

**Comparative transcriptome analysis of genes expressed in the skin
of channel catfish, common carp and pleco**

by

Jun Yao

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
December 12, 2015

Keywords: channel catfish, common carp, pleco, scale, transcriptome, comparative analysis

Copyright 2015 by Jun Yao

Approved by

Zhanjiang Liu, Chair, Professor of School of Fisheries, Aquaculture and Aquatic Sciences
Rex Dunham, Professor of School of Fisheries, Aquaculture and Aquatic Sciences
Nannan Liu, Professor of Entomology and Plant Pathology
Luxin Wang, Assistant Professor of Food Safety and Microbiology

Abstract

Channel catfish (*Ictalurus punctatus*), as a member of Siluriformes, has scaleless skin. However, it is hardy and resistant to many types of stresses including low oxygen, high density, handling, and infections from various pathogens. These characteristics make it a successful species for aquaculture. It is widely cultured in the world, especially in North America and Asia. As such, its genome research has been a priority of genetics research. Catfish is one of the aquaculture species on which the USDA NRSP-8 National Animal Genome Project Program has focused.

However, the evolutionary causes for the scaleless skin, and how the scaleless skin adapted to being resistant to disease infection is unknown. The catfish transcriptome has been well studied from various tissues, but the skin transcriptome has not been well characterized. The goals of this research are to conduct transcriptome analysis in the skin of scaleless catfish to generate the catfish skin transcriptome, and conduct comparative transcriptome analysis of scaled fish with scaleless fish to provide insights into the genomic causes for the evolutionary loss of scales, and for the compensatory expression of genes to account for the strong resistance against various infections without the physical protection of scales.

RNA-Seq was used to generate the short reads of the skin transcriptomes of channel catfish, common carp, and pleco. In all cases, hundreds of millions of short reads were generated,

and they were then assembled into contigs for bioinformatic analysis. Annotation of the transcriptomes allowed identification of 20,474 genes from channel catfish skin, 25,136 from the common carp skin, and 21,105 from the pleco skin.

The interspecific comparative transcriptome analysis between the scaled common carp and scaleless channel catfish allowed identification of 836 genes that were expressed in the skin of common carp but not in the skin of channel catfish. This pool of the 836 genes should be of interest for the identification of the genes important for scale formation.

The vast majority of catfishes are scaleless, but some such as pleco has bony dermal plate type of scales. The scaled catfish should provide an ideal natural model for comparative studies as to what genes have been lost during evolution in scaleless catfishes. However, scale regeneration was not possible with pleco. I conducted scale regeneration experiments in common carp to identify differentially expressed genes during scale regeneration. Compared to the normal common carp skin, a total of 1,173 differentially expressed genes were identified during carp scale regeneration. Functional annotation indicated that the differentially expressed genes included genes regulating the immune response, metabolism, collagen trimer and binding, cellular cytoskeletal structure, calcium ion binding, and formation or development related genes in scale, bone, hair, tooth and fin, providing a pool of genes involved in scale formation.

Although the interspecific transcriptome subtraction and the scale regeneration experiments both allowed the identification of a fraction of the skin transcriptome to be considered for candidates of scale formation, the list of these pools of genes are yet too long (836 and 1,173, respectively). I further reasoned that the genes important for scale formation should

satisfy both the conditions: 1) they should not be expressed in the scaleless catfish, and 2) they should be differentially expressed during scale regeneration. Therefore, a cross subtraction was conducted: coupling interspecific comparative transcriptome subtraction between common carp skin transcriptome and channel catfish skin transcriptome with the carp scale regeneration experiments. A total of 18 genes meeting the above two conditions were identified including 13 known genes and five unknown genes. Of the 13 known genes, 10 genes were up-regulated during scale regeneration, and three genes were down-regulated. The up-regulated genes include 14 kDa apolipoprotein, actinoporin-like protein, apolipoprotein A-I, apolipoprotein AIIb1, lymphocyte antigen 6D, protein FAM133-like, protein GAPT-like, secretory calcium-binding phosphoprotein 7, si:dkey-22i16.3 (fa93e10), and uroplakin 2-like. Although their involvement in scale formation requires functional analysis, this pool of 18 genes is of great interest for additional analysis for their involvement in scale formation.

Although direct analysis of scale regeneration was not feasible with pleco, analysis of its skin transcriptome should still be useful to provide insights for the involved genes for scale formation. To obtain the differentially expressed genes between the channel catfish and pleco, RNA-Seq was conducted with pleco skin. Comparative transcriptome subtraction between the transcriptomes of pleco and channel catfish skin allowed identification of 704 genes that were expressed in the skin of scaled pleco but not in scaleless channel catfish.

Of the 18 genes that were only expressed in scaled carp skin but not in channel catfish skin and are differentially expressed genes during scale regeneration, three were expressed in scaled pleco skin, but not in channel catfish. These genes are lymphocyte antigen 6D, secretory

calcium-binding phosphoprotein, and uncharacterized protein si:dkey-30j10.5. The sharing of these genes from both comparative subtraction analysis provided stronger evidence that these genes are likely involved in scale formation in teleosts. They may account, at least in part, for the evolutionary lack of scales in catfish. Future functional studies such as knock out experiments using CRISPR or its related technologies are warranted.

My dissertation work allowed generation and assembly of the transcriptomes from channel catfish, common carp, and pleco. Comparative transcriptome subtraction analysis, coupled to scale regeneration experiments, as demonstrated here, provided insights into genomic basis for the scaleless phenotype. My work also demonstrated the power of comparative transcriptome subtraction as an alternative for searching candidate genes of structural significance. Practically, the transcriptome resources I generated should be useful for various genome and transcriptome research, which would set the foundation for evolutionary analysis, as well as analysis for various biological and aquaculture-related traits. In addition, as the transcriptome resources accumulate from the scaled fishes and scaleless fishes, it will soon become possible to conduct comparative analysis as to what additional genes that are expressed in scaleless fishes, but not in scaled fishes, which provide the compensatory advantage allowing the scaleless fishes to be resistant against various infectious agents even though they lack the physical protection offered by scales.

Acknowledgments

I am deeply indebted to my graduate committee, friends, colleagues, and family members because I would never have been able to finish my dissertation without their guidance, help, encouragement and support. My deepest gratitude goes to my advisor, Dr. Zhanjiang Liu, for providing me the opportunity to work as a Ph.D. student in his laboratory. Throughout my four and half years of graduate work, Dr. Liu has always provided me with strong guidance for my research work, and kind encouragement and inspirations for my success. I would like to thank Dr. Rex Dunham for his advice of my research and helping me to develop my background in genetics. Many sincere thanks go to Dr. Nannan Liu and Dr. Luxin Wang, who are always willing to help and give their advice to me not only for my research but also for my life. My appreciation also extends to Dr. Charles Y. Chen for his advice and support.

I would like to thank Dr. Huseyin Kucuktas for his technical assistance. Many thanks to all the colleagues in the laboratory for their collaborations, help and friendship. Special thanks goes to Ludmilla Kaltenboeck for her support and friendship that made my time at Auburn enjoyable.

Finally, I would like to thank my husband, Jiaren Zhang, for his endless love, and unconditional support at all times, good or bad. I am grateful to my mother Zhenhua Liu, my parents in law Li Zhang and Jianli Hao, for their understanding, support, and encouragement, and their care of my lovely daughter, the little Rebecca.

Table of Contents

Abstract.....	ii
Acknowledgments.....	vi
List of Tables	ix
List of Figures	x
CHAPTER I. INTRODUCTION.....	1
CHAPTER II. LITERATURE REVIEW	5
Channel catfish and common carp	5
Fish scales	11
RNA sequencing (RNA-Seq).....	13
Scale regeneration studies in fish.....	20
CHAPTER III. MATERIAL AND METHODS.....	24
Overall strategy	24
Sample collection, RNA extraction and sequencing.....	25
Assembly and annotation of transcriptomes of channel catfish skin, common carp skin, and pleco skin.....	28
Interspecific comparative transcriptome analysis	30
Genes differentially expressed for carp scale regeneration as analyzed by RNA-Seq Analysis	32

CHAPTER IV. RESULTS.....	34
Generation of short reads from channel catfish skin, common carp skin and pleco skin	34
<i>De novo</i> assembly of transcriptomes of the channel catfish skin, common carp skin, and pleco skin	36
Gene identification and annotation of channel catfish skin transcriptome, common carp skin transcriptome and pleco skin transcriptome.....	38
Differentially expressed genes during carp scale regeneration.....	41
Comparative subtraction of the channel catfish skin transcriptome from the common carp skin transcriptome.....	55
The cross intersection of differentially expressed genes during carp scale regeneration and genes only expressed in common carp skin, but not in channel catfish skin	86
Comparative subtraction of the channel catfish skin transcriptome from the pleco skin transcriptome.....	90
CHAPTER V. DISCUSSIONS.....	117
CHAPTER VI. CONCLUSIONS	124
REFERENCE.....	126

List of Tables

Table 1	17
Table 2	34
Table 3	35
Table 4	36
Table 5	37
Table 6	37
Table 7	38
Table 8	39
Table 9	40
Table 10	40
Table 11	41
Table 12	44
Table 13	45
Table 14	57
Table 15	88
Table 16	92

List of Figures

Figure 1	6
Figure 2	7
Figure 3	9
Figure 4A	18
Figure 4B	18
Figure 5	43
Figure 6	56
Figure 7	89
Figure 8	91

CHAPTER I. INTRODUCTION

Catfish, belonging to the order of Siluriformes, contain over 4,000 species. It is one of the largest orders of teleosts, accounting for over 12% of all fish species (around 33,000 species) and over 6.2% of all vertebrates (about 64,000 species) (Eschmeyer and Fong, 2015; Wilson and Reeder, 2005). The vast majority of catfish are scaleless, but some have bony dermal plates covering the skin, such as Loricariids (plecos). However, the genetic basis for the evolutionary lack of scales is unknown. The evolutionary lack of scales are not only important for the understanding of osteogenesis and formation of appendages, but also for understanding of immune compensation of catfishes in dealing with various infectious agents without physical barrier of scales as a mechanism of protection. Therefore, the analysis of the genomic basis for the evolutionary lack of scales is of great interest not only to evolutionary biologists, but also to aquaculturists.

In spite of its biological significance, scale formation is not well understood. This is, in part, because of the low research support of fish biological studies. However, in recent years, technological advances, made such research more feasible. For instance, the direct phenotypic contrast is available within catfish. With many catfish being scaleless, such as channel catfish (*Ictalurus punctatus*) and blue catfish (*I. furcatus*), direct genomic and transcriptomic

comparisons are now possible by generating genomic or transcriptomic sequences between fish with scales and those without scales. However, although such studies could have been conducted by comparing channel catfish genome with those of scaled catfishes, biological analysis of scale formation is difficult with catfish bearing bony dermal plates. Thus, my dissertation project focused on the analysis of scale formation using channel catfish and common carp as a model system because genomic sequences of both species are available, and scale regeneration experiments are possible with common carp.

To provide insights into the genomic causes for the scalelessness in channel catfish, and further understand how the loss of scales is related to other disadvantage and/or advantages of catfish in their ability to adapt to the environment, I sequenced and annotated the transcriptome of skin in scaleless channel catfish and the scaled common carp (*Cyprinus carpio*). A direct comparison of the two transcriptome allows me to obtain a list of genes that are expressed in the scaled carp skin, but not in the scaleless channel catfish skin. However, because channel catfish and carp are phylogenetically different, plus many physiological and environmental effects, this list of genes may be long, which prohibits from practical analysis for genes responsible for the lack of scales.

Scale regeneration is readily possible with common carp. Therefore, I also conducted scale regeneration experiments with common carp, and determined differentially expressed genes during scale regeneration. Along with the list of genes that are only expressed in scaled carp but not in scaleless channel catfish, the differentially expressed genes during scale regeneration in

common carp allow me to pin down the candidate genes. The cross intersection of these two lists, at least, should significantly reduce the number of genes because the genes must satisfy two conditions: 1) they are only expressed in the skin of fish with scales such as carp, and 2) they also must be expressed differentially during scale regeneration. In addition, we also have relatively abundant genomic resources in channel catfish and carp such as the whole genome sequences. In addition to the genes, analysis of genes pathways of the identified genes should allow me further to pin down the candidate genes important for scale formation. My research, as presented here, provides insights into the genomic basis for the evolutionary scaleless phenotype, a characteristic shared by thousands of catfish species. In addition, my project may provide additional hints as to how scaleless catfish compensate their loss of scales in response to infectious agents. Literature reviews of fish species, fish scales, RNA sequencing and fish scale regeneration studies are given in chapter II.

