The globe transcriptome analysis reveals hybrid effects and regulatory divergence between (*Acanthopagrus schlegelii* × *Pagrus major*) hybrid F1 and parents

Shuyin Chen

Marine Fisheries Research Institute of Jiangsu Province

Zhiyong Zhang

Marine Fisheries Research Institute of Jiangsu Province

Hongjiu Ji

Marine Fisheries Research Institute of Jiangsu Province

Shixia Xu

Nanjing Normal university

Yunxia Yang (✉ yangyx@zjou.edu.cn)

Zhejiang Ocean University

Chaofeng Jia

Marine Fisheries Research Institute of Jiangsu Province

Fei Zhu

Marine Fisheries Research Institute of Jiangsu Province

Qian Meng

Marine Fisheries Research Institute of Jiangsu Province

Ruijian Sun

Marine Fisheries Research Institute of Jiangsu Province

Zhiwei Zhang

Marine Fisheries Research Institute of Jiangsu Province

Yongchao Zhao

Marine Fisheries Research Institute of Jiangsu Province

Research article

Keywords: Hybrid; Transcriptome; Differentially expressed genes

Posted Date: August 23rd, 2019

DOI: <https://doi.org/10.21203/rs.2.13504/v1>

Abstract

Background: Hybridization in two related homoploid species can mix the genotypes which have the potential for good environmental adaptation conditions or rapid growth condition. Since hybridization technique has been used in other animals to produce more rapidly growing individuals, it may be useful in commercial fish culture. The nutritional composition and rapid growth rate of *P. major*, *A. schlegelii* and hybrid F1 had been researched. However, gene expressions of these fish are less well studied. *de novo* assembly RNA-Seq is a better approach to understanding transcriptomes of animals and can offer much data with better coverage and sufficient sequence depth for transcriptomes. Results: In order to facilitate molecular research in parents and hybrids F1, we characterized the whole body transcriptome for identification of transcripts involved in growth, metabolism and immune since several attributes were coming from different samples (*Acanthopagrus schlegelii*, *Pagrus major* and hybrid F1 *Acanthopagrus schlegelii* × *Pagrus major*). We analyzed the transcriptome of As, Pm and hybrid F1 AP using Illumina high throughput sequencing platform. Our results suggested that physiological index of organism development we observed is consistent to the gene express in transcriptome. We have also compared the expression levels of transcripts in parents and hybrids F1. The results showed that AP had a quickly growth and strong resistant than parents Pm and As. Conclusion: This is the first research on transcriptome of As, Pm and hybrid F1 at one month age. And it will provide an important basic molecular data for studying molecular mechanisms involved in different sample.

Background

Hybridization in two related homoploid species can mix the genotypes which have the potential for good environmental adaptation conditions or rapid growth condition [1]. These situations have been confirmed in some plants and animals [2,3], including in fish [4]. Since hybridization technique has been used in other animals to produce more rapidly growing individuals, it may be useful in commercial fish culture [5]. And the ability of some hybrid fishes to grow more rapidly than the parent species has been reported by Ricker [6], Smirnov [7], Buss and Wright [8], and Garside and Christie [9].

Black seabream (*Acanthopagrus schlegelii*) (Bleeker, 1854) is one kind of marine carnivorous fish. It has high aquaculture potential because of its high economic value and excellent meat quality. Also it has strong ability to tolerate a wide range of environmental conditions, high stocking densities and strong resistance to diseases [10,11]. From the previous research, it is highly adaptable to a wide range of water temperatures and salinities [12,13]. But the long rearing period to market size and poor resistance to disease is the defect. Red seabream, *Pagrus major* (Temminck and Schlegel, 1843), another most important species in fisheries, has the advantages of large size, fast growth and high disease resistance [14], but it is susceptible to salinity and water temperature change. Hybrid progeny can obtain good parental characters through crossbreeding [15]. The hybrid F1 seabream (*Pagrus major* × *Acanthopagrus schlegelii*) has the rapid growth rate and good muscle nutritional composition [16]. The experiment of hybrid F1 showed that the hybrid progeny had better growth speed, hypoxia tolerance and stress resistance than its parents [17–20].

