Skin transcriptome analysis of the genes involved in mucous immunity and discovery of SSRs and SNPs in Crucian carp Carassius auratus

CURRENT STATUS: POSTED

ruixia wang
Henan Normal University

Xiucai Hu
Tianjin Agricultural University

aijun lv ✉ lajand@126.com
Tianjin Agricultural University
Corresponding Author

Jingfeng Sun
Tianjin Agricultural University

Yajiao Song
Henan Normal University

DOI:
10.21203/rs.2.14680/v1

SUBJECT AREAS
Epigenetics & Genomics

KEYWORDS
Carassius auratus; skin; transcriptome; mucin; immunity
Abstract

Fish skin is first line of defence against attachment and penetration of pathogens. Crucian carp (Carassius auratus) is one of the major freshwater species and important food fish in China, yet its molecular mechanism of skin immune response remains unclear. In this study, a de novo transcriptome assembly of crucian carp skin was performed by using Illumina Hiseq 2000 platform. A total of 49,154,776 unigenes were assembled, among which 60,824 (25.16%) unigenes were annotated against the NCBI database. Gene ontology and KEGG mapping assigned the terms to 37,103 (39.15%), 9,337 (23.23%) unigenes, respectively. The best represented KEGG categories were immune system (8,871, 20.50%) and signal transduction (7,805, 18.04%), and several mucin proteins (MUC1, MUC2, MUC5B, MUC5AC, MUC18, etc.) as the main membrane-bound and secreted components of the mucosal layer, implying the differentiation of immune responses in the crucian carp skin. Moreover, the results revealed that potential 28,928 SSRs and 249,964 SNPs were detected in crucian carp. This is the first report on transcriptome analysis in the skin of C. auratus, which contribute to understanding the molecular mechanisms of mucous immunogenectics response and epidermal mucus secretion of the skin in fish.

Background

The crucian carp Carassius auratus is a member of the family Cyprinidae (Teleostei) [1]. It is one of the important fish species in aquaculture, which is distributed widely in both Europe and Asia [2,3]. Crucian carp is also one of the major freshwater species in china, with an annual production of over 3.0 million tons [4,5], and are a highly commercial species due to its excellent growth traits, good taste and suitability in aquaculture[6]. However, exploration of skin immunity is developed with little attention to the crucian carp.
Fish skin is the gateway between animal body and external water environment, which play an important role in the pathogenesis. In addition to fish skin functioning as a mechanical barrier, it is a multi-functional organ with roles in ion regulation, excretion, ion and thermal regulation and metabolically active tissue [7-9]. Although skin is an important component of the mucosal immune system, a few studies on the skin transcription profile have been characterized in fish [10-12]. Evidence is accumulating that skin is a peripheral immune organ in teleosts [2, 13, 14]. However, there is a lack of genomic information such as the skin transcriptome of crucian carp. Transcriptomic tools are commonly used to measure the expression profiling of immune genes and to find new immune genes in fish [15,16]. The annotation was lower than those reported in high throughput sequencing studies conducted in other fish species, such as turbot (44.84%) [17] and mud loach (43.76%) [18], but comparable to that in crucian carp (17.44%) [5]. Fish skin plays various vital functions especially in immunity and defense against invading pathogens and environmental stressors [19, 20]. Pooling of RNA samples from multiple individuals followed by transcriptome analysis using 454 sequencing provides an excellent opportunity to generate large numbers of SNP markers [21]. Analysis of skin transcriptome of Sea trout (Salmo truttam) , Japanese flounder (Paralichthys olivaceus) and Mud loach (Misgurnus anguillicaudatus) showed putative genes involved in immunity and epidermal mucus secretion, suggesting the complexity of immune mechanisms in fish skin [12,22,18]. However, transcriptomic analysis of crucian carp skin remains to be performed.

In this study, we assembled and characterized the skin transcriptome of crucian carp using the Illumina HiSeq 2000 system analysis. A large of genes involved in immune reactions and mucus secretion were identified. Moreover, this study is the first to report the characterization of crucian carp skin associated with innate immunity and molecular
markers, the results will help to develop database for crucian carp and lay a solid scientific foundation for functional genomics in fish.