The long-term goal of this project is to understand the biological characteristics of catfishes, especially the genes responsible for the loss of scales, and how the loss of scales is related to other disadvantages and/or advantages of catfish in their ability to adapt to the environment. This project is a crucial first step toward such understanding by providing some initial candidate genes involved in scale loss in catfishes. Future studies and functional analysis will validate my results, and eventually leading to the demonstration of the causative genes for scale formation. Additional understanding of immunological changes associated with the loss of scales should also have implications in aquaculture.

For my dissertation project, I had the following specific objectives 1) Generate and assemble skin transcriptome from channel catfish; 2) Generate and assemble skin transcriptome from common carp (scaled); 3) Generate and assemble skin transcriptome from pleco (a member of Siluriformes, but with dermal bony plate); 4) Comparative subtraction of the channel catfish skin transcriptome from the common carp skin transcriptome, to determine a list of genes that are expressed in the scaled carp, but not in scaleless channel catfish; 5) Comparative subtraction of the channel catfish skin transcriptome from the pleco skin transcriptome, to determine a list of genes that are expressed in the scaled pleco, but not in scaleless channel catfish; 6) Conduct scale regeneration experiments with common carp to determine differentially expressed genes during scale regeneration; and 7) Cross subtraction of the interspecific transcriptome difference and the differentially expressed genes during scale regeneration to narrow the pool of genes as candidate genes involved in scale formation.

CHAPTER II. LITERATURE REVIEW

Channel catfish and common carp

Generally, fish are a group of finned aquatic vertebrate animals with the ability of breathing through gills. There are around 33,000 fish species (<http://www.fishbase.org>) in the world, accounting for about 50% of the total number of vertebrates. Fish spend all their lives in aquatic environments such as marine, fresh water, and brackish waters throughout the world. As shown in figure 1, fish species can be divided into three major groups based on the order of precedence in evolution. The first group is Aganatha, consists of primitive jawless fish, such as hagfish and lampreys. The second group is named Chondrichthyes, refers to a group of cartilaginous fish belonging to superclass Gnathostomata, with the features of possessing jaws, paired appendages and a cartilaginous skeleton, such as sharks and rays. The third group is Osteichthyes which contains the largest number of bony fish, characterized by the presence of lungs or swim bladder and bony endoskeleton. Osteichthyes can be further divided into two subclasses: Sarcopterygii (lobe-finned fish), including coelacanths and lungfish, and the largest and most diverse vertebrate group, Actinopterygii (ray-finned fish). Most of present teleost fish are in the infraclass Neopterygii which belonged to Actinopterygii. The basic phylogenetic relationships of these groups are shown in Figure 1.

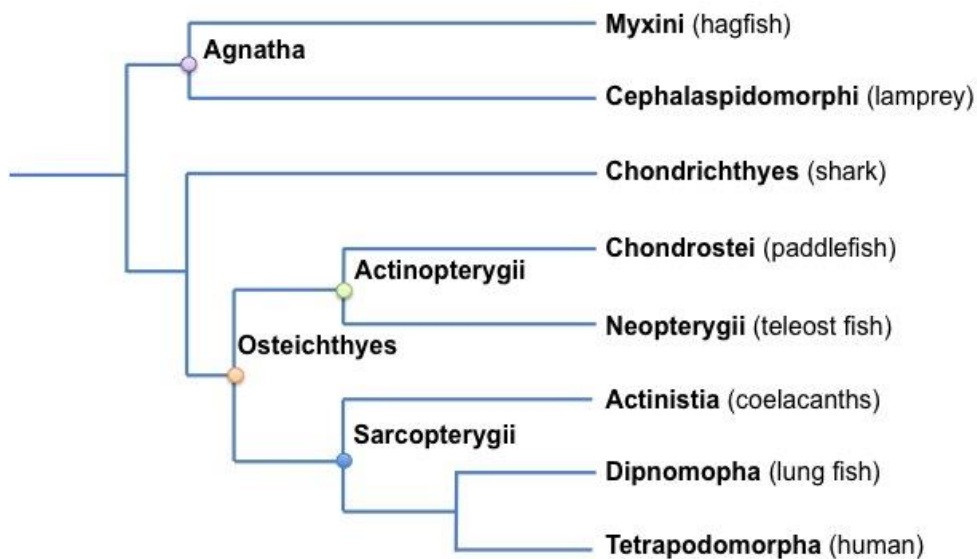


Figure 1. Higher level phylogenies of aquatic vertebrate animals. Pre-teleost groups include hagfish, lampreys, and sharks lacking the mineralized endoskeleton. The teleost group contains the largest numbers of bony fishes, while the more advanced groups such as coelacanth are lies between teleosts and tetrapods.

Teleosts include over 30,000 species. Many of them are extremely ecologically and economically important. For instance, they are important species for fisheries or aquaculture. Currently, fish products accounts for ~20% animal protein for human consumption (http://www.who.int/nutrition/topics/3_foodconsumption/en/index5.html), and approximately 50% of the seafood are produced from aquaculture. Many fish species are recreationally important species for sport fishing. In addition, with its crucial position in evolution, quite a few fish

species, such as zebrafish and medaka, have been used as model species for scientific research to address developmental biology questions.

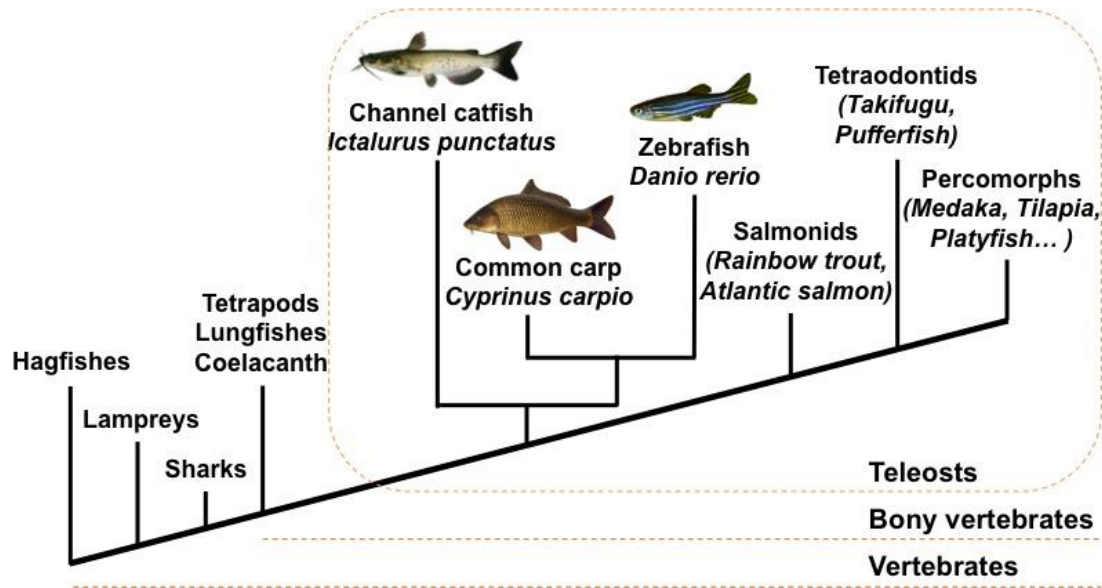


Figure 2. Phylogenetic tree for channel catfish and common carp.

Catfish, belonging to the class of Actinopterygii, the order of Siluriformes, is an extremely diverse clade with over 4,000 extant species (Armbruster, 2011), accounting for over 12% of all fish species (around 33,000 species) and over 6.2% of all vertebrates (about 64,000 species). Catfish are a group of freshwater fishes broadly distributed worldwide, and of considerable economic importance for aquaculture worldwide (Eschmeyer and Fong, 2015; Wilson and Reeder, 2005). They are so-named because their barbels resemble a cat's whisker. Catfish have four pairs of chemosensitive barbels surrounding the mouth, with one pair of barbel at each

corner of mouth, one pair of barbel adjacent to each nostril and two pairs of barbel under the jaw (Holtan, 1998). These barbels are sensors capable of sensing water waves and tastes. Catfish are usually grey and white in color and their colors vary as an adaptive trait to their environment. The body of Catfish is roughly cylindrical shape with sharp spines in their dorsal fin and two pectoral fins. Most catfish lack scales, but some catfish, such as plecos (*Pterygoplichthys disjunctivus*) and striped raphael catfish (*Platydoras armatulus*), have bony dermal plates covering their skin (Arce et al., 2013; Armbruster, 2004; Ferraris Jr and Vari, 2012). Because of being scaleless, catfish are of great interest to evolutionary biologists.

Channel catfish (*Ictalurus punctatus*, family Ictaluridae, Order Siluriformes) is an important commercial aquaculture species widely cultured in a number of countries in America, Europe and Asia (FAO, http://www.fao.org/fishery/culturedspecies/Ictalurus_punctatus/en; Page and Burr, 1991). Particularly, channel catfish is the primary aquaculture species in the United States, accounting for around 1/3 to 1/2 of all U.S. aquaculture production. As a bottom-dwelling species, channel catfish favor to spend most of their lives at the bottom of fresh water volume, such as rivers, lakes, streams, reservoirs as well as ponds. They prefer substrates like sand, gravel and mud, thus are able to tolerate a wide range of environmental conditions. For example, even though the optimal water temperature for growth is around 29 °C, channel catfish can survive at a wide range of temperatures from just above freezing to nearly 38 °C. Channel catfish become sexually mature in three years, and their optimal spawning temperature is around 27 °C.

However, channel catfish can spawn with variable temperatures ranging from 21 °C to 30 °C (Wellborn, 1988).

One of the most distinct characteristics for channel catfish is scaleless. However, channel catfish have a thick skin with irregular and dark spots on it, and a thick layer of mucosa. This mucosal layer functions both as a physical barrier for nutrient absorption and waste secretion, and in mediating immune responses (Li et al., 2013; Li et al., 2012; Peatman et al., 2013). Other appendages for channel catfish include four pairs of barbels around the mouth, a fork tail, a soft-ray dorsal fin containing hard spines, a pair of sharp and hard pectoral fins, a pair of pelvic fins, a long and soft-ray anal fin and an edge-rounded adipose fin.

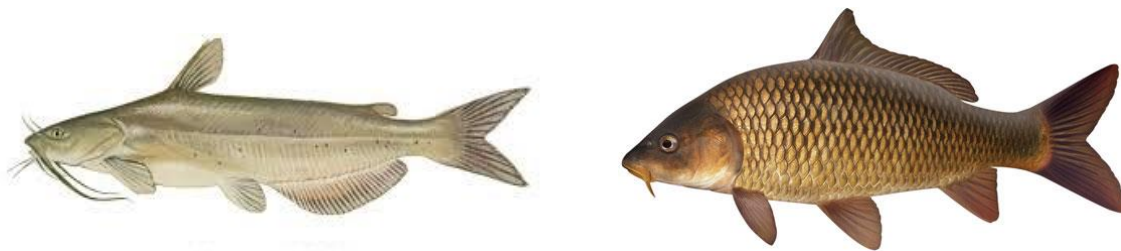


Figure 3. Pictures of *Ictalurus punctatus* (left picture) and *Cyprinus carpio* (right picture).

(Pictures came from website

<http://www.dpi.nsw.gov.au/fisheries/pests-diseases/freshwater-pests/species/carp> and

<http://www.indianaoutside.com/indiana-fish-identification/channel-catfish/>).