The nutritional composition and rapid growth rate of *P. major*, *A. schlegelii* and hybrid F1 had been researched. However, gene expressions of these fish are less well studied. *de novo* assembly RNA-Seq is a better approach to understanding transcriptomes of animals and can offer much data with better coverage and sufficient sequence depth for transcriptomes. In this study, we report the *de novo* assembly transcriptome of *A. schlegelii* (As), *P. major* (Pm) and hybrid F1 (*A. schlegelii* × *P. major*, AP) for the first time. The high-quality reads were assembled into unique transcripts. Then these transcripts were annotated using different database to identify the putative pathways and genes responsible for their body properties.

Results

Illumina paired-end sequencing and *De novo* assembly

The fish were used as sequencing sample in biological duplicates. The pre-processing of raw reads was done for removal of adaptor sequences and low-quality reads and the average clean reads were 47.21M ± 7.05M (mean ± SD). After quality trimming, 97.37% ± 0.52% of reads were retained, indicating a high quality data set (>85% reads with ≥Q30). The GC contents were 50.02% ± 0.28%. The final assembly consisted of 151,982, 166,373, 216,483 transcripts and the assembled transcripts were further clustered into 120,855, 129,258, 155,311 unigenes with As, Pm and AP samples. The transcriptome assembly details are given in Table 1.

Functional annotation of unigenes

We annotated unigenes by NR, NT, KO, SwissPort, PFAM, GO, KOG databases. The detail information are showed in Table 2. NR and NT is the non-redundant NCBI collection and protein sequence database. And the Nt database can align more nucleic acid in this study. Swiss-Prot is a curated protein sequence database which strives to provide a high level of annotation (such as the description of the function of a protein, its domain structure, post-translational modifications, variants, etc), a minimal level of redundancy and high level of integration with other databases. And this database successfully annotated the protein also contain over 20%. Of these, 34008, 30691 and 31770 were successfully GO mapped. GO terms for biological processes, molecular function and cellular components were all highly represented in annotated genes (Fig. 1).

Also the KEGG database was used to analyze the the active biological pathways. These assembled unigenes were assigned to 32 biological pathways through this process. Among them, the highest pathway assigned to the unigenes is Signal transduction, followed by Endocrine system, Immune system, Nervous system, Signaling molecules and interaction, Cellular commiunity, Transport and catabolism and Digestive system (Fig. 2).

Lastly, the top-hit species distribution was dominated by *Larimichthys crocea* with over 18380 (Pm), 17746 (As) and 20797 (AP) hits, and *Stegastes partitus*. Other species included *Notothenia coriiceps*,

Oreochromis niloticus, and *Haplochromis burtoni*, which were the third, fourth and fifth most highly represented species, respectively (Fig. 3).

Differential gene expression analysis

The estimation of transcript abundance or expression levels of *de novo* assembled unigenes from each sample transcriptome was calculated based on FPKM values using RSEM. And the DEGs analysis were used to compare the hybrid F1 with As and Pm transcriptome: AP vs As and AP vs Pm. The results showed that the genes expressed in different groups had significantly different (Fig. 4). There are 13,441 and 10,075 genes significantly expressed in two groups (Fig. 5). GO and KEGG were used to analyse the function of genes.

In AP vs As group, most DEGs belonged to chemokine activity, chemokine receptor binding, carbohydrate binding, G-protein coupled receptor binding, cytokine receptor binding, cytokine activity, nucleic acid binding, immune system process, immune response, and so on (Fig. 6). It indicated that in hybrid F1 and parent, immune system had changed. In AP vs Pm group, these metabolic process, cellular metabolic process, catalytic activity, nitrogen compound metabolic process, intracellular part, organic cyclic compound binding, and so on (Fig. 6), indicated that hybrid F1 compared with its parent, had a strong development process.