Results

De novo assembly and transcriptome annotation of the skin of crucian carp

In order to characterize the skin transcriptome of crucian carp, total RNA samples isolated from the skin of crucian carp were subjected to library construction and high-throughput sequencing using the Illumina Illumina Hiseq 2000 platform. Taken together, a total of 4,915,776 unigenes were finally assembled from the filtered short reads. The total number, mean length, N50 value of assembled unigenes were 129,797, 649 and 1,407 bp, respectively (Table 1).

To further understand the crucian carp skin transcriptome, the assembled unigenes were aligned to the NCBI nr database. In this study, our results that all the assembled 60,824 (25.16%) unigenes were annotated by BLASTX searches against the NCBI non-redundant protein database. The similarity distributions of best blast hits are displayed in (Figure 1A). The E-value distribution of the top hits indicated that 15.35% contigs showed excellent matches (E-value=1e-30 ~ 1e-15) and that 19.63% had highly and moderately significant homology (E-value= 0, Fig 1B). The species distributions of best blast hits are shown in Figure 1C. Most of the matched contigs had hits in the Cyprinidae family: Danio rerio, Brachidanio rerio, 79.54% and Carp Cyprinus carpio, 2.25%. Moreover, 18.21% exhibited similarity to other organisms, mainly eukaryotes: fishes, mammals and birds.

Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis

Gene Ontology (GO) category enrichment analysis showed that cellular process, metabolic process, biological regulation and response to stimulus were the four most abundant GO
term in crucian carp skin, as shown in Figure 2. A total of 337, 407 GO terms were assigned to 37,103 unigenes (39.15% of the total assembled sequences). The numbers of mapped GO terms for biological process, molecular function and cellular component were (172,651, 51.17%), (47,757, 14.15%) and (116,999, 34.68%), respectively. Within the biological process category, contigs involved in cellular process (27,381, 15.86%) and single-organism process (22,494, 13.03%) were highly represented. Binding (24,504, 51.31%) and catalytic activity (13906, 8.18%) were dominant groups within the molecular function category. In the cellular component category, cell (26,447, 22.60%) and cell part (26,446, 22.60%) were the most represented subcategories.

The annotated unigenes were further grouped using the Cluster of Orthologous Groups (COG) database. As shown in Fig. 3, 41,200 unigenes were classified into 25 functional families. The subfamily was translation, ribosomal structure and biogenesis (3,011, 7.34%), followed by signal transduction mechanisms (2,631, 6.39%), and cell cycle control, cell division, chromosome partitioning (2,619, 6.36%). A number of unigenes (124, 0.30%) were assigned to defense mechanisms that might be closely related to the crucian carp immune defense.

With the Kyoto Encyclopedia of Genes and Genomes (KEGG) database, we can further analyzed the genes-related signaling pathway, it showed that a total of 43,269 (17.90%) of the unigenes were assigned to 259 pathways, and found the several significant immune-related pathways. Within the top five KO categories including metabolism, genetic information processing, environmental information processing, cellular process and organism system, the subcategory immune system (8,871, 20.50%) was the most highly represented, followed by signal transduction (7,805, 18.04%), nervous system (4,708, 10.88%), cellular community (4,261,9.85%), transport and catabolism (3,776, 8.73%), endocrine system (3,324,7.68%), signaling molecules and interaction (3,234, 7.47%);
Epidermal mucus secretion

Based on the top BLAST hit descriptions, contigs annotated as mucin genes were extracted from skin crucian carp transcriptomes and used for further analysis. The BLASTX results identified contigs as showing significant homology to fish mucin genes in Table 3, 9 types of mucins were identified in the skin. Several mucin proteins (MUC1, MUC2, MUC5B, MUC5AC, MUC18 etc.) as the main structural components of the mucosal layer, implying the differentiation of immune responses in the crucian carp skin. Moreover, multiple alignment of the amino acid sequences of Muc5AC, Muc2 and Muc18 were scanned against a NCBI database, and amino acid sequence comparison revealed the conservation among mucins from Crucian carp, cavefish, rainbow trout, zebrafish, frog, mouse and human (Fig4). Additionally, a search for functional motifs in the PROSITE database revealed three conserved domains in Crucian carp mucins Muc5AC, Muc2 and Muc18 for C8 domain (pfam08742), CT domain (smart00041) and Ig_PVR domain (cd05719), respectively. Phylogenetic tree analysis revealed that the translated amino acid sequences of Crucian carp Muc5AC, Muc2 and Muc5B were closely clustered with the corresponding sequences from zebrafish and other vertebrates, it shows that in vertebrates there are two clusters: membrane-bound and secreted gel-forming mucins cluster, implying an epidermal mucus secretion system in skin immune responses (Fig. 5).