Common carp (*Cyprinus carpio*) (Figure 3, right), belonging to the family Cyprinidae, Order Cypriniformes, and Class Actinopterygii, is an outstanding commercially important

freshwater aquaculture species around the world, accounting for 10% of total freshwater aquaculture production (Bostock et al., 2010). First group of domesticated common carp is believed to be the descendant of wild ancestor lived in Caspian and Aral Sea basins. After that, common carp were dispersed to both European and Asian countries (Balon, 1995). Common carp are further introduced into other parts of the world including North America, southern Africa, Australia and New Zealand (FAO, http://www.fao.org/fishery/culturedspecies/Cyprinus_carpio/en). During the mid-18th century, common carp were brought to the United States as a food fish.

Different from channel catfish, common carp wear large and thick cycloid scales as physical barrier, and their dorsal fin and anal fins contain sharp spines. There are commonly 32 to 38 scales lying on the lateral line of the body. In addition, the caudal fin of common carp is deeply emarginated (Kottelat and Freyhof, 2007), and they have two pairs of barbels. Common carp inhabit at bottom of rivers, lakes and water reservoirs, but they can also swim to middle and upper layer of water volume to seek food. Common carp can tolerate a wide range of water conditions. Take oxygen and temperature as examples, Common carp can survive in both hypoxia (low oxygen concentration) and hyperoxia (super saturated oxygen level) conditions. Generally, the optimal water temperature for common carp growth is between 23-30 °C, and best temperature for spawning is above 17-18 °C. However, in tropical areas, common carp are even able to breed throughout the year.

Fish scales

Fish scales, primarily composed of calcium, are commonly derived from dermis layer of skin and function as a physical barrier to protect fish body from extrinsic invasion. Together with skin, they play an important role in fish locomotion through controlling wave propagation and storing energy to make swimming more efficient (Zhu et al., 2012). All fish scales, and even other types of dermal derived skeleton in living vertebrates share a common ancestor, odontodes (Huysseune and Sire, 1998; Reif, 1982; Sire and Huysseune, 2003). However, the morphology of scales is diverse in shape, size, structure, and arrangement, and most fish scales can grow along with fish growing. Therefore, ichthyologists use fish scales as a coordinate to determine life history, and understand the relatedness of different lineages (Lagler, 1962).

Fish scales are generally classified into four major categories: placoid, cosmoid, ganoid, and elasmoid (including cycloid and ctenoid) (Zhu et al., 2012). These types of scale evolved in an evolutionary order: Primitive jawless fish hagfish and lampreys have no scales; cartilaginous sharks and rays have dentinal placoid scales, which are structurally similar with mammalian teeth; coelacanth (*Latimeria chalumnae*) scales are named ‘cosmoid’; in primitive and basal actinopterygii species, such as sturgeons and gars, they have ganoid scales; most actinopterygii species have elasmoid scales, with the exception of some fish species possessing dermal bony scales (Lagler, 1962) and scutes.

Elasmoid scales are thin, transparent, imbricated, lamellar and collagenous, and having been found in a large number of teleosts (Sire and Akimenko, 2004). Each scale is derived and inserted in its own pocket in the superficial dermis, and covered by epidermis layer (Bereiter-Hahn and Zylberberg, 1993; Sire and Huysseune, 2003). Elasmoid scales are arranged to follow the direction from head to tail, making them more flexible, and reducing drag when swimming in the water. There are two major forms for elasmoid scales: cycloid scales and ctenoid scales. Cycloid scales are simply rounded and smooth on the outer or posterior edge, and they are more popular type in many teleosts, such as carp and salmon. While ctenoid scales, with a set of fine teeth along the posterior edge can be found in a small group of teleosts, such as perch. Generally, in order to generate a new elasmoid scale, fibroblast-like cells in the dermis layer will accumulate at the epidermal-dermal boundary and form a dermal papilla, and then calcium is deposited between two layers of scale-forming cells (Lanzing and Wright, 1976; Sire et al., 1997). In some catfish species, such as common pleco, the scales have been modified and merged into a dermal bony plate to form an armour. This type of shield-like scale is called scutes, and these fish are so called armoured catfish. Scutes type scales can be also found in sea horses and pinecone fish. Cartilaginous fish, such as shark and rays, wear placoid scales. The structure of this type of scale is similar with teeth, with hard enamel layer outside, a dentine layer and a soft central pulp, and as such placoid scales are also called dermal denticles. Different from elasmoid scales which are present between dermis layer and epidermis layer, placoid scales pass through the epidermis layer, and can't enlarge with the growing of fish body. Instead, new scales

will fill up the space opened on the skin during fish growth. With these dermal denticles covering on the skin, sharks are protected from predators or injuries, and even reduce drag and improve swimming efficiency (Raschi and Tabit, 1992). The cosmoid scales are rare type of scale, and are composed of four layers: dense bone, spongy bone, hard dentine and enamel. In contrast, the ganoid scales, which are evolutionarily emerged after cosmoid scales, only have a calcified layer and a fibrous layer.

During evolution, environment has significant effect on fish scale morphology. For example, fish inhabit in fast flowing waters, such as trout, have small scales, thus to reduce the influence of flow drag to swim rapidly. In contrast, fish living in motionless waters tend to have larger type of scales, such as carps. Some fish species, such as herring, have different size of scales, with larger scales on head and smaller scales on tail, indicating their different demands for flexibility (<http://www.earthlife.net/fish/scales.html>). Overall, the evolution trend is from large and multilayered scales to smaller, thinner, and less number of layered, all of which would reach the goal of decreasing weight, improving flexibility, maneuverability and increasing swimming efficiency (Bruet et al., 2008; Colbert, 1955).

RNA sequencing (RNA-Seq)

With the constant progress in sequencing technologies, particular the robust development of next generation sequencing, increasing number of whole genome sequences have been released.

For instance, there are 13 fish genomes have been deposited in Ensembl genome database. These fish species include zebrafish (*Danio rerio*), Atlantic cod (*Gadus morhua*), medaka (*Oryzias latipes*), tilapia (*Oreochromis niloticus*), fugu (*Takifugu rubripes*), cave fish (*Astyanax mexicanus*), stickleback (*Gasterosteus aculeatus*), spotted gar (*Lepisosteus oculatus*), coelacanth (*Latimeria chalumnae*), platyfish (*Xiphophorus maculatus*), Amazon molly (*Poecilia formosa*), lamprey (*Petromyzon marinus*) and tetraodon (*Tetraodon nigroviridis*). The availability of these genomic resources establishes concrete groundwork for biological studies at the genome level. However, extra efforts, at transcriptome level, are necessary to be made for the expression analysis of functional genes harbored in the genome. Generally, a transcriptome is composed of the complete set of transcripts, including mRNA, rRNA, tRNA, and other non-coding RNA in a given cell, or tissue, or the entire organism. Different from the genome, transcriptome is dynamic depending on developmental stages, physiology condition, and the environment in a particular moment (Martin and Wang, 2011; Wang et al., 2009). With these characteristics, transcriptome analysis, including the transcriptional structure of genes, and the expression profile change of each gene during development under different physiology or environmental conditions, is of great importance to annotate the functional elements (including functional genes, and other untranslated fragments) of the genome (Wang et al., 2009).

A number of technologies have been developed to analyze transcriptional level activities in past years. These technologies can be classified into two categories: hybridization-based and sequencing-based approaches. For hybridization-based approach, which is mainly applied in

traditional and commercial high-density oligo microarrays, probes are fixed onto a solid surface and then hybridize to DNA or RNA in samples. This hybridization-based approach has advantages of its relatively inexpensive cost and high throughput screening capacity. However, it is limited by the requirement of a reference genome, only with which the information of genes or probes on arrays could be designed. Beyond that, the complicated normalization method is another obstacle for application when comparing expression levels across different experiments (Russo et al., 2003).

Sanger sequencing technology was widely applied in screening of cDNA or EST libraries at the early stage of sequencing-based transcriptional analysis. However, with the rapid development of next generation high-throughput sequencing platforms, sequencing-based approach has entered a new era. In comparison with traditional EST sequencing by Sanger technology, which can provide only 10 fold or even less coverage for expressed genes, RNA sequencing (RNA-Seq) using next generation sequencing technologies are able to generate much greater depth and higher coverage of transcript sequencing. For example, in a regular RNA-Seq experiment using Illumina HiSeq 2000 platform, more than 400 million reads with the length of 100 bp each read can be obtained. Assuming there are 25,000 genes in the species of interest, an average of 16,000 reads could be gained for each gene, which is a much higher coverage than output from sequencing EST libraries using Sanger technology. In addition, with the ability of snap-shot a near-complete transcriptome, RNA-Seq is able to capture rare transcripts, such as regulatory elements.

RNA-Seq, as a deep sequencing technology for gene expression profiling, is popular and broadly applied in a variety of genomic and transcriptome analysis, not only because of its capability of providing high sequencing coverage, but also because of its capability for new gene discoveries. Different from microarrays, which can only detect transcripts hybridizing to probes designed in advance and fixed on the array surface, RNA-Seq is able to capture all the transcripts at the time of analysis, thus providing a blue print of whole transcriptome of interested organisms including both known and unknown genes and regulatory RNAs. In addition, RNA-Seq is capable of identification of SNPs and splicing variants in the transcribed region (Sun et al., 2014; Wang et al., 2013), and these SNPs and splicing variants could be very informative because they may be related to gene regulation and functions. Compared with microarray technology, RNA-Seq uses simpler sample preparation procedure, requires smaller amount of input sample, detects broader range of expression levels, is more sensitive to detect low-abundance transcripts, and requires less financial resources on a per gene basis. RNA-seq technology has shown to be accurate for quantitative analysis of transcription, generating results highly correlated with those provided by quantitative PCR (qPCR) (Li et al., 2013; Liu et al., 2013; Nagalakshmi et al., 2008; Wang et al., 2009). In conclusion, RNA-Seq provides manifest advantages over existing approaches for both mapping and quantifying transcriptome. A comparison of pros and cons between RNA-Seq and traditional microarray-based approaches are summarized in Table 1.

Table 1. Advantages of RNA-Seq compared with microarray.

Technology	RNA-Seq	Microarray
Principle	High-throughput sequencing	Hybridization
RNA amount	Low	High
Based on genomic sequence	No	Yes
Throughout	High	Low
Background noise	Low	High
Range to quantify gene expression level	Large	Small
Distinguish allelic expression	Yes	Limited
Detect isoforms or splicing	Yes	Limited
Cost	Low	High

The regular procedure for RNA-Seq is summarized in Figure 4A. Generally, total RNA with polyadenylation is firstly reversed transcribed to cDNA fragments with adaptor attached to one or both ends. After amplification, high-throughput RNA sequencing then generates hundreds of millions of reads each lane, with read lengths vary from 36 bp to 200 bp (Luo et al., 2012; Metzker, 2010). Currently, there are three dominant next generation sequencing technologies provided by 454 Roche Applied Science (<http://www.454.com>), Illumina (<http://www.illumina.com>), and the Applied Biosystems SOLiD (<http://www.appliedbiosystems.com>), respectively. Among these, Illumina sequencer and ABI SOLiD can generate more reads than Roche 454 sequencer. For example, Roche 454 output about 0.7 Gigabase (Gb) data per 10-hour run, with a cost of \$5,000, while Illumina Hiseq 2000 sequencer can generate 50-60 Gb data per lane, with a cost of \$3,000.