Discussion

During the development phase of larval fish, many biological processes, such as differentiation, cellular proliferation, growth, can be regulated to obtain the phenotype of an adult fish [21]. These regulations are carried out through the expression of genes involved in larval ontogenesis processes. Gene expression analysis can be comprehensively analyze through transcriptome using high throughput RNA-Seq tool. Previous physiologic studies showed that hybrid F1 seabream (*Pagrus major* × *Acanthopagrus schlegelii*) grow quickly, with enhanced muscle nutritional composition [16–19]. However, the transcriptome information about the larval As, Pm and F1 are lacked. In the present study, we presented the main findings from transcriptomic investigations of As, Pm and AP during postembryonic (one month age) development²¹. From the results of RNA-Seq, the whole data is nearly 6G of each sample (Table 1). The depth of sequencing is enough and the quantity of our sequencing data are convincing for next analysis.

We used the unique transcripts as the database to analysis the development processes of each sample. From the gene function annotation, KEGG results showed that most genes belongs to signal transduction, endocrine system, immune system, nervous system, signaling molecules and interaction, and so on, in each sample (Fig. 2). The molecular results were consistent with physiological development. The multiple biological processes were combined with consistencies of results obtained from the annotated expressed genes in each sample, reveals the complexity of the transcriptome regulation during larval ontogenesis processes. The authors used the transcriptome analysis to show that more than two hundred genes

involved in visual pigmentation, metabolism, digestive function and epithelial development exhibited differential expression during the early development of gilthead sea bream [22]. From previous physiological research about As, Pm and AP samples, the fish began growing quickly, and the digestive system, circulatory system, and excretory system are gradually formed during the fourth day after hatching [23]. The transcriptome results were consistent with physiological research. Our results showed that the many genes belonged to endocrine system, immune system, nervous system development have expressed in each sample. As we all know, immune system is very important to juvenile fish. When the immune system formed, they can prevent the disease invading about environment change or viral infection. During the agricultural production, 25–35 days fish will be sold to farm. These fish need to face the different environment to leave off the incubation environment. In this period, the fish have good condition for transport far away. Our result supposed that the immune system of the fish is more active at this phase, and the fish body has a strong ability to resist the invasion of external pathogens. Besides, metabolism process, such as carbohydrate metabolism, lipid metabolism, amino acid metabolism, and nucleotide metabolism, all had the active gene-expression (Fig. 2).

For different species have individual difference in the uniform period, different expressed genes (DEGs) are the focus of our attention. The DEGs between offspring (AP) transcriptome and parent sample (As, Pm) transcriptome were compared. From the top 20 pathway, AP had a high gene expression in spliceosome, DNA replication, cell cycle, ribosome biogenesis in eukaryotes, parkinson's disease, p53 signaling pathway, ECM-receptor interaction and mismatch repair pathway than Pm (Table 3). These genes in each pathway most have high expression during the individual growth. It also suggested that the cell in fish had cell multiplication to make the body grow. Then, the high expressed genes in fat digestion and absorption and oxidative phosphorylation pathway suggested that the fish need more energy for growing. Also between As and AP compared, AP also has a quickly growth. During the early development, the cell quickly divided into more cells to make body growth. The genes belong to DNA replication and cell cycle pathway can show this situation. During the replication, it need ECM-receptor interaction and mismatch repair to make the organism keep the biological function, such as cell adhesion [24]. The developmental integrity of fish cells plays an important role in the resistance of the whole body to external viruses. Besides the quickly growth, there are much other information we can get from AP and As. There are 111 genes of viral carcinogenesis which had high expressed in AP than in As (Table 3). The high expressed genes showed a strong cell-proliferation ability in AP. As we know, tumor viruses can promote an abnormal cell-proliferation via modulating cellular cell-signaling pathways via expression of many potent oncoproteins [25–27]. Our results showed that 111 genes belonging to oncoproteins had high expression in AP. It can suppose that AP had stronger cell-proliferation than As, suggesting that AP maybe grow quickly than As. Another special pathway in AP compared with As, is that 106 genes of herpes simplex infection pathway had higher expression. Herpes simplex virus (HSV) is the main cause of herpes infections that lead to the formation of characteristic blistering lesion. HSV can interfere with host immune responses by expressing multiple viral accessory proteins [28]. Influenza A is a contagious respiratory disease caused by influenza virus infection, which had 91 genes highly expressed in AP sample than in As. Influenza virus has one kind of proteins which is a multifunctional virulence factor