Immune-related pathways

The immune responses in the skin of crucian carp have not been characterized. To obtain a better overview of the immune system in crucian carp, we further identified immune-relevant unigenes from the transcriptome. All the immune pathways annotated in the crucian carp skin transcriptome are listed in Table 4. KEGG annotation identified 259 genes related to the immune system: these genes were distributed in 15 pathways. The
dominant pathways included chemokine signaling, leukocyte transendothelial migration and Fc gamma R-mediated phagocytosis. The highest numbers of retrieved genes in each pathway were found in the RIG-I-like receptor signaling pathway (95.83%), antigen processing and presentation (88.10%) and T cell receptor signaling pathway (78.87%) contained the highest ratios of identified genes versus the total number of known genes in the reference pathway. Thus, these results have provided an overview of the pathways involved in the immune functions of crucian carp skin.

**Identification of Simple sequence repeats (SSRs) and single nucleotide polymorphisms (SNPs)**

In total, 28,928 potential SSRs were identified in 22,829 (19.33%) unigenes. As shown in Table 5, there are 4,801 sequences with more than one SSR, among which 1,445 SSRs are in compound formation. Di-nucleotide repeats are the most frequent form of SSRs (12,099, 41.82 %), and then follow by mononucleotide repeats (11,811, 40.83 %), trinucleotide repeats (4,133, 14.29 %), quadnucleotide repeats (475, 1.64%), pentanucleotide repeats (274, 0.95 %), and hexanucleotide repeats (136, 0.47 %). Moreover, a total of 249,964 putative SNPs were identified, including 143,361 transitions and 106,603 transversions. The proportions of transition substitutions were 28.70% for A/G and 28.65% for C/T, compared with smaller proportions of transversions for A/T (12.46%), G/T (11.30%), C/G (7.53%) and A/C (11.36%) (Table 6).

**Discussion**

The innate immune system is the only known defense weapon of invertebrates and a fundamental defense mechanism of fish [26]. The important defensive role of the skin in the innate immune system is well known and has been studied in several fish species [27-29, 18, 30]. Moreover, crucian carp, mainly omnivorous fish that feed on plants, preferring to live in groups and choosing food, and has long been used for food and medical
purposes in eastern Asia. Therefore, characterization of its transcriptome will be of great value for the breeding, cultivation and disease prevention and control of this species. In recent years, the transcriptome analysis via next-generation sequencing technology has brought new insight into the knowledge of whole transcriptomes in many organisms [31,32,5,33-35]. The immune functions of fish skin have attracted intensive interests of the research community, and a large number of antimicrobial and bioactive substances have been identified in the skin mucus of fish [36-38]. In this present study, we sequenced and analyzed the skin transcriptome of crucian carp using Illumina Hiseq 2000 platform analysis, all differently expressed genes (DEGs) were grouped in GO enrichments and KEGG pathways for functional annotation. With the aim of clarifying the secretary processes and skin immune functions in fish, we obtained that all the assembled 60,824 (25.16%) unigenes were annotated by BLASTX searches against the NCBI non-redundant protein database, GO mapping (37,103, 39.15%) and KEGG annotation 43,269(17.90%). The classification of GO categories including biological process, molecular function and cellular component provide overall information about biological functions of genes expressed in the skin of crucian carp. For molecular functions, the majority of the transcripts were associated with binding, catalytic activity, and enzyme regulator activity. For cellular component, the final products of these transcripts including cell, organelle and membrane were located mainly within cell, which is probably not surprising because the skin tissue was sampled for RNA extraction. For biological processes, most of these transcripts were related to cellular process and metabolic process, which was in agreement with most of other studies [35,39]. The activities of these biological processes perhaps the basis of the quick protein biosynthesis and secreting ability of fish epidermis.

