protein phosphorylation and enzyme activity, gene transcription, the dynamics of cytoskeleton constituents and also in cell proliferation or differentiation (Jiang et al., 2011).

Sex-specific natural selection favors traits that increase the survival or general reproductive success of individuals of the respective sex, whereas sexual selection favors traits involved specifically in mating (or fertilization) success. This includes traits that are relevant to within-sex competition, such as male–male or sperm competition, as well as those related to mating preference, such as female mate choice. In most other species, the male and female genomes differ by only a few genes located on sex-specific chromosomes (such as the Y chromosome of mammals). This implies that the vast majority of sexually dimorphic traits result from the differential expression of genes that are present in both sexes (Ellegren & Parsch, 2007).

The study of sex-biased gene expression is relevant to several biological disciplines. Elucidating the molecular mechanisms that lead to sex-biased

gene expression is essential to understanding gene regulation, epigenetics, and developmental biology. Because natural and sexual selection may act differently on females and males, the study of sex-biased gene expression is also of great interest to evolutionary biologists and may provide insight into the evolution of genes, genomes, and sex chromosomes (Grath & Parsch, 2016).

RNA-Seq analysis should provide a wide range of data that, in conjunction with genetic and epigenetic data, will help elucidate the regulatory mechanisms controlling sex-biased expression. Although almost all previous expression studies have been conducted at the RNA level, high-throughput proteomic methods, such as quantitative mass spectrometry, will allow sex-biased expression to be determined at the protein level, which can reveal the effects of post-transcriptional regulation.

5.ACKNOWLEDGEMENTS

As my advisor, Prof. Yoon Kwon Nam deserves thanks for many things. Notably, for creating the research environment in which I have performed my graduate studies. I would like to thank him for his guidance which made my maturation as a scientist, his guidance, continues support for my studies, his patience has taught me a great deal in scientific research and life. I feel honored to join his team for two primary reasons: the areas of research in which he was involved and his disciplined and ethical approach to scientific studies. He has provided guidance at key moments in my study while also allowing me to work independently the majority of the time. I will be always grateful to him for accepting me as his student.

Besides my advisor, I would like to give my respect and many thanks to the members of the thesis committee, Prof. Dong Soo Kim and Prof. Seung Pyo Gong for sparing their precious time and review this thesis.

I would like to express my sincere gratitude to Dr. Chan Hee Kim who have influenced and enhanced my research, shared his immense knowledge, spare his precious time and patience. I feel honored to have a chance to learn many things from him about science and it would not be possible to conduct this research without his precious support.

I would like to give my respect and sincere gratitude to my former advisor in Ankara University, Prof. Hasan Hüseyin Atar. who has been helpful in providing advice many times and supporting me to come and study in Korea. And I would like to thank my former advisor in Pukyong National University, Prof. Sung Koo Kim who excepted me in his laboratory. I am thankful for his support and advices during my studies.

I would like to thank Mr. Ji for his help and guidance who was working for ICFO (International Cooperative Fishery Organization) which is the scholarship that made my studies possible in Korea.

I have crossed paths with many graduate students. I would like to thank to my lab friends Berkay, Eun Jong, Jiyeon, Jin Woo and Hyeon Jong for their kindness and for helping me during our studies.

I would like to thank my parents, for encouraging me, believing in me and supporting my curiosity to be a scientist since the early days of my life. And my little brother, for reminding me who I was whenever I feel lost.

One's life has value so long as one attributes value to the life of others, by means of love, friendship, indignation and compassion. Some special words of gratitude goes to my friends who have been a major source of support when things would get a bit discouraging: Nagi, Ülgen, Özgün, Dave, Boris, Steph, Wyatt, Megan and Murat, thank you for all the moments we've shared and things that I've learned from our friendship.

6. REFERENCES

Apweiler, Rolf, et al. "UniProt: the universal protein knowledgebase." Nucleic acids research 32.suppl_1 (2004): D115-D119.

Barnes, Frank L., et al. "Blastocyst development and birth after in-vitro maturation of human primary oocytes, intracytoplasmic sperm injection and assisted hatching." Human Reproduction 10.12 (1995): 3243-3247.

Bassett Jr, Douglas E., Michael B. Eisen, and Mark S. Boguski. "Gene expression informatics—it's all in your mine." Nature genetics 21.1s (1999): 51.

Bizuayehu, Teshome T., et al. "Sex-biased miRNA expression in Atlantic halibut (*Hippoglossus hippoglossus*) brain and gonads." Sexual Development 6.5 (2012): 257-266.