that interfere IFN-mediated antiviral response. It can inhibit activation of transcription factors such as NF-kappa B [29]. Except the virus infect, the parasite also can disrupt the intestinal mucus layer of fish, followed by apoptosis of host epithelial cells. 82 genes belonging to amoebiasis pathway had higher expression in AP than in As. The ingestion of contaminated water and food can lead the pathogenesis of amoebiasis. Intestinal tissue destruction causes severe dysentery and ulcerations in amoebic colitis [30]. And several amoebic proteins, such as lectins, cysteine proteases, and amoebapores, are associated with the invasion process [31]. The fusion proteins have aberrant transcriptional function for fusion transcription factors alter expression of target genes, and thereby result in the tumorigenesis. They are all led by transcriptional misregulation. From the DEGs, There are 77 genes higher expressing in AP sample, belong to Transcriptional misregulation. During rearing, the juvenile fish often were attracted by virus in environment. These pathways maybe provided the challenge ability for fish to resist the environment. Also during these pathways, there were many immune related genes which had high express in fish. These results suggested that the AP, compared with As and Pm, has stronger immune system. The further studies about the As, Pm and AP immune response needs to be carried out in future.

Materials And Methods

Fish rearing and collection

The As (*A. schlegelii*), Pm (*P. major*) and hybrid F1 AP (*A. schlegelii* × *P. major*) fertilized eggs were incubated in the cage of cement pool at 22.4 °C, and the pH was about 8.2, and the hatched larvae were developed synchronously. They were temporarily maintained at the Jiangsu Province Seedling breeding and breeding Technology Center. The larvae fish were preserved using RNA later and stored at -80°C for RNA extraction.

RNA extraction and qualification

Total RNA were extracted using Trizol extraction method [32]. RNA quantification were detected by 1.5% agarose gels. The more detail information about RNA quantification, RNA purity, concentration and integrity were carried out. RNA purity was assessed using the NanoPhotometer® spectrophotometer (IMPLEN, CA, USA). Concentration was measured using Qubit® RNA Assay Kit in Qubit® 2.0 Fluorometer (Life Technologies, CA, USA). And Integrity was checked by the RNA Nano 6000 Assay Kit of the Agilent Bioanalyzer 2100 system (Agilent Technologies, CA, USA).

Library preparation for transcriptome sequencing

A total amount of 1.5 µg RNA per sample was used for sample preparations. Sequencing libraries were generated using NEBNext® Ultra™ RNA Library Prep Kit for Illumina® (NEB, USA) following manufacturer's recommendations. Briefly, mRNA was purified from total RNA using poly-T oligo-attached magnetic beads. Fragmentation was carried out using divalent cations under elevated temperature in NEBNext First

Strand Synthesis Reaction Buffer (5×). First strand cDNA was synthesized using random hexamer primer and M-MuLV Reverse Transcriptase (RNase H). Second strand cDNA synthesis was subsequently performed using DNA polymerase I and RNase H. Remaining overhangs were converted into blunt ends via exonuclease / polymerase activities. After adenylation of 3' ends of DNA fragments, NEBNext Adaptor with hairpin loop structure were ligated to prepare for hybridization. In order to select cDNA fragments of preferentially 150~200 bp in length, the library fragments were purified with AMPure XP system (Beckman Coulter, Beverly, USA). Then 3 µl USER Enzyme (NEB, USA) was used with size-selected, adaptor-ligated cDNA at 37°C for 15 min followed by 5 min at 95 °C before PCR. Then PCR was performed with Phusion High-Fidelity DNA polymerase, Universal PCR primers and Index (X) Primer. At last, PCR products were purified (AMPure XP system) and library quality was assessed on the Agilent Bioanalyzer 2100 system. The clustering of the index-coded samples was performed on a cBot Cluster Generation System using TruSeq PE Cluster Kit v3-cBot-HS (Illumina) according to the manufacturer's instructions (Novogene Experimental Department). After cluster generation, the library preparations were sequenced on an Illumina HiSeq platform and paired-end reads were generated.