Fish skin has vital biological functions including chemical and physical protection, sensory
activity, behavioral purposes, thermoregulation, hormone metabolism, maintenance of fluid balance and osmotic homeostasis [11,40]. Energy metabolism and cellular differentiation were the two major subclasses of metabolism analysis. Energy metabolism related pathways such as oxoacid metabolic process and carboxylic acid metabolic process were obviously detected after 2 h duration stress. In the previous ischemic/hypoxic brain studies, energy metabolism was considered to play a key role in protecting the organs from the consequences of energy deprivation [41]. MAP kinases (MAPKs) play significant roles in the immune system by regulating key cellular events including cell migration and phagocytosis [42, 43]. And a broad range of functions in energy metabolism were enriched in KEGG pathway, such as arginine and proline metabolism, fructose and mannose metabolism. In a summery, the transcriptome profile analysis would provide valuable information for future studies on development changes and broad stress response in teleosts.

Fish possess numerous distinct and complex defense mechanisms to protect themselves from these pathogenic infections amongst which fish skin mucus acts as the first line of physical defense against pathogens [44]. It has been studied in several fish species [36, 28, 17]. Recently, the skin transcriptome of Atlantic salmon (Salmo salar), Mud Loach (Misgurnus anguillicaudatus) and Sea trout (Salmo trutta m. trutta) were recently assembled and several mucin genes were identified [29,18,12]. Mucins are high molecular weight glycoproteins and are the main constituents of fish skin mucus, which can produce a protective gel to bind a range of bacteria and which constitutes an important part of the mucosal defense against infection [45].

Although immune-related genes in Japanese flounder have been previously characterized from ESTs or EST-based microarray chips [47, 48], studies on the immune system of this species are limited due to the lack of transcriptomic and genomic resources. Analysis of
skin transcriptome of Sea trout (Salmo truttam) and Mud loach (Misgurnus anguillicaudatus) showed putative genes involved in immunity and epidermal mucus secretion, suggesting the complexity of immune mechanisms in fish skin [12,18]. Although immune genes and pathways in fish tissues such as gill, liver, spleen, head kidney and larvae of turbot [17] and head kidney of grass carp [49] have been previously characterized using RNA-seq, our results that T cell receptor signaling pathway (78.87%) and B cell receptor signaling pathway (35.85%). B lymphocytes and T lymphocytes participate in specific antigen defense. B cells are involved in adaptive humoral immunity through antibody production, antigen presentation, and memory B cells development after antigen-mediated activation. B cell activation is achieved through the binding of antigen to the B cell receptor (BCR) located on the outer surface of B cells [50]. T cells are involved in cell-mediated immunity through phagocyte activation, antigen-specific cytotoxic T-lymphocytes, and the release of various cytokines in response to an antigen. T cell activation is achieved when the T cell receptor (TCR) recognizes antigens presented by major histocompatibility complex (MHC) molecules [51]. In this transcriptome study, 38 unigenes were mapped to the B cell signaling pathway that consists of 106 known genes, 56 unigenes were mapped to the T cell signaling pathway that consists of 71 known genes. Consequently, B cell receptor and T cell receptor signaling pathway members detected in the crucian carp transcriptome will enable investigations into the mechanisms in this pathway.

Conclusions

This study investigated the skin transcriptome of the crucian carp using the Illumina Hiseq 2000 platform. A total of 49,154,776 unigenes were finally assembled, 60,824 (25.16%) of which were annotated by BLAST searches. A large number of contigs were classified according to GO and KEGG terms, which represent multiple signaling pathways and
processes. These data provide a rich source to discover and identify immune-relevant genes, and the new mucin sequences (e.g., MUC5AC, MUC2, MUC18) identified here will facilitate our understanding of the mechanisms involved in the immune response. Furthermore, the potential SSRs and SNPs found in this study are important resources with respect to future development of a linkage map or marker assisted breeding programs for the crucian carp.

Methods

Fish

Crucian carp (average weight 50±5g, body length 13±2cm) were purchased from a local fish farm (Xinxiang, China). All procedures involving the handling and treatment of fish used during this study were approved by the Henan normal university institutional animal care and use committee (HNU-IACUC) prior to initiation. Fish were anesthetized by immersion in a bath containing 10 mg/L benzocaine. Fish were acclimatized at 20±2 °C in 45 L tanks with aerated fresh water for 15 days. Fish were fed with commercial dry feed distributed manually twice a day. After 2 weeks of acclimation, fish were randomly sampled to detect pathogenic bacteria from the liver, spleen and kidney. Thirty healthy fish were randomly used for the experiments and the samples from dorsal skin around the central part of the body were collected with sterile scissors. Skin samples from three fish were pooled and immediately frozen in liquid nitrogen and then stored at -80 °C for total RNA extraction.