Blagden, Sarah P., et al. "Drosophila Larp associates with poly (A)-binding protein and is required for male fertility and syncytial embryo development." Developmental biology 334.1 (2009): 186-197.

Bleil, Jeffrey D., and Paul M. Wassarman. "Mammalian sperm-egg interaction: identification of a glycoprotein in mouse egg zonae pellucidae possessing receptor activity for sperm." Cell 20.3 (1980): 873-882.

Burkart, Anna D., et al. "Ovastacin, a cortical granule protease, cleaves ZP2 in the zona pellucida to prevent polyspermy." J Cell Biol 197.1 (2012): 37-44.

Chen, Bertha, et al. "Elastin metabolism in pelvic tissues: is it modulated by reproductive hormones." American journal of obstetrics and gynecology 192.5 (2005): 1605-1613.

Cho, Young Sun, and Yoon Kwon Nam. "Transgene chgH-rfp expression at developmental stages and reproductive status in marine medaka (*Oryzias dancena*)." Fisheries and Aquatic Sciences 19.1 (2016): 41.

Cho, Young Sun, et al. "Functional ability of cytoskeletal β-actin regulator to drive constitutive and ubiquitous expression of a fluorescent reporter throughout the life cycle of transgenic marine medaka *Oryzias dancena*." Transgenic research 20.6 (2011): 1333-1355.

Cohen, Jacques, et al. "Impairment of the hatching process following IVF in the human and improvement of implantation by assisting hatching using micromanipulation." Human Reproduction 5.1 (1990): 7-13.

Contractor, Rooha G., et al. "Evidence of gender-and tissue-specific promoter methylation and the potential for ethinylestradiol-induced changes in Japanese medaka (*Oryzias latipes*) estrogen receptor and aromatase genes." Journal of Toxicology and Environmental Health, Part A 67.1 (2004): 1-22.

Date, Kimie, et al. "Pancreatic α-amylase controls glucose assimilation by duodenal retrieval through N-glycan-specific binding, endocytosis, and degradation." Journal of Biological Chemistry (2015): jbc-M114.

Ellegren, Hans, and John Parsch. "The evolution of sex-biased genes and sex-biased gene expression." Nature Reviews Genetics 8.9 (2007): 689.

Epifano, Olga, et al. "Coordinate expression of the three zona pellucida genes during mouse oogenesis." Development 121.7 (1995): 1947-1956.

Finn, Robert D., et al. "Pfam: the protein families database." Nucleic acids research 42.D1 (2013): D222-D230.

Fong, C. C., et al. "iTRAQ-based proteomic profiling of the marine medaka (*Oryzias melastigma*) gonad exposed to BDE-47." Marine pollution bulletin 85.2 (2014): 471-478.

Gong, Zhiyuan, et al. "Development of transgenic fish for ornamental and bioreactor by strong expression of fluorescent proteins in the skeletal muscle." Biochemical and Biophysical Research Communications 308.1 (2003): 58-63.

Grabherr, Manfred G., et al. "Full-length transcriptome assembly from RNA-Seq data without a reference genome." Nature biotechnology 29.7 (2011): 644.

Grath, Sonja, and John Parsch. "Sex-biased gene expression." Annual Review of Genetics 50 (2016): 29-44.

Haas, Brian J., et al. "De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis." Nature protocols 8.8 (2013): 1494

Huang, T. T. F., A. D. Fleming, and R. Yanagimachi. "Only acrosome-reacted spermatozoa can bind to and penetrate zona pellucida: A study using the guinea pig." Journal of Experimental Zoology 217.2 (1981): 287-290.

Inoue, Koji, and Yoshio Takei. "Asian medaka fishes offer new models for studying mechanisms of seawater adaptation." Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 136.4 (2003): 635-645.

Jiang, Hongfei, et al. "Calcium-binding protein S100P and cancer: mechanisms and clinical relevance." Journal of cancer research and clinical oncology 138.1 (2012): 1-9.

Jin, Wei, et al. "The contributions of sex, genotype and age to transcriptional variance in *Drosophila melanogaster*." Nature genetics 29.4 (2001): 389.

Kanehisa, Minoru, et al. "The KEGG resource for deciphering the genome." Nucleic acids research 32.suppl_1 (2004): D277-D280.

Kitajima, K., Y. Inoue, and S. Inoue. "Polysialoglycoproteins of Salmonidae fish eggs. Complete structure of 200-kDa polysialoglycoprotein from the unfertilized eggs of rainbow trout (*Salmo gairdneri*)." Journal of Biological Chemistry 261.12 (1986): 5262-5269.