Data analysis

Raw data (raw reads) of fastq format were firstly processed through in-house perl scripts. In this step, clean data (clean reads) were obtained by removing reads containing adapter, reads containing poly-N and low quality reads from raw data. At the same time, Q20, Q30, GC-content and sequence duplication level of the clean data were calculated. All the downstream analyses were based on clean data with high quality. Transcriptome assembly was accomplished using Trinity [33] by default and all other parameters set default.

Gene function was annotated based on the following databases: Nr (NCBI non-redundant protein sequences); Nt (NCBI non-redundant nucleotide sequences); Pfam (Protein family); KOG/COG (Clusters of Orthologous Groups of proteins); Swiss-Prot (A manually annotated and reviewed protein sequence database); KO (KEGG Ortholog database) and GO (Gene Ontology).

Differential expression analysis of two groups was performed using the DESeq R package (1.10.1) [34,35]. Gene expression levels were estimated by RSEM [36] for each sample. DESeq provide statistical routines for determining differential expression in digital gene expression data using a model based on the negative binomial distribution. The resulting P values were adjusted using the Benjamini and Hochberg's approach for controlling the false discovery rate. Genes with an adjusted P-value <0.05 found by DESeq were assigned as differentially expressed.

Also the Gene Ontology (GO) enrichment analysis of the differentially expressed genes (DEGs) was implemented by the Goseq R packages based Wallenius non-central hyper-geometric distribution [37] which can adjust for gene length bias in DEGs. KEGG [38] is a database resource for understanding high-level functions and utilities of the biological system, such as the cell, the organism and the ecosystem, from molecular-level information, especially large-scale molecular datasets generated by genome

sequencing and other high-throughput experimental technologies (<http://www.genome.jp/kegg/>). We used KOBAS [39] software to test the statistical enrichment of differential expression genes in KEGG pathways.

Ethical Statement

All samples and methods used in this study were conducted in accordance with the Laboratory Animal Management Principles of China. All experimental protocols in this study were approved by the Ethics Committee for Animal Experiments of the Institute of Jiangsu Province Seedling breeding and breeding Technology Center. All fish handling was performed under tricaine methanesulfonate (MS222) anesthesia.

Declarations

Authors' Contribution

S. C., Y. Y., Z. Z., S. X. designed the study. H. J., C. J., F. Z. performed dissections. C. J., F. Z., R. S., Z. Z., Q. M. performed the experiment. Y. Y., and Y. Z. analyzed the data. S. C. and Y. Y. wrote the manuscript. All authors reviewed the manuscript.

Funding

This research was supported by Project of Creation of Agricultural Major New Varieties of Jiangsu Province(PZC201744), Fund for Independent Innovation of Agricultural Science and Technology of Jiangsu Province(CX(17)2021) and National Science Foundation of Jiangsu Province (BK20181202).

Competing Interests

The authors declare no competing interests.

Reference

1. Stemshorn KC, Reed FA, Nolte AW, Tautz D. Rapid formation of distinct hybrid lineages after secondary contact of two fish species (cottus sp.). *Mol. Ecol.* 2011;20:1475–91. <https://doi.org/10.1111/j.1365-294X.2010.04997.x>.
2. Rieseberg LH. Hybrid origins of plant species. *Ann. Rev. Ecol. Syst.* 1997;28:359–89. <https://doi.org/10.1146/annurev.ecolsys.28.1.359>.
3. Gross BL, Rieseberg LH. The ecological genetics of homoploid hybrid speciation. *J. Heredity.* 2005;96:241–52. <https://doi.org/10.1093/jhered/esi026>.