RNA extraction, RNA-Seq sample preparation

Total RNA was extracted from the skin samples using Trizol reagent (Invitrogen) and was further purified using an RNeasy Mini kit (Qiagen) according to the manufacturer's instructions. The quantity, purity and integrity of RNA were measured on a Nanodrop-2000 spectrophotometer. Samples with higher quality (absorbance ratios at 260 nm/280
nm>1.9) were selected for high-throughput sequencing. Crucian carp skin HiSeq sequencing library was constructed following the manufacturer’s protocol. By using an Illumina Hiseq 2000 platform with instruments in pairs, and making the usual 90 bp readings.

**De novo assembly, comparative analysis, and functional annotation of the transcriptome**

Transcriptome sequencing was conducted by using an Illumina Hiseq 2000 sequencing platform. To better assemble the entire transcriptome de novo, a paired-end (PE) sequencing strategy was used. Transcriptome de novo assembly is carried out with short reads assembling program Trinity package. The raw data generated from Illumina sequencing were first quality-filtered to eliminate adaptor sequences, low-quality bases (Q < 20) and unpaired reads [23]. The longest sequences in each cluster were reserved and designated as unigenes. The assembled unigenes of the crucian carp was compared to the unique nucleotide sequences of crucian carp deposited in the NCBI databases using the BLASTn algorithm. The Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and COG annotations were performed using blastx search against the Nr, Swiss-Prot KEGG and COG databases. Sequences were checked and confirmed for homologs in the GenBank nr database [46] using the program BLASTX and BLASTN (http://blast.ncbi.nlm.nih.gov/Blast.cgi). The outputs of blast searching against the NCBI nr protein database were imported into Blast2GO program [24] for GO term mapping. Multiple sequence alignments were carried out by using the Genedoc program, and phylogenetic analysis was performed and analyzed by using the MEGA 5.1 software.

**Identification of SSR and SNP**

MicroSAtellite (http://pgrc.ipk-gatersleben.de/misa/) was used to identify putative SSRs in the unigenes from the assembled transcirpt. The parameters set to ≥10 repeat units for
mononucleotide SSRs, ≥6 repeat units for dinucleotide, and ≥5 repeat units for trinucleotide, tetranucleotide pentanucleotide, and hexanucleotide SSRs [5]. The compound repeats which composed of two or more microsatellite sequences separated by 100 bases were identified, respectively. To identify putative SNP in the transcriptome of Japanese flounder, quality SNP was used to identify candidate SNP markers [25].