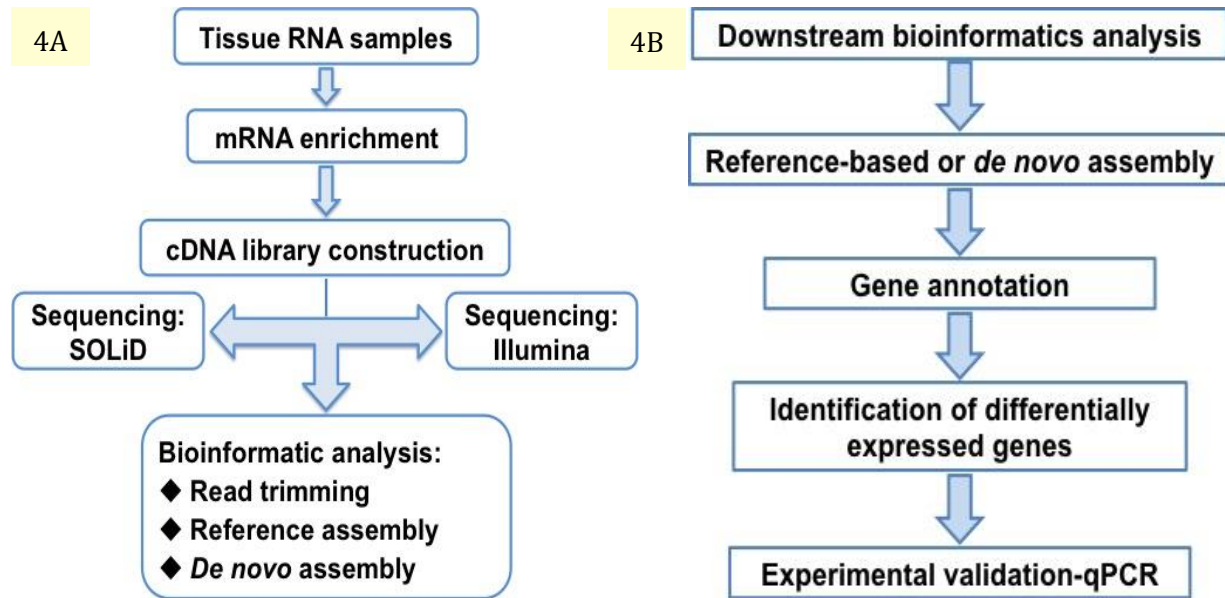


Figure 4A. The flowchart indicating general steps for RNA-seq analysis; **Figure 4B.** The flowchart shows the general downstream bioinformatics analysis of RNA-Seq data.

After raw reads are generated from high-throughput sequencing platform, the following steps are to conduct downstream analyses using bioinformatic tools. As shown in Figure 4B, four major steps are involved in a typical transcriptome expression analysis: transcriptome assembly, gene annotation, identification of differentially expressed genes, and experimental validation using qPCR. Based on the availability of reference genome for interesting species, there are two types of transcriptome assembly strategy: reference-based assembly and *de novo* assembly.

Reference assembly is conducted through aligning short reads to a reference genome and assemble the transcriptome based on that alignment, whereas *de novo* assembly is to achieve the transcriptome using only short reads and these reads are assembled based on their overlapping sequences. Because reference-based assembly requires a reference genome, it is usually more

useful for species with a reference genome sequence such as human and zebrafish. For most aquaculture species, because the reference genome sequences are not yet available, *de novo* assembly is the most often used. However, with rapid progress of genomic sequencing, reference guided assembly will become more and more popular.

There are many software packages available for the assembly of RNA-Seq datasets. The reference-based assembly can be achieved through using Cufflinks (Trapnell et al., 2010) or Scripture (Guttman et al., 2010). For *de novo* assembly, a number of software are on the list, such as ABySS (Birol et al., 2009), Trans-ABySS (Robertson et al., 2010), SOAPdenovo (Li et al., 2009), Velvet (Zerbino and Birney, 2008) and Oases (Schulz et al., 2012). Each package has its own advantages and disadvantages. However, Trinity (<http://trinityrnaseq.github.io>) (Grabherr et al., 2011) is designed specifically for transcriptome assembly with greatly improved performance compared with the other *de novo* assemblers, thus has been widely used (Chauhan et al., 2014; Haas et al., 2013; Li et al., 2014; Wu et al., 2015; Xu et al., 2015).

Generally, gene expression analysis in an RNA-Seq experiment is conducted through quantification of short reads mapped on the specific transcribed region. Specifically, this is achieved by aligning the short reads to the assembled transcriptome, followed by counting the number of reads aligned to the gene, and normalized as RPKM (Reads Per Kilobase exon Per million reads). Differentially gene expression and fold change can be then calculated through comparing RPKM values among samples under various conditions, such as between treated sample and the control sample. With statistics methods embedded in most software packages,

difference in expression levels can be tested statistically. In general, the higher the expression level, the greater likelihood for the samples to be statistically different if the fold changes are similar. That is because it is more sure when comparing 100 with 500 (a five fold difference) than comparing 1 with 5 (also a five fold difference).

Scale regeneration studies in fish

Fish scales constitute a vital physical barrier to protect fish from extrinsic invasion and environmental changes. In many cases, fresh fish scales are able to be immediately regenerated when they are either lost, severe damage or experimentally removed. Several studies have been reported for scales regeneration in fish species, including zebrafish (*Danio rerio*) (de Vrieze et al., 2011; Sire and Akimenko, 2004), medaka (*Oryzias latipes*) (Kondo et al., 2001), sea bream (*Sparus auratus*) (Vieira et al., 2011), goldfish (*Carassius auratus*) (Iimura et al., 2012; Ohira et al., 2007; Thamamongood et al., 2012) and cichlid (*Hemichromis bimaculatus*) (Sire and Géraudie, 1984). It was indicated in these studies that, when scales were removed in the experiment, both epidermal cells and superficial dermis layer were lost alongside the scales. Within a few hours, the wounds in skin were healed and re-covered by mucus and epithelia cells from the margin of wounds (Iger and Abraham, 1990; Quilhac and Sire, 1999; Vieira et al., 2011). In general, as long as the scales are removed, fresh scales grow at the maximum speed in

the first five days, and mature scales could completely replace the old ones in a few weeks (Bereiter-Hahn and Zylberberg, 1993; Ohira et al., 2007).

There are four major stages for scale regeneration: stage A, B, C and D. At stage A (day 1-2), as long as scales are removed, re-epithelialization around injured area and differentiation of scale-forming cells initiate. After that, calcified external layer (bony layer) matrix (stage B: days 3-5), basal-plate matrix (stage C: days 6-14), and finally partial mineralization of the basal plate (stage D: days 14-28) are rapidly produced (Ohira et al., 2007). Though the regeneration of scales is of great interest for researchers, the molecular mechanism underlying is not clear yet. To date, most studies of scale regeneration are focusing on morphological process using histology and immunohistochemistry methods, however, few studies have been conducted at the transcriptomic level to determine gene expression profiles during scale regeneration.

It has been reported that scales, teeth and hair share common evolutionary features (Sharpe, 2001). For example, Kondo et al., in 2001, provided a concrete evidence that EDAR (ectodysplasin-A receptor) is essential for scale development in fish, and suggests it is evolutionary conserved, because this pathway is also required for the development of other epithelial appendages in vertebrates, such as hair and teeth (Sharpe, 2001). As such, studies conducted on the developmental pathways in teeth and hair, could shed light on molecular mechanism of development of fish scales. In previous studies, teeth and hair are reported sharing a number of common developmental pathways, such as the Hedgehog, Bone morphogenetic protein (BMP) and Wnt signalling pathways, which is consistent with theory that they possess

common early development stage involving epithelial–mesenchymal interactions (Thesleff and Sharpe, 1997).

In fish, a number of studies have been conducted on the genes or pathways related to scale formation or regeneration. In cichlid fish, Sire and G éraudie (1984) observed a population of scleroblasts in the empty dermal pocket during scale regeneration process. In medaka and zebrafish, several studies have been reported that the mutation of ectodysplasin (EDA) or ectodysplasin-A receptor (EDAR) leads to the loss or abnormal scales (Harris et al., 2008; Iida et al., 2014; Kondo et al., 2001). The rs-3 (reduced scale-3) locus, which was knocked out in zebrafish, has now been shown to encode EDAR and it is involved in the TNF pathway. Its mutation resulted in the failure of fish scale development (Kondo et al., 2001). In addition, Sire and Akimenko (2004) indicated that sonic hedgehog (SHH) gene and its downstream pathway might be involved in the process of scale morphogenesis and differentiation in the formation of the epidermal basal layer . Monnot et al. (1999) reported the epidermal expression of the apolipoprotein E (APOE) gene in the process of fin regeneration and development in zebrafish, suggesting that APOE could play an important role in scale differentiation and particularly in the epidermal-dermal interaction regions. de Vrieze et al. (2011) reported the expression of MMP9 (Matrix metalloproteinase 9) and MMP2 (Matrix metalloproteinase 2) during scale regeneration in zebrafish through using qRT-PCR, situ hybridization and immunocytochemistry analyses, suggesting MMPs and anticipated osteoclasts play an crucial role in scale resorption and remodeling during regeneration. The expression patterns of osteoblast-related protein, including

runt-related gene 2 (RUNX2), secreted protein acidic and rich in cysteine (SPARC) and bone morphogenetic protein (BMP), during regeneration of goldfish scale were examined by Iimura et al. (2012), strongly suggesting the molecular mechanism of scleroblast is similar to that of osteoblasts and odontoblasts in mammalian species. Thamamongood et al. (2012) conducted the expression analysis of osteoblast-specific genes, including distal less homeobox 5 (DLX5), runt related transcription factor 2a (RUNX2A), runt related transcription factor 2b (RUNX2B), osterix, receptor activator of the NF- κ B ligand (RANKL), type I collagen, alkaline phosphatase (ALP) and osteocalcin during spontaneous scale regeneration in gold fish, showing these gene expression profile shared similar pattern with intramembranous bone regeneration. Recently, Vieira et al. (2011) conducted transcriptome level analysis at the initial stages of scale removal and re-epithelialization in sea bream using microarray as well as histology analyses, indicating a number of candidate genes involved in cell cycle regulation, cell proliferation and adhesion, immune response and antioxidant activities.

CHAPTER III. MATERIAL AND METHODS

Overall strategy

The major objective of my dissertation is to generate the skin transcriptome of channel catfish, common carp, and pleco, to explore crucial genes or gene pathways that are highly involved in scale regeneration, but not expressed in the skin of scaleless fish. In order to achieve this goal, common carps were subjected to scale regeneration experiments. Gene expression was analyzed before and after scale removal at different time points during the process of scale regeneration. Transcriptome level comparative analyses were conducted between common carp skin and channel catfish skin, as well as between pleco skin and channel catfish skin.

In this dissertation, taking advantage of the high throughput of RNA-Seq, I first generated the transcriptome assemblies of channel catfish, common carp, and pleco, and then conducted comparative transcriptome subtractions. The interspecific differentially expressed genes between the skin tissues of channel catfish and common carp were identified. In a similar fashion, comparison of the transcriptome before scale regeneration with those during scale regeneration allowed identification of differentially expressed genes during scale regeneration. Combination of both subtractive analysis, i.e., the intersection of interspecific comparative transcriptome

analysis and differentially gene expression analysis for common carp scale regeneration using RNA-Seq was crucial for providing insights into the genomic causes for the scalelessness in channel catfish, setting the foundation for the understanding of the molecular mechanisms underlying scale development in teleost fish during evolution.

Sample collection, RNA extraction and sequencing

All procedures involved in handling and treatment of fish in this study were approved by the Auburn University Institutional Animal Care and Use Committee (AU-IACUC) prior to the experiment. Fish, channel catfish, common carp, or pleco, were maintained in experiment tanks (length 3.2 m, width 0.5 m, and depth 1.0 m) and acclimatized for 7 days before experiment.

Sample collection of channel catfish skin

Channel catfish were reared at the Auburn University Fish Genetics Research Unit before sampling. Ten channel catfish were euthanized with 300 mg/l MS-222 before skin collection. Skin tissues were all collected from the left side of each fish, frozen with dry ice immediately, and then stored in a -80 °C freezer until RNA extraction. Equal amount of skin tissues from 10 fish was pooled together for RNA isolation.

Scale removal and sample collection of common carp skin

Common carps were reared at the Auburn University Fish Genetics Research Unit before experiment. Water temperature was maintained at 20-25 °C during the scale regeneration experiment. Before removing the scales and collect skin samples at each time point, experimental common carps were anesthetized using MS222 solution (100 ppm, pH 7.0). To determine differentially expressed genes during carp scale regeneration, scales were removed from a 3 cm-by-3 cm area on the left side of each fish. Skin samples were taken from the descaled area of nine fish at each time point of 0h, 12h, 24h, 3 days, 5 days, 7 days, 14 days, and 21 days after scale removal. Daily observations were made during the regeneration experiments. As reported, regenerated fresh scales started to emerge approximately 5 days after scale removal (Ohira et al., 2007) and were more visible around 12-14 days. Skin samples collected at each time point were immediately frozen with dry ice and stored in a -80 °C freezer until RNA extraction. Since nine common carps were randomly picked up at each time point, a total of 72 common carps were used in the experiment.