4. Meyer A, Salzburger W, Schartl M. Hybrid origin of a swordtail species (Teleostei: Xiphophorus clemenciae) driven by sexual selection. *Mol. Ecol.* 2006;15:721–30. <https://doi.org/10.1111/j.1365-294X.2006.02810.x>.
5. John JG. Growth of a Blue X Channel Catfish Hybrid as Compared to its Parent Species. *The Progressive Fish-Culturist*. 1966;28:142–5. [https://doi.org/10.1577/1548-8640\(1966\)28\[142:GOABXC\]2.0.CO;2](https://doi.org/10.1577/1548-8640(1966)28[142:GOABXC]2.0.CO;2).
6. Ricker WE. Hybrid sunfish for stocking small ponds. *T. AM. Fish. Soc.* 1948;75:94–6.
7. Stairnov AL. Some characteristics of interspecific hybrids of autumn chums x pinks: *Oncorhynchus keta* (Walbaum) infrasp. *Autumnalis berg* x *O. gorbusha* (Walbaum) fam. *Salmonidae*. *Doklady Akademii Nauk SSSR*. 1953;91:29–33.
8. Buss K, James EW. Results of species hybridization within the family *Salmonidae*. *Progressive Fish-Culturist*. 1956;18:149–58.
9. Garside ET, Christie WJ. Experimental hybridization among three coregonine fishes. *T. AM. Fish. Soc.* 1962;91:196–200.
10. Hong WS, Zhang QY. Review of captive bred species and fry production of marine fish in China. *Aquaculture*. 2003;227:305–18. [https://doi.org/10.1016/S0044-8486\(03\)00511-8](https://doi.org/10.1016/S0044-8486(03)00511-8).
11. Shao Q, Ma J, Xu Z, Hu W, Xu J, Xie S. Dietary phosphorus requirement of juvenile black sea bream, *Sparus macrocephalus*. *Aquaculture*. 2008;277:92–100. <https://doi.org/10.1016/j.aquaculture.2008.01.029>.
12. Partridge GJ, Jenkins GI. The effect of salinity on growth and survival of juvenile black bream (*Acanthopagrus butcheri*). *Aquaculture*. 2002;210: 219–30. [https://doi.org/10.1016/S0044-8486\(01\)00817-1](https://doi.org/10.1016/S0044-8486(01)00817-1).
13. Jiang HL, Wu XF. Comparative study on temperature tolerance and growth between *Pagrus major* (Temminck et Schlegel) and *Spraus macrocephalus* and F1 hybrid. *Modern Fisheries Information* (in Chinese). 2009;24:27–29.
14. Bui H, Khosravi S, Fournier V, Herault M, Lee K. Growth performance, feed utilization, innate immunity, digestibility and disease resistance of juvenile red seabream (*Pagrus major*) fed diets supplemented with protein hydrolysates. *Aquaculture*. 2014;418–419:11–16. <https://doi.org/10.1016/j.aquaculture.2013.09.046>.
15. Bryden CA, Heath JW, Heath D. Performance and heterosis in farmed and wild Chinook salmon (*Oncorhynchus tshawytscha*) hybrid and purebred crosses. *Aquaculture*. 2004;235:249–61. <https://doi.org/10.1016/j.aquaculture.2004.01.027>.
16. Wang S, Li P, Chen S, Zhang Z, Yin S, Zhu F, et al. Lipid content and fatty acid profile of *Acanthopagrus schlegeli*, *Pagrosomus major* and their hybrids. *Food Science*. 2016;37:81–5.
17. Lin M, Miao L, Li M, Yu Y, Xu W. Chromosome composition and AFLP analysis of *Pagrus major*, *Spraus macrocephalus* and the hybrids [J]. *Acta Oceanologica Sinica*. 2014;36:117–21. <https://doi.org/10.3969/j.issn.0253-4193.2014.08.013>.