References

1. Rhee J S, Jeong C B, Kim D H, et al. Immune gene discovery in the crucian carp Carassius auratus[J]. Fish & shellfish immunology, 2014, 36(1): 240-251.
2. Zhang Y, Liu J, Peng L, Ren L, Zhang H, Zou L, Liu W, Xiao Y. Comparative transcriptome analysis of molecular mechanism underlying gray to red body color formation in red crucian carp (Carassius auratus, red var.). Fish Physiol Biochem 2017; 43:1387-98.
3. Rhee JS, Jeong CB, Kim DH, Kim IlC, Lee YS, Lee C. Immune gene discovery in the crucian carp Carassius auratus. Fish & Shellfish Immunology 2014; 36: 240-51.
4. Fan J, Shan Q, Wang J, Liu S, Li L, Zheng G. Comparative pharmacokinetics of enrofloxacin in healthy and Aeromonas hydrophila-infected crucian carp (Carassius auratus gibelio). J Vet Pharmacol Ther 2017; 40: 580-2.
5. Liao X, Cheng L, Xu P, Lu GQ, Wachholtz M, Sun XW, Chen SL. Transcriptome Analysis of Crucian Carp (Carassius auratus), an Important Aquaculture and Hypoxia-Tolerant Species. Plos One 2013; 8: 62308.
6. Xiao J, Zou TM, Chen YB, Chen L, Liu SJ, Tao M, Zhang C, Zhao RR, Zhou Y, Long Y, You CP, Yan JP & Liu Y. Coexistence of diploid, triploid and tetraploid crucian carp (Carassius auratus) in natural waters. BMC Genet 2011; 12: 20.
7. Elliott DG. Integumentary system. In: Ostrander GK, editor. The Laboratory Fish. London: Academic Press; 2000. pp. 95-108.
8. Marshall WS, Bellamy D. The 50 year evolution of in vitro systems to reveal salt
transport functions of teleost fish gills. Comp Biochem Physiol. 2010; 155(3): 275-280.
9. Esteban MA. An overview of the immunological defenses in fish skin. ISRN Immunology Volume 2012. Article ID 853470, 29 pages.
10. Xu Z, Parra D, Gómez D, Salinas I, Zhang YA, Jørgensen LVG, Heinecke R, Buchmann K, LaPatra S & Sunyer J. Teleost skin, an ancient mucosal surface that elicits gut-like immune responses. Proceedings of the National Academy of Sciences of the United States of America 2013; 110: 13097-102.
11. Rakers S, Gebert M, Uppalapati S, Meyer W, Maderson P, Sell AF, Kruse C. & Paus R. 'Fish matters': the relevance of fish skin biology to investigative dermatology. Exp Dermatol 2010; 19: 313-24.
12. Malachowicz M, Wenne R, Burzynski A. De novo assembly of the sea trout (Salmo trutta m. trutta) skin transcriptome to identify putative genes involved in the immune response and epidermal mucus secretion[J]. PloS one, 2017, 12(2): e0172282
13. Zhu WB, Wang LM, Dong ZJ, Chen XT, Song FB, Liu N, Yang H. & Fu JJ. Comparative transcriptome analysis identifies candidate genes related to skin color differentiation in red Tilapia. Sci Rep 2016; 6: 31347.
14. Hu Y, Li A, Xu Y, Jiang B, Lu G, Luo X. Transcriptomic variation of locally-infected skin of Epinephelus coioides reveals the mucosal immune mechanism against Cryptocaryon irritans. Fish & Shellfish Immunol 2017; 66: 398-410.
15. Li GX, Zhao YL, Wang J, Liu BZ, Sun XL, Guo S, Feng JX. Transcriptome profiling of developing spleen tissue and discovery of immune-related genes in grass carp (Ctenopharyngodon idella). Fish & Shellfish Immunology 2017; 60: 400-10.
16. Zhou W, Zhang Y, Wen Y, Ji W, Zhou Y, Ji YC, Liu XL, Wang WM, Asim M, Liang XF, Ai TS, Lin L. Analysis of the transcriptomic profilings of Mandarin fish (Siniperca chuatsi) infected with Flavobacterium columnare with an emphasis on immune responses. Fish &
17. Pereiro P, Balseiro P, Romero A, et al. High-throughput sequence analysis of turbot (Scophthalmus maximus) transcriptome using 454-pyrosequencing for the discovery of antiviral immune genes[J]. PloS one, 2012, 7(5): e35369.

18. Long Y, Li Q, Zhou B, Song G, Li T, Cui Z. De novo assembly of mud loach (Misgurnus anguillicaudatus) skin transcriptome to identify putative genes involved in immunity and epidermal mucus secretion. PLoS One 2013; 8: 0056998.

19. Raj V S, Fournier G, Rakus K, et al. Skin mucus of Cyprinus carpio inhibits cyprinid herpesvirus 3 binding to epidermal cells[J]. Veterinary research, 2011, 42(1): 92.

20. Buchmann K. Immune mechanisms in fish skin against monogeneans-a model[J]. Folia Parasitologica, 1999, 46(1): 1-8.

21. Wheat C W. Rapidly developing functional genomics in ecological model systems via 454 transcriptome sequencing[J]. Genetica, 2010, 138(4): 433-451.

22. Huang L, Li G, Mo Z, Xiao P, Li J, Huang J. De Novo assembly of the Japanese flounder (Paralichthys olivaceus) spleen transcriptome to identify putative genes involved in immunity. PloS one. 2015;10(2):e0117642. doi: 10.1371/journal.pone.0117642. PubMed PMID: 25723398; PubMed Central PMCID: PMC4344349.

23. Grabherr M G, Haas B J, Yassour M, et al. Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data[J]. Nature biotechnology, 2011, 29(7): 644-652.

24. Conesa A, Götz S, García-Gómez J M, et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research[J]. Bioinformatics, 2005, 21(18): 3674-3676.