Sample collection of pleco skin

Pleco were euthanized with 300 mg/l MS-222 before skin tissues collection. Then skin tissues were collected from five pleco, frozen with liquid nitrogen immediately, and then stored in -80 °C freezer until RNA extraction. Equal amount of skin tissue from each of the fivepleco fish were pooled together for RNA isolation.

RNA extraction

Total RNA was isolated from the skin tissues of channel catfish, common carps and pleco using the RNeasy Plus Universal Kit (Qiagen, CA) according to the manufacturer's instructions, respectively. The extracted RNA samples were stored at -80 °C prior to RNA sequencing. Concentration and the integrity of RNA samples were measured on Bioanalyzer 2100 (Agilent technologies, USA) using an RNA Nano Bioanalysis chip. RNA-Seq was conducted as previously described (Liu et al., 2012; Liu et al., 2011).

RNA-Seq of channel catfish skin sample

RNA Sequencing of channel catfish skin samples were carried out at the HudsonAlpha Genomic Services Lab (Huntsville, AL) with Illumina HiSeq 2000 platform and Auburn University Genomics Sequencing Core Lab (Auburn, AL) with Illumina HiSeq 1500 platform. cDNA libraries were prepared using the Illumina TruSeq RNA Sample Preparation Kit (Illumina, San Diego, CA) according to the protocol. The libraries were amplified with 15 cycles of PCR and contained indexes within the adaptors. Amplified library products were quantified and diluted using KAPA kit. In the end, the libraries were sequenced using next generation sequencing platform with 100 bp paired-end (PE) reads.

RNA-Seq of common carp skin sample

Same as channel catfish, sequencing libraries of common carps were prepared using the Illumina TruSeq RNA Sample Preparation Kit (Illumina, San Diego, CA) according to the TruSeq protocol at Auburn University Genomics Sequencing Core Lab (Auburn, AL). The

libraries were amplified with indexes and adaptors attached by PCR with 15 cycles. Amplified library yields were quantified and diluted using KAPA kit. In the end, the libraries were sequenced using Illumina HiSeq 1500 platform with 100 bp lengths and paired-end (PE) reads at Auburn University Genomics Sequencing Core Lab (Auburn, AL).

RNA-Seq of pleco skin sample

Library preparation and RNA-Seq of pleco skin samples were carried out at the HudsonAlpha Genomic Services Lab (Huntsville, AL). The libraries were prepared using Illumina TruSeq RNA Sample Preparation Kit (Illumina, San Diego, CA) and the amplified library yields were quantified and diluted using KAPA kit. Then the RNA-Seq was performed using next generation sequencing platform Illumina HiSeq 2000 with 100 bp paired-end (PE) reads.

Assembly and annotation of transcriptomes of channel catfish skin, common carp skin, and pleco skin

De novo assembly of sequencing short reads

Short reads were assembled from sequenced channel catfish, common carp, and pleco, respectively. Sequencing reads were generated in FASTQ format with the feature of sequence order, quality scores and paired reads information. Raw reads were trimmed using Trimmomatic (<http://www.usadellab.org/cms/?page=trimmomatic>) (Bolger et al., 2014) with the quality control

of removing adaptor sequences, ambiguous nucleotides, quality scores less than 15 and reads length below 30 bp. The command were listed as below: “Java -jar trimmomatic-0.30.jar PE --phred33 -trimlog trimlogfile_name inputdata_forward.fq inputdata_reverse.fq output_forward_paired.fq output_forward_unpaired.fq output_reverse_paired.fq output_reverse_unpaired.fq LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:30”, the option “LEADING:3” means remove leading low quality or N bases below quality 3; the option “TRAILING:3” means remove trailing low quality below quality 3; the option “SLIDINGWINDOW:4:15” means scan the read with a 4-base wide sliding window, cutting when the average quality per base drops below 15; the option “MINLEN:30” means drop reads less than 30 bp.

In order to reduce memory requirement and improve assembly efficiency for the large RNA-seq data sets, data normalization was performed after trimming (Haas et al., 2013). The command for data normalization is

```
“TRINITY_RNASEQ_ROOT/util/insilico_read_normalization.pl --seqType fq --JM 400G --max_cov 100 --left input_left.fq --right input_right.fq --PARALLEL_STATS -output output_file --CPU 10”.
```

The option “--seqType” are the type of input reads; the option “--max_cov 100”, means targeted maximum coverage for reads; the option “--PARALLEL_STATS”, means the the paired reads statistics generated in this step.

After normalization, *de novo* assembly was performed using Trinity (Grabherr et al., 2011)

with command of “Trinity --seqType fq --JM 400G --CPU 10 --left
input_forward_paired.normalized.fq --right
input_reverse_paired.normalized.fq -output output_file”.

Gene annotation

The assembled contigs for each species were used as queries, respectively, to search against the National Center for Biotechnology Information (NCBI) Non-Redundant (NR) protein database and Ensemble zebrafish protein database using the BLASTX program. The cutoff E-value was set at 1e-5, and only the best match of gene ID and gene description was annotated for each contig.

Interspecific comparative transcriptome analysis

Interspecific comparative transcriptome analysis between common carp skin and channel catfish skin

Interspecific comparative skin transcriptome analysis between common carp (scaled) and channel catfish (scaleless) were conducted using TBLASTX with the maximal E-value of 1e-5. Interspecific comparative transcriptome analysis were conducted with the following steps: 1) A list of unique protein-coding transcripts from the common carp skin transcriptome were compiled and queries against the channel catfish skin transcriptome; 2) For carp contigs with no

matches to the channel catfish transcriptome assembly but with matches to the NR database, the sequences of the top hits were retrieved from the NR database and used to query the channel catfish skin transcriptome assembly. 3) Those carp transcripts that remained unidentified were identified as genes that were expressed in the carp skin, but not expressed in the channel catfish skin.

Interspecific comparative transcriptome analysis between pleco skin and channel catfish skin

Comparative transcriptome analysis between scaled pleco skin and scaleless channel catfish skin were conducted using BLAST program with the cut off E-value of $1e-5$. Interspecific comparative transcriptome analysis were performed with the following steps which is similar to the comparative analysis between common carp skin and channel catfish skin: 1) A list of unique protein-coding transcripts compiled from pleco skin transcriptome were used as queries to search against the channel catfish skin transcriptome; 2) For pleco contigs with no matches to the channel catfish transcriptome assembly but with matches to the NR database, the sequences of the top hits were retrieved from the NR database and used as queries to BLAST against the channel catfish skin transcriptome assembly. 3) Those pleco transcripts in 'no hit' pool were identified as genes that only were expressed in the pleco skin, but not expressed in the channel catfish skin.

Genes differentially expressed for carp scale regeneration as analyzed by RNA-Seq

Analysis

Identification of differentially expressed contigs

In order to identify differentially expressed genes in scale regeneration of common carp, the trimmed reads of each time point were mapped onto the assembled reference transcriptome using CLC Genomics Workbench. The mapping parameters were set as: a minimum of 90% of the length aligned and a maximum of two mismatches. After counting the total mapped reads for each transcript, the normalized RPKM (Reads Per Kilobase of exon model per Million mapped reads) was calculated, and then fold changes was determined based on the normalized RPKM of each sample. Differentially expressed genes were defined as at least two-fold change in expression and FDR (false discovery rate) corrected p-value < 0.05. The significant expressed contigs with previously annotated gene name were carried forward for further analysis. The following functional GO terms and related pathways of the differently expressed genes were identified by GO analysis, enrichment analysis and manual literature review.

Gene ontology analysis

In order to identify the ontology of genes that are only expressed in common carp skin but not in channel catfish skin, as well as the genes that are only expressed in pleco skin but not in channel catfish skin, and the differentially expressed genes during common carp scale regeneration. Gene Ontology (GO) annotation analysis was performed based on the zebrafish

BLAST results using Blast2GO (version 3.0), which is an automated and high throughput tool for the gene ontology terms analysis (Conesa et al., 2005). The annotation result was categorized with respect to Biological process, Molecular function, and Cellular component at level 2.

Enrichment analysis

In order to identify overrepresented GO terms in the differentially expressed gene set compared with the broader reference assembly, enrichment analysis of differentially significant expressed GO terms was performed using Ontologizer 2.0. The Parent-Child-Union method and Term-For-Term method with a Benjamini-Hochberg multiple testing correction (Bauer et al., 2008) were used for the enrichment analysis, respectively. GO terms for each gene were obtained by applying zebrafish annotations for the UniGene set. The differences of the frequency of gene ontology terms in the differentially expressed genes were compared with the overall common carp skin transcriptome assembly. The threshold was set as FDR value < 0.05 .

CHAPTER IV. RESULTS

Generation of short reads from channel catfish skin, common carp skin and pleco skin

Sequencing of short reads from channel catfish skin

Illumina-based RNA-sequencing (RNA-seq) was conducted to obtain short reads of transcripts from pooled skin samples of channel catfish. Sequencing was performed using Illumina HiSeq 2000 platform from HudsonAlpha Genomic Services Lab and Illumina HiSeq 1500 platform from Auburn University Genomics Sequencing Core Lab. A total of 219.7 million 100 bp paired-end high quality reads were generated from channel catfish skin sample (Table 2). After removing ambiguous nucleotides, low-quality sequences (quality scores < 15) and sequence short reads length shorter than 30 bp, the remaining high quality reads were obtained for assembly and further analysis.

Table 2. Summary of data generated for channel catfish skin transcriptome.

Platform	Sequencing label	No. of raw reads	Read length (bp)	No. of bases (Gbp)
HudsonAlpha	S7	96,727,002	100	9.7
	S8	96,874,116	100	9.7
Auburn	S1	26,091,876	100	2.6
Total	-	219,692,994	100	22.0

Sequencing of short reads from common carp skin

Illumina sequencing (RNA-seq) was carried out on pooled common carp skin samples. Reads from different time points samples were distinguished through using of multiple identifier (MID) tags. A total of 467.6 million 100 bp high quality reads were generated from common carp skin samples on an Illumina HiSeq 1500 platform (Table 3). Greater than 41 million reads were generated for each of the eight libraries. After trimming raw reads, the remaining high-quality reads were carried forward for assembly and analysis.

Table 3. Summary of data generated for common carp skin transcriptome.

Sequencing label	No. of raw reads	Read length (bp)	No. of bases (Gbp)
P1	56,134,924	100	5.6
P2	47,517,358	100	4.8
P3	59,259,544	100	5.9
P4	77,327,322	100	7.7
P5	84,498,032	100	8.4
P6	49,176,462	100	4.9
P7	41,332,026	100	4.1
P8	52,396,968	100	5.2
Total	467,642,636	100	46.6

Sequencing of short reads from pleco skin

RNA sequencing from pooled pleco skin samples was conducted to obtain short reads of transcripts. Sequencing was applied using Illumina HiSeq 2000 platform from HudsonAlpha Genomic Services Lab. A total of 213.8 million 100 bp paired-end high quality reads were generated from pleco skin sample (Table 4). After trimming, high quality reads were obtained for assembly and further analysis.

Table 4. Summary of data generated for pleco skin transcriptome.

Platform	Sequencing label	No. of raw reads	Read length (bp)	No. of bases (Gbp)
HudsonAlpha	S5	107,286,888	100	10.7
	S8	106,546,826	100	10.6
Total	-	213,833,714	100	21.3

***De novo* assembly of transcriptomes of the channel catfish skin, common carp skin, and pleco skin**

De novo assembly of channel catfish skin transcriptome

De novo assembly of the expressed short reads of channel catfish skin transcriptome was conducted using Trinity. There were 298,183 contigs assembled with an average length of 782 bp. The N50 contig size is 1,416 bp; the maximum contig length is 45,625 bp; and 63,505 contigs were longer than 1,000 bp. The assembly statistics were summarized in Table 5.

Table 5. Summary of *de novo* assembly for channel catfish skin transcriptome.