18. Jiang H, Wu X, Shen P. Hybrid breeding on *Pagrosomus major*×*Spraus macrocephalus*. *Fishery Modernization*. 2008;6:33–5. <https://doi.org/10.3969/j.issn.1007-9580.2008.06.009>.
19. Jiang J.. Preliminary studies on the anti-adversity in cultured *Pagrosomus Major*×*Sparus Macrocephalus* F1[D]. Ningbo University. 2011.
20. Park I. S. et al. Comparative of morphometric traits of hybrids between red sea bream (*Pagrus major*) and the black sea bream (*Acanthopagrus schlegelii*). *J Fish Sci Techol*. 2006;9:44–7. <https://doi.org/10.5657/fas.2006.9.1.044>.
21. Mazurais D, Darias M, Zambonino-Infante JL, Cahu C. Transcriptomics for understanding marine fish larval development 1. *Revue Canadienne De Zoologie*. 2011;89:599–611. <https://doi.org/info:doi/10.1139/z11-036>.
22. Sarropoulou E, Kotoulas G, Power D, Geisler R. Gene expression profiling of gilthead sea bream during early development and detection of stress-related genes by the application of cDNA microarray technology. *Physiol. Genomics*. 2005;23:182–91. <https://doi.org/doi:10.1152/physiolgenomics.00139.2005>.
23. Chen S, Zhang Z, Liu H, Jia C, Xu J, Zhu F, et al. Backcross Breeding and Observation of Early Larva Development in Red Bream *Pagrosomus major* and Black Bream *Acanthopagrus schlegelii*. *Fisheries Science*. 2017;36:693–99. <https://doi.org/10.16378/j.cnki.1003-1111.2017.06.001>.
24. Domínguezgiménez P, Brown NH, Martínbermudo MD. Integrin-ECM interactions regulate the changes in cell shape driving the morphogenesis of the *Drosophila* wing epithelium. *Journal of Cell Science*. 2007;120:1061–71. <https://doi.org/10.1242/jcs.03404>.
25. Raska I, Shaw PJ, Cmarko D. New insights into nucleolar architecture and activity. *Int Rev Cytol* 2006;255:177–235. [https://doi.org/10.1016/S0074-7696\(06\)55004-1](https://doi.org/10.1016/S0074-7696(06)55004-1).
26. Derenzini M, Montanaro L, Trere D. What the nucleolus says to a tumor pathologist. *Histopathology*. 2009;54:753–62, <https://doi.org/10.1111/j.1365-2559.2008.03168.x>.
27. Maggi LB, Weber JD. Nucleolar adaptation in human cancer. *Cancer Invest*. 2005;23:599–608. <https://doi.org/10.1080/07357900500283085>.
28. Hayward AR, Read GS, Cosyns M. Herpes simplex virus interferes with monocyte accessory cell function. *The Journal of Immunology*. 1993;150:190–96.
29. Corn RA, Aronica MA, Zhang F, Tong Y, Stanley SA, Kim SR, et al. T cell-intrinsic requirement for NF-kappa B induction in postdifferentiation IFN-gamma production and clonal expansion in a Th1 response. *J. Immun.* 2003;171:1816–24. <https://doi.org/10.4049/jimmunol.171.4.1816>.
30. Martínezpalomo A. The pathogenesis of amoebiasis. *Parasitology Today*. 1987;3:111–18. [https://doi.org/10.1016/0169-4758\(87\)90048-2](https://doi.org/10.1016/0169-4758(87)90048-2).
31. Bansal D, et al. An ex-vivo Human Intestinal Model to Study *Entamoeba histolytica* Pathogenesis. *Plos Neglected Tropical Diseases*. 2009;3:e551. <https://doi.org/10.1371/journal.pntd.0000551>.
32. Chen S, Zhang G, Shao C, Huang Q, Liu G, Zhang P, et al. Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nat. Genet.* 2014;46:253–60. <https://doi.org/10.1038/ng.2890>.

33. Grabherr MG, Hass BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nat Biotechnol.* 2011;29:644–52. <https://doi.org/10.1038/nbt.1883>.
34. Wang L, Feng Z, Wang X, Wang X, Zhang X. DEGseq: an R package for identifying differentially expressed genes from RNA-seq data. *Bioinformatics.* 2010;26:136–38. <https://doi.org/10.1093/bioinformatics/btp612>.
35. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome biol.* 2010;11:106. <https://doi.org/10.1186/gb-2010-11-10-r106>.
36. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. *BMC Bioinformatics.* 2011;12:323. <https://doi.org/10.1186/1471-2105-12-323>.
37. Young MD, Wakefield MJ, Smyth GK, Oshlack A. Gene ontology analysis for RNA-seq: accounting for selection bias. *Genome Biology.* 2010;11:14. <https://doi.org/10.1186/gb-2010-11-2-r14>.
38. Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, et al. KEGG for linking genomes to life and the environment. *Nucleic Acids Res.* 2008;36:480–4. <https://doi.org/10.1093/nar/gkm882>.
39. Mao X, Cai T, Olyarchuk JG, Wang L. Automated genome annotation and pathway identification using the KEGG Orthology (KO) as a controlled vocabulary. *Bioinformatics.* 2005;21:3787–93. <https://doi.org/10.2307/1592215>.