25. Tang J, Vosman B, Voorrips R E, et al. QualitySNP: a pipeline for detecting single nucleotide polymorphisms and insertions/deletions in EST data from diploid and polyploid
26. Magnado.ttir B. Innate immunity of fish (overview). Fish Shellfish Immunol. 2006; 20(2): 137-51. doi: 10.1016/j.fsi.2004.09.006 PMID: 15950491

27. Aranishi F, Nakane M. Epidermal proteinases of the Japanese eel. Fish Physiol Biochem. 1997; 16(6): 471-478.

28. Hjelmeland K, Christie M, Raa J. Skin mucus protease from rainbow trout, Salmo gairdneri Richardson, and its biological significance. J Fish Biol. 1983; 23(1): 13-22.

29. Micallef G, Bickerdike R, Reiff C, Fernandes JM, Bowman AS, Martin SA. Exploring the transcriptome of Atlantic salmon (Salmo salar) skin, a major defense organ. Mar Biotechnol. 2012; 14(5): 559-69. doi: 10.1007/s10126-012-9447-2 PMID: 22527268

30. Du Y, Yi M, Xiao P, Meng L, Li X, Sun G, et al. The impact of Aeromonas salmonicida infection on innate immune parameters of Atlantic salmon (Salmo salar L). Fish Shellfish Immunol. 2015; 44(1): 307-15. doi: 10.1016/j.fsi.2015.02.029 PMID: 25725402

31. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics[J]. Nature reviews genetics, 2009, 10(1): 57.

32. Ji P, Liu G, Xu J, et al. Characterization of common carp transcriptome: sequencing, de novo assembly, annotation and comparative genomics[J]. PloS one, 2012, 7(4): e35152.

33. Salem M, Rexroad C E, Wang J, et al. Characterization of the rainbow trout transcriptome using Sanger and 454-pyrosequencing approaches[J]. BMC genomics, 2010, 11(1): 564.

34. Samanta M P, Tongprasit W, Istrail S, et al. The transcriptome of the sea urchin embryo[J]. Science, 2006, 314(5801): 960-962.

35. Shen G M, Dou W, Niu J Z, et al. Transcriptome analysis of the oriental fruit fly (Bactrocera dorsalis)[J]. PloS one, 2011, 6(12): e29127.

36. Subramanian S, Ross N W, MacKinnon S L. Myxinidin, a novel antimicrobial peptide
from the epidermal mucus of hagfish, Myxine glutinosa L[J]. Marine biotechnology, 2009, 11(6): 748.

37. Zhang C X, Huang K X. Apoptosis induction on HL-60 cells of a novel polysaccharide from the mucus of the loach, Misgurnus anguillicaudatus[J]. Journal of ethnopharmacology, 2005, 99(3): 385-390.

38. Kasai K, Ishikawa T, Komata T, et al. Novel l-amino acid oxidase with antibacterial activity against methicillin-resistant Staphylococcus aureus isolated from epidermal mucus of the flounder Platichthys stellatus[J]. The FEBS journal, 2010, 277(2): 453-465.

39. Bai X, Mamidala P, Rajarapu S P, et al. Transcriptomics of the bed bug (Cimex lectularius)[J]. PloS one, 2011, 6(1): e16336.

40. Le Guellec D, Morvan-Dubois G, Sire J Y. Skin development in bony fish with particular emphasis on collagen deposition in the dermis of the zebrafish (Danio rerio)[J]. International Journal of Developmental Biology, 2003, 48(2-3): 217-231.

41. Schurr A. Energy metabolism, stress hormones and neural recovery from cerebral ischemia/hypoxia[J]. Neurochemistry international, 2002, 41(1): 1-8.

42. Crean J K G, Finlay D, Murphy M, et al. The role of p42/44 MAPK and protein kinase B in connective tissue growth factor induced extracellular matrix protein production, cell migration, and actin cytoskeletal rearrangement in human mesangial cells[J]. Journal of Biological Chemistry, 2002, 277(46): 44187-44194.

43. Huang G, Shi L Z, Chi H. Regulation of JNK and p38 MAPK in the immune system: signal integration, propagation and termination[J]. Cytokine, 2009, 48(3): 161-169.

44. Wang S, Wang Y, Ma J, et al. Phosvitin plays a critical role in the immunity of zebrafish embryos via acting as a pattern recognition receptor and an antimicrobial effector[J]. Journal of Biological Chemistry, 2011: jbc. M111. 247635.