***De novo* assembly of Illumina sequence data from channel catfish skin using Trinity.**

Number of contigs	298,183
Maximum length (bp)	45,625
Large contigs (≥ 1000 bp)	63,505
N50 (bp)	1,416
Average contig length (bp)	782.26

De novo assembly of common carp skin transcriptome

In order to obtain a comprehensive and reliable assembly, Trinity was used for *de novo* assembly of the expressed short reads of common carp skin transcriptome. Trinity generated approximately 395,135 contigs with an average length of 878 bp with a N50 contig size of 1,700 bp. A total of 102,725 contigs were longer than 1,000 bp, allowing for reliable representations of transcriptome, and the maximum contig length was 49,445 bp. The assembly statistics were summarized in Table 6.

Table 6. Summary of *de novo* assembly for common carp skin transcriptome.

***De novo* assembly of Illumina sequence data from common carp skin using Trinity.**

Number of contigs	395,135
Maximum length (bp)	49,445
Large contigs (≥ 1000 bp)	102,725
N50 (bp)	1,700
Average contig length (bp)	878

De novo assembly of pleco skin transcriptome

Similarly, *de novo* assembly of short reads was performed for pleco skin transcriptome using Trinity. A total of 265,414 contigs were assembled with an average length of 772 bp and a N50 size of 1,465 bp. A total of 54,921 contigs were longer than 1,000 bp and the maximum contig length was 20,992 bp. The assembly statistics were summarized in Table 7.

Table 7. Summary of *de novo* assembly for pleco skin transcriptome.

<i>De novo</i> assembly of Illumina sequence data from pleco skin using Trinity.	
Number of contigs	265,414
Maximum length (bp)	20,992
Large contigs (≥ 1000 bp)	54,921
N50 (bp)	1,465
Average contig length (bp)	771.89

Gene identification and annotation of channel catfish skin transcriptome, common carp skin transcriptome and pleco skin transcriptome

Gene identification and annotation of channel catfish skin transcriptome

In order to annotate the channel catfish skin transcriptome assembly for the downstream gene expression analysis, the assembled contigs were used as queries to search against the NCBI Non-Redundant (NR) protein database using BLASTX program with E-value cutoff of $1e-5$, and only the best match was determined as gene annotation for each contig. Of 298,183 assembled

contigs for the channel catfish skin transcriptome, a total of 85,404 contigs had significant BLAST hits against the NR protein database, corresponding to 20,474 unique proteins (Table 8). A total of 4,524 contigs had unigene hits from NR protein databases to unnamed, uncharacterized and hypothetical proteins in the databases.

Table 8. Gene annotation of channel catfish skin transcriptome.

Summary of gene identification and annotation of channel catfish assembled contigs	
Number of contigs	298,183
Number of contigs with putative gene matches	85,404
Unigene matches	20,474
Known gene matches	15,950
Unname, Uncharacterized, or hypotetical gene matches	4,524

Gene identification and annotation of common carp skin transcriptome

Similarly, in order to annotate the common carp skin transcriptome assembly for downstream interspecific comparative transcriptome analysis and differentially expressed genes during common carp scale regeneration, the assembled contigs were used as queries to search against the NCBI Non-Redundant (NR) protein database using BLASTX program with E-value cutoff of $1e^{-5}$, and only the best match was annotated determined as gene annotation for each contig. Of the 395,135 assembled contigs, a total of 135,883 common carp contigs had significant BLAST hits against the NR database, corresponding to 25,136 unique proteins, with 19,901 known genes and 5,235 unnamed or uncharacterized or hypothetical proteins (Table 9).

Table 9. Gene annotation for common carp skin transcriptome.

Summary of gene identification and annotation of common carp assembled contigs	
Number of contigs	395,135
Number of contigs with putative gene matches	135,883
Unigene matches	25,136
Known gene matches	19,901
Unname, Uncharacterized, or hypothetical gene matches	5,235

Gene identification and annotation of pleco skin transcriptome

For pleco skin transcriptome, similarly, BLASTX was performed to annotate assembled contigs of pleco for downstream interspecific comparative analysis. After BLAST annotation, a total 265,414 contigs have significant BLAST hits against 21,105 unique genes retrieved from NR protein database, with an E-value cutoff of $1e-5$. Among these 21,105 unigene matches, 16,476 genes were known gene matches, while 4,629 contigs had hits to unnamed or uncharacterized or hypothetical proteins (Table 10).

Table 10. Gene annotation for pleco skin transcriptome.

Summary of gene identification and annotation of pleco assembled contigs	
Number of contigs	265,414
Number of contigs with putative gene matches	76,580
Unigene matches	21,105
Known gene matches	16,476
Unname, Uncharacterized, or hypothetical gene matches	4,629

Differentially expressed genes during carp scale regeneration

During scale regeneration at all the time points, a total of 1,173 genes (unique annotated contigs with significant BLAST identities) were identified as differentially expressed genes, with greater than 2-fold expression difference and FDR corrected p-value of less than 0.05. Of these, a total of 638 (54.4%) genes were differentially expressed at 12 h after scale removal, with 422 up-regulated genes and 216 down-regulated genes (Table 11). Similarly, a total of 513 (43.7%), 477(40.7%), 394(33.6%), 331(28.2%), 343(29.2%) and 213(18.2%) genes showed differential expression at 24 h, 3 days, 5 days, 7 days, 14 days, and 21 days after scale removal, respectively (Table 11). As shown in Table 11, At 24 h, 332 genes were up-regulated, and 181 genes were down regulated. At 3 days, 228 genes were up-regulated, and 249 genes were down-regulated. At 5 days, as many as 162 genes were up-regulated and 232 genes were down-regulated. At 7 days, a total of 97 genes were up-regulated and 234 genes were down-regulated. After 14 days since scales removal, 118 genes were up-regulated and 225 genes were down-regulated. At 21 days, only 59 genes were up-regulated and 154 genes were down-regulated.

Table 11. The number of differentially expressed genes at different time points during scale regeneration.

Time point	12h	24h	3d	5d	7d	14d	21d
Up-regulated	422	332	228	162	97	118	59
Down-regulated	216	181	249	232	234	225	154
Total (DEG)	638	513	477	394	331	343	213

Gene ontology and enrichment analysis of differentially expressed genes during common carp scale regeneration

GO analysis was performed using differentially expressed genes during carp scale regeneration. GO term distribution showed that cellular process (GO:0009987), binding (GO:0005488), and cell (GO:0005623) were the most common annotation terms in Biological Process, Molecular Function, and Cellular Component GO term categories at level 2, respectively (Figure 5).

Enrichment analysis with Parent-Child-Union method detected a total of 219 significantly overrepresented GO terms, with FDR-corrected P value < 0.05 . In brief, the top five significant GO terms of biological processes include hydrogen transport, translation, proton transport, generation of precursor metabolites and energy, and ATP metabolic process. The top five significant GO terms for cellular component include respiratory chain, proteasome complex, macromolecular complex, proton-transporting two-section ATPase complex, and mitochondrial membrane part. The top three of molecular function terms include ion and molecule transporting such as hydrogen ion transmembrane transporter activity, RNA binding, and structural molecule activity (Table 12).

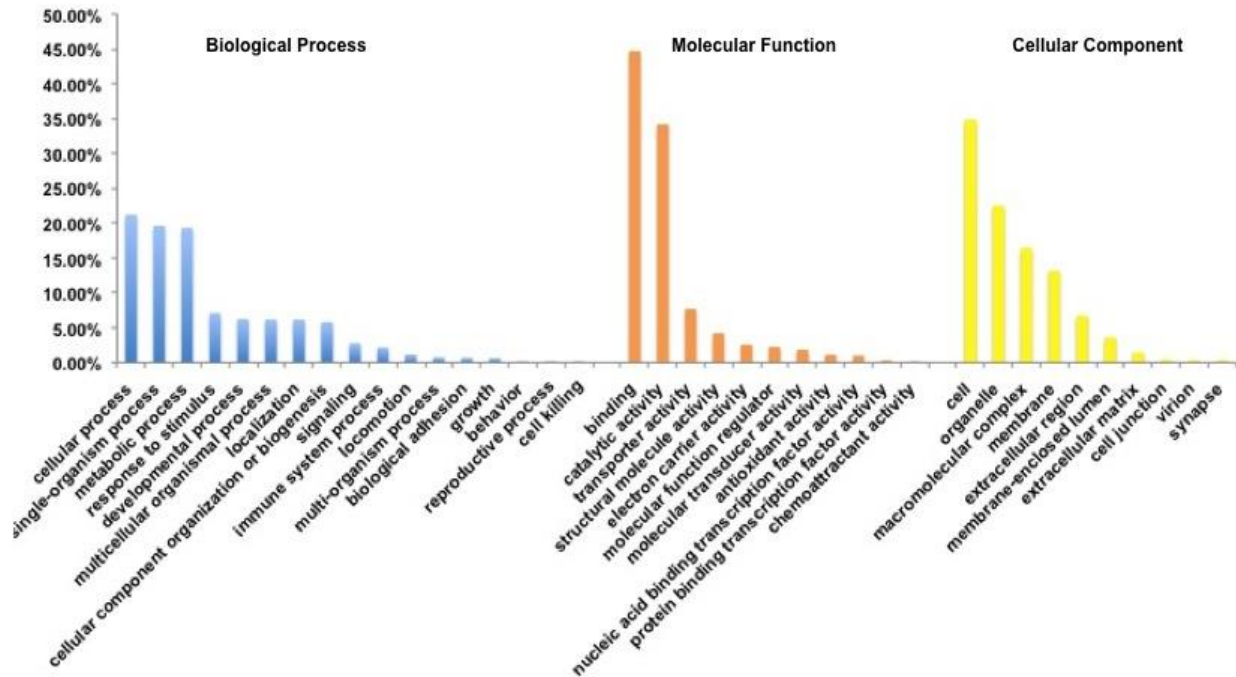


Figure 5. Gene ontology (GO) term categorization and distribution of all differently expressed genes in common carp skin during scale regeneration process. GO-terms were processed using Blast2GO and categorized at level 2 under biological process, molecular function and cellular component GO term categories, respectively.

Table 12. Summary of GO term enrichment analysis using Parent-Child-Union method by Ontologizer 2.0. $FDR \leq 0.05$ was considered significant. Pop. count is the number of genes associated with the term in the population set. Study count is the number of genes associated with the term in the study set.

GO ID	Name	NSP	Adj P-value	Rank	Pop. Count	Study Count
GO:0015078	Hydrogen ion transmembrane transporter activity	M	7.39E-19	1	79	37
GO:0070469	Respiratory chain	C	7.39E-19	2	37	22
GO:0003723	RNA binding	M	1.22E-15	3	463	66
GO:0006818	Hydrogen transport	B	2.26E-15	4	49	22
GO:0000502	Proteasome complex	C	2.26E-15	5	53	28
GO:0006412	Translation	B	4.55E-14	6	313	56
GO:0032991	Macromolecular complex	C	5.97E-13	7	2140	218
GO:0015992	Proton transport	B	7.05E-13	8	49	22
GO:0005198	Structural molecule activity	M	2.39E-12	9	331	59
GO:0016469	Proton-transporting two-sector atpase complex	C	3.77E-12	10	43	20
GO:0006091	Generation of precursor metabolites and energy	B	9.67E-12	11	114	33
GO:0044455	Mitochondrial membrane part	C	1.22E-11	12	56	20
GO:0046034	ATP metabolic process	B	1.15E-09	13	33	18

As shown in Table 13, based on GO term annotation, enrichment analysis and manual literature search, representative key genes were identified and classified into seven categories: 1) immune response; 2) wound healing; 3) metabolism; 4) Scale/bone/hair/tooth/fin regeneration or development related; 5) Collagen trimer/binding; 6) cytoskeleton; and 7) Calcium ion binding.

Table 13. Differentially expressed genes in common carp skin during scale regeneration in seven functional classifications. Gene expressions were presented as fold-change relative to control (0h) samples (only absolute fold change value ≥ 2 was shown here). The value of ‘0’ indicated that no transcripts were detected or the RPKM equal to zero after normalization in the RNA-Seq dataset.