Lau, Karen, et al. "Identification and expression profiling of microRNAs in the brain, liver and gonads of marine medaka (*Oryzias melastigma*) and in response to hypoxia." PloS one 9.10 (2014): e110698.

Lee, Sang Yoon, Dong Soo Kim, and Yoon Kwon Nam. "Gene structure and estrogen-responsive mRNA expression of a novel choriogenin H isoform from a marine medaka *Oryzias dancena*." Fisheries and aquatic sciences 15.3 (2012): 221-231.

Liu, Zhiqian, Rose Oughtred, and Simon S. Wing. "Characterization of E3Histone, a novel testis ubiquitin protein ligase which ubiquitinates histones." Molecular and cellular biology 25.7 (2005): 2819-2831.

Lu, Jianguo, et al. "Gonadal transcriptomic analysis of yellow catfish (*Pelteobagrus fulvidraco*): identification of sex-related genes and genetic markers." Physiological genomics 46.21 (2014): 798-807.

Margulies, Marcel, et al. "Genome sequencing in microfabricated high-density picolitre reactors." Nature 437.7057 (2005): 376..

Miura, Chiemi, et al. "Trypsin is a multifunctional factor in spermatogenesis." Proceedings of the National Academy of Sciences 106.49 (2009): 20972-20977.

Murata, Kenji, et al. "Cloning of cDNA and estrogen-induced hepatic gene expression for choriogenin H, a precursor protein of the fish egg envelope (chorion)." Proceedings of the National Academy of Sciences 94.5 (1997): 2050-2055.

Murayama, Yoshinobu, et al. "Mouse zona pellucida dynamically changes its elasticity during oocyte maturation, fertilization and early embryo development." Human Cell 19.4 (2006): 119-125.

Omran, Heymut, et al. "Ktu/PF13 is required for cytoplasmic pre-assembly of axonemal dyneins." Nature 456.7222 (2008): 611.

Parsch, John, and Hans Ellegren. "The evolutionary causes and consequences of sex-biased gene expression." Nature Reviews Genetics 14.2 (2013): 83.

Pearson, William R. "An introduction to sequence similarity ("homology") searching." Current protocols in bioinformatics 42.1 (2013): 3-1.

Peterson, Esther A., et al. "Characterization of a SEPT9 interacting protein, SEPT14, a novel testis-specific septin." Mammalian Genome 18.11 (2007): 796-807.

Pillai, Suja, Vinod Gopalan, and Alfred King-Yin Lam. "Review of sequencing platforms and their applications in phaeochromocytoma and paragangliomas." Critical reviews in oncology/hematology 116 (2017): 58-67.

Qian, Xi, et al. "RNA-Seq technology and its application in fish transcriptomics." Omics: a journal of integrative biology 18.2 (2014): 98-110.

Romaniuk, Paul J. "Characterization of the RNA binding properties of transcription factor IIIA of *Xenopus laevis* oocytes." Nucleic acids research 13.14 (1985): 5369-5387.

Schedina, Ina Maria, et al. "The gonadal transcriptome of the unisexual Amazon molly *Poecilia formosa* in comparison to its sexual ancestors, *Poecilia mexicana* and *Poecilia latipinna*." BMC genomics 19.1 (2018): 12.

Shibata, Yasushi, et al. "Expression of gonadal soma derived factor (GSDF) is spatially and temporally correlated with early testicular differentiation in medaka." Gene Expression Patterns 10.6 (2010): 283-289.

Takehana, Yusuke, et al. "Co-option of Sox3 as the male-determining factor on the Y chromosome in the fish *Oryzias dancena*." Nature communications 5 (2014): 4157.

Tanaka, Minoru, et al. "Establishment of medaka (*Oryzias latipes*) transgenic lines with the expression of green fluorescent protein fluorescence exclusively in germ cells: a useful model to monitor germ cells in a live vertebrate." Proceedings of the National Academy of Sciences 98.5 (2001): 2544-2549.

Tardif, Steve, et al. "Zonadhesin is essential for species specificity of sperm adhesion to the egg's zona pellucida." Journal of Biological Chemistry (2010): jbc-M110.

Teng, Mingxiang, et al. "A benchmark for RNA-seq quantification pipelines." Genome biology 17.1 (2016): 74.

Whitfield, L. Simon, Robin Lovell-Badge, and Peter N. Goodfellow. "Rapid sequence evolution of the mammalian sex-determining gene SRY." Nature 364.6439 (1993): 713.

Xiao, Sheng-Jian, et al. "TiSGeD: a database for tissue-specific genes." Bioinformatics 26.9 (2010): 1273-1275.