45. Linden SK, Sutton P, Karlsson NG, Korolik V, McGuckin MA. Mucins in the mucosal
barrier to infection. Mucosal Immunol. 2008; 1(3): 183-97. doi: 10.1038/mi.2008.5 PMID: 19079178

46. Ren L, Cui J, Wang J, et al. Analyzing homoeolog expression provides insights into the rediploidization event in gynogenetic hybrids of Carassius auratus red var. Cyprinus carpio[J]. Sci Rep, 2017, 7(1).

47. Aoki T, Hirono I, Kondo H, et al. Microarray technology is an effective tool for identifying genes related to the aquacultural improvement of Japanese flounder, Paralichthys olivaceus[J]. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics, 2011, 6(1): 39-43.

48. Kurobe T, Yasuike M, Kimura T, et al. Expression profiling of immune-related genes from Japanese flounder Paralichthys olivaceus kidney cells using cDNA microarrays[J]. Developmental & Comparative Immunology, 2005, 29(6): 515-523.

49. Chen J, Li C, Huang R, et al. Transcriptome analysis of head kidney in grass carp and discovery of immune-related genes[J]. BMC veterinary research, 2012, 8(1): 108.

50. Kurosaki T, Shinohara H, Baba Y. B cell signaling and fate decision[J]. Annual review of immunology, 2009, 28: 21-55.

51. Smith-Garvin J E, Koretzky G A, Jordan M S. T cell activation[J]. Annual review of immunology, 2009, 27: 591-619.

Tables
Due to technical limitations, the tables have been placed in the Supplementary Files section.

Figures
Figure 1

Characteristics of the homology search of crucian carp transcripts. (A) similarity distribution, (B) E-value and (C) species distribution of the best hits based on the NCBI nr database.
Gene ontology (GO) classification for the assembled unigenes in crucian carp. Results of blastx searches against the NCBI nr protein database were imported into Blast2GO software for GO term mapping and annotation. The number and ratio of sequences assigned to level 2 GO terms from sub GO categories including biological process, molecular function and cellular component were shown.
Figure 3

The cluster of orthologous groups (COG) classification of the unigenes from crucian carp.
Multiple alignment of the amino acid (aa) sequences of Muc5AC, Muc2 and Muc18 from Crucian carp, cavefish, rainbow trout, zebrafish, frog, mouse and human. (A) Muc5AC cavefish (Sinocyclocheilus grahami, XP_016148695.11), rainbow trout (Oncorhynchus mykiss, XP_021463199.1), Western clawed frog (Xenopus tropicalis, XP_017948652.1), human (Homo sapiens, CAC83674.1), Muc5B genes of zebrafish (Danio rerio, XP_021333301.1), mouse (Mus musculus, XP_006508716.2). Conserved domains C8 domain (ID 10555326 aa 192-238) is
marked in the crucian carp sequence. (B) Muc2, cavefish (Sinocyclocheilus anshuiensis, XP_016346170.1), rainbow trout (Oncorhynchus mykiss, XP_021442066.1), human (Homo sapiens, AAA59164.1), mouse (Mus musculus, Mus musculus) and Western clawed frog (Xenopus tropicalis, XP_017948654.1). Conserved domains CT domain (ID 214482

GHB_like superfamily) aa 68-122 is marked in the crucian carp sequence. (C) Muc18, human (Homo sapiens, NP_006491.2), mouse (Mus musculus, NP_075548.2), Western clawed frog (Xenopus tropicalis, XP_004916103.1), cavefish (Sinocyclocheilus rhinoceros, XP_016148695.11), zebrafish (Danio rerio, XP_021337028.1), Salmo salar (Atlantic salmon, XP_014069113.1). Conserved domains Ig_PVR domain (ID 319296 aa 146-191) is marked in the crucian carp sequence.
Figure 5

Phylogenetic tree analysis of 26 mucins from different organisms based on amino acid sequences. Crucian carp mucins are highlighted in a black triangle. GenBank accession numbers are provided for each sequence. Unrooted tree was generated using neighbor-joining method by the MEGA 5.1 software. The numbers next to the branches indicate percentage values for 1000 bootstrap replicates.

Supplementary Files

This is a list of supplementary files associated with the primary manuscript. Click to download.
Tables.docx