Carp Contig ID	Gene Description	Fold Change (vs. 0h)						
		12h	24h	3d	5d	7d	14d	21d
<i>Immune response</i>								
comp165462_c0_seq1	Apolipoprotein M						0/184	
comp204590_c1_seq4	Barrier-to-autointegration factor	6.16	11.65	10.47				
comp205370_c0_seq3	CC chemokine-1	38.89	42.06					
comp199635_c0_seq4	C-C motif chemokine 19-like precursor	-64.83	-29.17	-38.26	-22.47	-62.6		-11.28
comp206307_c0_seq5	C-C motif chemokine 27 precursor			5.15	4.99	4.7	3.36	
comp184731_c0_seq1	Chemokine (C-C motif) ligand 19b precursor	-14.56	-11.46	-10.3	-25.25	-18.17	-7.28	-3.15
comp203349_c1_seq12	Chemokine (C-X-C motif) ligand 18b precursor	-4.36	-2.74	-4.03	-2.9	-2.32		2.33
comp211696_c0_seq2	Cis-aconitate decarboxylase	177.29	189.31					
comp210078_c1_seq8	Complement component C7 isoform X2	-5.14						
comp197403_c0_seq3	CXC chemokine cxcb2		9.96	6.19				
comp209707_c1_seq1	CXC chemokine receptor-1	163.08	118.52					
comp143005_c0_seq1	C-X-C motif chemokine 11		34.02					
comp197403_c0_seq10	C-X-C motif chemokine 11-like	6.33	11.4	5.71				
comp198199_c0_seq3	C-X-C motif chemokine 3				18.69			
comp180678_c0_seq1	Deoxynucleoside triphosphate triphosphohydrolase SAMHD1			71.24				
comp200487_c0_seq1	G-protein coupled receptor 84 isoform X1	107.96	114.6					
comp176038_c0_seq2	Interferon-induced transmembrane protein 1	622.3	436.33	9.57				
comp206458_c2_seq1	Interferon-induced transmembrane protein 3		8.2					
comp211175_c0_seq2	Interleukin 1 beta 2-2		25.28					

comp197259_c0_seq7	Interleukin 8	4.63	7.99	-4.24	-4.82		-4.64	
comp208181_c0_seq2	Interleukin enhancer binding factor 2	3.49	2.72					
comp211175_c0_seq6	Interleukin-1 beta	124.72	413.78					
comp213585_c1_seq12	Interleukin-1 receptor 2	148.88						
comp205911_c0_seq2	Interleukin-6a		9.68					
comp211021_c1_seq14	MHC class I antigen			2.02	2.2		2.53	2.3
comp211021_c1_seq9	MHC class I protein					2.14	3.95	3.83
comp206430_c1_seq8	MHC class II alpha chain	-3.6	-3.42				2.02	
comp206430_c1_seq6	MHC class II antigen, partial	-11.88	-150.3	-153.0	743/0	743/0	3.73	743/0
comp186994_c0_seq1	MHC class II, partial	-19.22	1729/0	-910.4	1729/0	1729/0	-2.29	1729/0
comp190523_c0_seq2	Polymeric immunoglobulin receptor-like	65.63	55.54					
comp213814_c2_seq3	Protein kinase HCK	18.03	12.61					
comp212048_c2_seq7	Putative Tumour necrosis factor alpha two		0/239					
comp215119_c0_seq7	Pyrin-like		10.81	8.42				
comp203318_c0_seq8	Transmembrane protein 173					-3.37	-2.59	
comp206756_c0_seq10	Tyrosine-protein kinase Fgr	14.45	10.79					
comp201905_c0_seq3	Tumor necrosis factor-inducible gene 6 protein precursor	26.14	35.25					
comp199657_c0_seq1	Invariant chain like protein 1					54	8.93	21.09

Response to wound/wound healing

comp202492_c1_seq2	Chemokine CCL-C11a precursor	-6.69	-3.31	-2.64				
comp208790_c0_seq3	Coagulation factor VIII	29.1						
comp182241_c0_seq1	Fibrinogen beta chain precursor						0/366	
comp165150_c0_seq1	Fibrinogen gamma chain precursor						0/388	
comp205890_c0_seq11	Inducible nitric oxide synthase	0/1191	0/1747					
comp211878_c1_seq2	Lectin, galactoside-binding, soluble, 2a		-3.22					
comp210632_c0_seq8	Prostaglandin-endoperoxide synthase 2 precursor	3.79						
comp200710_c0_seq3	Sdc4 protein, partial	3.98						
comp201964_c0_seq1	Small gtpase rhoa	2.93						2.76

comp202166_c0_seq1	Ras-related C3 botulinum toxin substrate 2	2.98						
comp208942_c0_seq1	Rho-related gtp-binding protein rhoc	4.84						
<i>Metabolism</i>								
comp203855_c0_seq1	3-ketoacyl-coa thiolase, mitochondrial					-3.4	-3.4	
comp208834_c0_seq2	Acyl-protein thioesterase 2		2.79	3.45				
comp202055_c0_seq3	Adenylate kinase 2, mitochondrial isoform X1			3.3				
comp212225_c0_seq2	Adenylosuccinate synthase, like	2.85						
comp206774_c0_seq9	Adenylosuccinate synthetase isozyme 1				-17.67			
comp203727_c0_seq2	Aldehyde dehydrogenase family 9 member A1-A	-7.41	-10.65					
comp196224_c0_seq2	Amylase-3 protein precursor							0/345
comp211723_c0_seq1	ATP synthase subunit alpha, mitochondrial					-4.55	-5.33	-4.03
comp199219_c1_seq1	Atp2a1 protein		-2.01	-9.49	-4.53	-4.5	-4.79	
comp199219_c1_seq2	Atpase, Ca ⁺⁺ transporting, cardiac muscle, fast twitch 1	-3.8	-2.21	-25.22	-9.13	-6.99	-4.08	-3.13
comp208939_c0_seq2	ATP synthase subunit beta, mitochondrial				-2.26	-4.63	-4.66	-3.74
comp95151_c0_seq1	ATP-dependent 6-phosphofructokinase, muscle type	-6.6	-4.08	-31.95	-8.07	-8.55	-4.51	
comp196427_c0_seq7	Cholesterol 25-hydroxylase-like protein	3.31	3.37					
comp195926_c0_seq3	Cyclin-selective ubiquitin carrier protein E2-C			6.44	4.82			
comp201777_c0_seq27	Cytidine deaminase	130.19	100.22					
comp208677_c0_seq3	Cytosolic non-specific dipeptidase isoform X1		3.96	6.1				
comp201669_c0_seq1	DEAD (Asp-Glu-Ala-Asp) box polypeptide 18	3.81	3.02					
comp208568_c0_seq3	Diacylglycerol O-acyltransferase 2	-8.22		-11.19	-9.31	-36.64	-6.01	
comp198972_c0_seq9	Epididymal secretory protein E1 precursor	3.8	2.95	2.26				
comp203580_c0_seq2	Eukaryotic initiation factor 4A-III-like	2.34						
comp211370_c0_seq2	G1 to S phase transition 1, like	4.27	2.92					
comp197709_c0_seq14	Gig2-like protein cauib		79.1	174.8	68.6			83.32
comp196951_c1_seq5	Glutaredoxin-1	3.45	5.31	4.09	2.49	2.16		
comp200038_c0_seq1	Glutathione S-transferase theta 1b	-9.33						
comp205057_c3_seq1	Glutathione-specific gamma-glutamylcyclotransferase 1	2.62						

comp200210_c0_seq1	Guanine nucleotide-binding protein G(i) subunit alpha-1-like	4.14							
comp202992_c1_seq1	Guanine nucleotide-binding protein G(i) subunit alpha-2	3.23							
comp187893_c0_seq1	Hydroxysteroid (17-beta) dehydrogenase 12a	21.13	11.69	9.84	9.17				
comp196296_c1_seq3	Intraflagellar transport protein 22 homolog					-2.19	-8.58	-4.11	-3.46
comp192355_c1_seq3	Adenylate kinase isoenzyme 1	-3.35	-3.01	-12.14	-5.33	-8.67	-6.96	-4.86	
comp207418_c1_seq5	Lipid phosphate phosphohydrolase 1 isoform X1	-6.43	-3.5	-4.82	-5.23	-4.91			
comp204631_c0_seq4	L-threonine 3-dehydrogenase, mitochondrial	6.7	5.68						
comp199551_c0_seq2	Lysozyme C	8.71							
comp209290_c2_seq1	Myh9 protein	-3.04	-2.35	-2.28	-2.32				
comp201210_c0_seq4	Myosin heavy chain fast skeletal type 1								6.29
comp200305_c0_seq2	Myosin heavy chain, fast skeletal muscle								5.64
comp208489_c0_seq5	NAD kinase b isoform X2	27.26							
comp213641_c0_seq1	Phosphofructokinase, muscle b isoform X1	-4.64	-4.91	-35.73	-12.03	-12.7	-10.15	-4.88	
comp206390_c0_seq1	Phosphoglucosmutase-1	-3.58	-2.45	-6.96	-4.94	-4.67	-3.34		
comp203069_c0_seq1	Phosphoglycerate mutase 2	-4.33	-3.67	-36.13	-16.27	-10.34	-6.46	-3.50	
comp203760_c1_seq1	Phosphoglycolate phosphatase	14.76	7.84						
comp199076_c0_seq5	Phosphoserine phosphatase isoform X1					-15.78	-27.74		
comp214337_c1_seq1	ATP-dependent RNA helicase DDX5 isoform X2	2.20							
comp205786_c0_seq4	Protein disulfide-isomerase A3 precursor			2.00					
comp201725_c0_seq1	Pyruvate dehydrogenase E1 alpha 1 isoform X1							-3.81	
comp215596_c2_seq2	Pyruvate dehydrogenase E1 component subunit beta, mitochondrial			-2.68	-2.43	-5.86	-6.65	-3.13	
comp201211_c0_seq1	Rab32 protein	5.57							
comp198901_c0_seq2	Ras-related protein Rab-10	4.28							
comp211262_c1_seq6	Ras-related protein Rab-18-B	3.59							
comp215366_c2_seq4	Ras-related protein Ral-B	2.79							
comp198555_c0_seq3	Si:ch73-199e17.1 isoform X1				6.17				
comp187059_c0_seq1	Slow myosin heavy chain 2			-4.61		-7.02	-5.98	-2.78	
comp203791_c0_seq14	Spermidine/spermine N1-acetyltransferase	7.90	4.45			2.11			
comp205448_c0_seq1	Triosephosphate isomerase B			-3.22	-2.94	-4.05	-3.79	-3.61	

comp207859_c0_seq4	Ubiquinol-cytochrome c reductase core protein I					-7.29	-27.31	-11.09
comp211335_c0_seq1	Very-long-chain enoyl-coa reductase			-32.5	-8.20	-9.25	-8.70	
comp194279_c1_seq2	Zgc:158768	-7.19	-15.56	-10.14	-2.86	-2.22		
<i>Scale/bone/hair/tooth/fin regeneration or development related</i>								
comp203286_c0_seq5	Apolipoprotein E	-8.13	-8.17			2.39	2.31	
comp192992_c0_seq1	Apolipoprotein A-I						0/483	
comp192992_c0_seq2	Apolipoprotein aib1						0/351	
comp187501_c0_seq1	Apolipoprotein C Ia						32.22	
comp187501_c0_seq2	Apolipoprotein C Ib						348.06	
comp210669_c1_seq7	Aquaporin 3					2.22	3.53	2.13
comp208252_c0_seq1	Aquaporin 3b					29.82	24.81	21.31
comp176367_c0_seq1	Catenin beta-1 isoform X2			2.82	3.5			
comp193535_c0_seq5	Cathepsin B preproprotein	2.44	2.7	2.99				
comp208662_c0_seq6	Cathepsin K	2.36	3.67	3.26	2.46	2.76	3.37	3.27
comp216088_c0_seq4	Cathepsin L1 isoform X1	5.61	6.63	5.94	4.18	2.64	2.53	
comp208542_c5_seq3	Collagen alpha-1(X) chain				134.85	664.47	3353.5	
comp216017_c0_seq4	Collagen alpha-1(XVII) chain isoform X1			2.98				
comp188889_c0_seq1	Collagen type I alpha 1	-3.7	-3.79	-2.71			2.74	
comp212701_c0_seq2	Decorin variant 1	-3.24	-2.18	-2.53	-2			
comp203225_c0_seq1	Ectonucleoside triphosphate diphosphohydrolase 5					0/379	0/540	0/587
comp206788_c0_seq5	Envoplakin	10.05	6.06		5.68			
comp213870_c0_seq1	Epidermal growth factor receptor precursor	3.17						
comp199420_c0_seq2	Epithelial cell adhesion molecule					2.33		
comp205338_c0_seq3	Eukaryotic translation initiation factor 3 subunit H-A		2.19	2.18				
comp200220_c1_seq2	Galectin-1 isoform X1	157.98	226.12	34.26				
comp198754_c1_seq1	Gp130	6.1	5.18					
comp200591_c0_seq4	Granulin	30.85	32.08	40.87	16.93	7.52	3.5	3.22
comp209916_c0_seq3	Heat shock cognate 70	11.72	6.51	3.66	2.85	3.56	2.31	3.01