Tables

Table 1 Summary of Illumina paired-end sequencing and *de novo* assembly of each sample.

Sample	Raw Reads	Clean Reads	Clean Bases	Transcript	Unigene	Q30(%)	GC Content(%)
As_s1	53110260	46176972	6.93G			88.68	50.41
As_s2	52038966	43917894	6.59G	151982	120855	87.01	50.27
Pm_s1	70319018	61169348	9.18G			88.54	49.81
Pm_s2	51425784	43621038	6.54G	166373	129258	87.06	50.07
AP_h1	53419960	46523082	6.98G			88.7	49.68
AP_h2	49987854	41868274	6.28G	216483	155311	85.81	49.86

Table 2 Annotation by NR, NT, KO, SwissPort, PFAM, GO, KOG in each sample.

database	AP		As		Pm	
	Number of Unigenes	Percentage (%)	Number of Unigenes	Percentage (%)	Number of Unigenes	Percentage (%)
NR	43880	28.25	37335	30.89	40131	31.04
NT	54280	34.94	48132	39.82	50744	39.25
KO	20987	13.51	19331	15.99	20365	15.75
SwissProt	33418	21.51	30419	25.16	32434	25.09
PFAM	34022	21.90	30689	25.39	31701	24.52
GO	34008	21.89	30691	25.39	31770	24.57
KOG	19656	12.65	18619	15.40	19319	14.94
all Databases	12360	7.95	11968	9.90	12257	9.48
at least one Database	66589	42.87	56761	46.96	59987	46.4
Total Unigenes	155311	100	120855	100	129258	100

Table 3 Top 20 pathway in AP vs As group and AP vs Pm group.

AP vs Pm group		AP vs As group	
pathway_term gene_number		pathway_term gene_number	
Spliceosome	73	Viral carcinogenesis	111
DNA replication	31	Herpes simplex infection	106
Cell cycle	66	Influenza A	91
Ribosome biogenesis in eukaryotes	43	Amoebiasis	82
Parkinson's disease	66	Transcriptional misregulation in cancer	77
p53 signaling pathway	33	Protein digestion and absorption	71
ECM-receptor interaction	52	ECM-receptor interaction	67
Mismatch repair	18	Measles	63
Fat digestion and absorption	20	Complement and coagulation cascades	59
Circadian rhythm	23	Ribosome biogenesis in eukaryotes	54
Oxidative phosphorylation	59	Systemic lupus erythematosus	50
Circadian rhythm - fly	10	Pertussis	47
Pyrimidine metabolism	45	Phototransduction	43
Antigen processing and presentation	28	Leishmaniasis	39
Graft-versus-host disease	11	Staphylococcus aureus infection	36
Cell cycle - yeast	38	Steroid biosynthesis	18
Base excision repair	20	Allograft rejection	15
Small cell lung cancer	42	Graft-versus-host disease	14
Phototransduction	18	Intestinal immune network for IgA production	14
One carbon pool by folate	15	Asthma	12

Figures

Figure 1

Distribution of each group, As sample, Pm sample and AP sample, sequence annotations database across the GO terms at level 2. Determined across the entire dataset for Biological Process, Cellular Component, and Molecular Function.

Figure 2

KEGG pathway unigene assignments.

Figure 3

Species distribution of successfully annotated sequences across the top 5 species

Cluster analysis of differentially expressed genes

Figure 4

The cluster heatmap in AP, As and Pm sample.

Figure 5

Volcano plot of DEGs in AP vs As and AP vs Pm groups.

Figure 6

GO classification of DEGs in AP vs As and AP vs Pm groups