comp202671_c1_seq8	High mobility group box 1			5.42	9.88	17.6	7.45	19.83
comp206644_c0_seq10	HMG-box transcription factor Sox9b	-26.43	-21.89	-9.43	-14.56	-28.63	-9.02	
comp202266_c0_seq5	Homeobox protein Dlx3b					3.31		
comp212595_c0_seq1	Hsp60	5.63	4.17					
comp211281_c1_seq1	Interferon-induced transmembrane protein 5			583.47	648.25	524.31	343.44	
comp213869_c0_seq1	Interleukin-6 receptor subunit beta isoform X1	2.47						
comp203231_c2_seq6	Keratin	-2.65		3.6	4.06	3.76	2.8	2.06
comp200551_c0_seq3	Keratin, type I cytoskeletal 13-like				6.08	12.66	19.79	2.96
comp199449_c0_seq7	Keratin, type II cytoskeletal 8-like					2.14	2.51	
comp199449_c0_seq1	Krt5 protein			3.63	3.15			
comp210602_c1_seq1	Legumain precursor			4.02	3.52			
comp212820_c1_seq4	Lipid phosphate phosphohydrolase 3 isoform X2	2.25						
comp213452_c2_seq7	Matrix metalloproteinase 14 (membrane-inserted) beta					2.92	3.21	
comp212221_c0_seq2	Matrix metalloproteinase 2						3.18	
comp208493_c0_seq6	Matrix metalloproteinase-9 precursor	104.31	142.39	17.17	8.34	11.62		
comp209330_c0_seq5	Periostin	-14.22	-18.93	-19.32	-13.4	-10.45		
comp210120_c1_seq1	Peroxiredoxin-5, mitochondrial-like	7.05	16.09					
comp207798_c0_seq5	Probable phosphatase phospho1				0/464			
comp203635_c0_seq1	Protein odd-skipped-related 1	-9.07	-10.37	-13.07	-15.7	-6.5		
comp203707_c0_seq4	Protein wntless homolog isoform X4					2.05		
comp181242_c0_seq1	Secretory calcium-binding phosphoprotein 7 precursor				75.29	102.98	186.6	12.96
comp197988_c0_seq10	Si:dkey-22i16.3 precursor (fa93e10)				0/2055	0/1778	0/1559	
comp201964_c0_seq1	Small gtpase rhoa	2.93						2.76
comp205077_c1_seq4	Sparc	-3.67	-4.08	-3.37			3.44	
comp200039_c0_seq2	Spp1 protein precursor						210.61	
comp210930_c0_seq3	Stromal cell-derived factor 1 precursor	-3.21	-2.22	-9.01	-5.12	-6.54	-2.56	
comp192461_c1_seq1	Tenascin C			11.44	12.81	12.55	6.76	
comp210664_c0_seq3	Tnc protein				8.59	7.55		
comp211880_c0_seq7	Transforming growth factor, beta-induced						2.37	

comp213900_c1_seq1	Transmembrane protein 100	12.29	6.72					
comp204517_c0_seq8	TSC22 domain family protein 3		-2.54	-5.95	-6.5	-4.29	-3.38	-2.97
comp206898_c0_seq3	Tumor necrosis factor receptor superfamily member 11B	327.15	241.88					
comp193378_c0_seq3	Uncharacterized protein LOC335309 isoform X1	11.94	13.71	14.29	12.23	8.44		7.29
comp200001_c1_seq1	Urokinase plasminogen activator surface receptor-like		15.4	23.29				
<i>Collagen trimer/binding</i>								
comp215903_c3_seq1	Collagen alpha-2(I) chain precursor	-5.59	-5.65	-5.32	-2.03		2.61	2.22
comp215803_c0_seq3	Collagen alpha-2(V) chain precursor				2.12	2.75	2.38	
comp208549_c0_seq1	Collagen alpha-2(XI) chain isoform X5				0/646			
comp203681_c0_seq2	Collagen type I alpha 2	-3.96	-4	-2.37			3.16	
comp215832_c0_seq2	Collagen type I alpha 3	-4.68	-3.87	-2.45			2.64	
comp194466_c0_seq1	Complement C1q subcomponent subunit A precursor						2.2	
comp175852_c0_seq1	Complement C1q subcomponent subunit B			2.22	2.24		2.45	
comp176397_c0_seq1	Complement C1q subcomponent subunit C			2.29		2.38	2.65	2.11
comp205832_c1_seq4	Hsp47	-4.9						
comp208462_c0_seq1	Lumican	-4.48	-4.1	-3.65	-2.51			
comp203643_c0_seq1	Lysyl oxidase homolog 2 precursor						3.64	
comp209888_c0_seq2	Procollagen C-endopeptidase enhancer precursor						2.22	
<i>Cytoskeletal</i>								
comp201493_c0_seq1	Capping protein (actin filament) muscle Z-line, alpha 1					6.4	8.61	6.29
comp208589_c0_seq3	Tubulin alpha-1 chain		2.48	2.64	2.37			
comp203318_c0_seq5	Tubulin alpha-1A chain	7.18	4.94		3.8			
comp208589_c0_seq2	Tubulin alpha-1A chain-like isoform X2	2.39		2.52	2.58			
comp208589_c0_seq2	Tubulin alpha-1A chain-like isoform X2	2.39		2.52	2.58			
comp198555_c0_seq2	Tubulin alpha-1D chain-like isoform X1		2.58	3.26	3.59	4.69	2.12	
comp200458_c0_seq2	Tubulin beta-6 chain	20.36	19.83	8.25	5.19			
comp210995_c1_seq3	Tubulin, beta 2c	4.34	3.54	2.8	2.74			

comp210520_c0_seq4	Actin related protein 2/3 complex, subunit 5A	3.73	3.37	2.99				
comp214803_c1_seq1	Actin-related protein 2/3 complex subunit 1A		2.18					
comp200087_c1_seq2	Actin-related protein 2-A	3.92	3.7					
comp181280_c0_seq1	Beta-actin	2.22	2.24	2.09	2.23			
comp202949_c0_seq1	Cell division cycle associated 8			7.51				
comp211727_c1_seq3	Cell division cycle protein 20 homolog			8.41				
comp197134_c2_seq2	Ezrin a	27.43						
comp205597_c0_seq1	Fructose-bisphosphate aldolase B		-12.34		2.73	2.83		
comp203054_c2_seq4	Gapdh protein, partial	-5.07	-3.33	-22.75	-8.51	-5.94	-3.68	
comp212385_c0_seq2	Ivns1abpa protein	-2.39	-3.9	-4.79	-3.39	-2.65		
comp200007_c0_seq1	Mid1-interacting protein 1A	4.32						
comp212439_c1_seq3	Moesin	2.5	2.22					
comp204937_c4_seq4	Non-muscle cofilin 1	-2.17			2.11			
comp202044_c0_seq15	Profilin-1-like	2.15	2.85					-2.51
comp206755_c0_seq4	Profilin-2-like isoform X2			2.54	2.31			
comp198777_c0_seq2	Protein BRICK1 isoform X1				8.33			
comp204592_c0_seq8	Thymosin beta-11-like	-2.9	-2.35					
comp200922_c1_seq2	Tropomodulin-4	-4.32	-3.7	-14.82	-7.22	-5.94	-6.77	-3.07
comp210995_c1_seq4	Tubulin beta-4B chain-like			3.08				
comp215831_c0_seq2	Girdin	8.58	6.35					
<i>Calcium ion binding</i>								
comp196590_c0_seq1	Actn3a	-3.33		-21.37	-3.65	-4.37	-4.05	-3.82
comp203339_c0_seq9	Allograft inflammatory factor 1-like	18.35	20.44					
comp192307_c0_seq1	Alpha-actinin	-2.97		-13.74	-5.58	-17.26	-7.65	-4.75
comp204073_c1_seq2	Alpha-actinin-1 isoform X1					-4.75	-5.02	
comp211566_c0_seq5	Alpha-actinin-3-like	-5.3	-4.11	-20.95	-11.48	-10.64	-8.47	
comp203634_c0_seq9	Annexin A2a		2.45	2.18	2.04			
comp213663_c0_seq3	Cadherin-like protein 26 isoform X1	-3.13	-5.86					

comp212610_c0_seq1	Calmodulin	-2.9						
comp214911_c0_seq2	Calpain small subunit 1				2.31			
comp187084_c0_seq1	Calpain-1 catalytic subunit	-37.7	-25.4	-3.46	-3.51			
comp207461_c0_seq8	Calrl protein		2.44	2.27				
comp202096_c0_seq1	Calsequestrin 1a precursor	-5.27	-5.25	-27.15	-6.69	-5.56	-4.29	-3.09
comp207284_c0_seq4	Desmoglein 2, tandem duplicate 1 precursor				2.21			
comp207708_c2_seq2	EF-hand domain-containing protein D2	2.82						
comp208883_c0_seq2	EH-domain containing 1a	5.46						
comp201826_c0_seq1	Fast skeletal myosin light chain 1a	-3.86	-3.32	-32.63	-9.9	-9.01	-6.02	-2.08
comp202094_c2_seq1	Fast skeletal myosin light chain 3	-4.54	-3.31	-22.53	-8.32	-4.43	-2.89	
comp215454_c1_seq1	Fstl1b protein, partial					2.43	3.75	2.04
comp165049_c0_seq1	Group 10 secretory phospholipase A2				17.51	18.18	21.39	
comp199834_c0_seq1	Hypothetical protein		2.91				-2.58	
comp196050_c0_seq1	Ictacalcin	-2.29	-3.59					
comp202021_c0_seq5	Matrix metalloproteinase 13	573.98	2451.7					
comp198677_c0_seq5	M-calpain							2.66
comp202107_c2_seq5	Myosin regulatory light chain	-4.43	-3.65	-22.39	-6.37	-5.04	-3.75	
comp215269_c0_seq2	Myosin regulatory light chain 2B, cardiac muscle isoform-like			-5.31		-10.36	-14.59	
comp210191_c0_seq3	Parvalbumin	-4.96	-3.94	-27.23	-7.85	-4.96	-2.86	
comp194539_c1_seq10	Parvalbumin-2	-2.9	-3.75	-65.96	-20.4	-6.98	-15.95	-72.84
comp201172_c0_seq4	Parvalbumin-7	-2.74	-4.75	-16.66	-14.93	-17.08	-5.22	
comp210973_c0_seq2	Peptidyl-prolyl cis-trans isomerase FKBP14 precursor	-13.85						
comp211833_c0_seq2	Plastin-2	3.3	4.46					
comp213459_c1_seq1	Protein PFC0760c-like isoform X1					-41.48		
comp173334_c0_seq1	Protein S100-A11					2.76	6.63	
comp208409_c1_seq4	Protein S100-A14	2.13						
comp197406_c0_seq10	Protein S100-A16	2.29						
comp211897_c0_seq3	Reticulocalbin-3 precursor	-11.15						
comp199219_c2_seq3	Sarcoplasmic/endoplasmic reticulum calcium atpase 1-like isoform X1	-4.19	-3.24	-21.91	-32.82	-11.3	-6.84	-2.47

comp204168_c1_seq6	Sarcoplasmic/endoplasmic reticulum calcium atpase 2			-3.92	-3.58	-3.5	-2.98	
comp201716_c1_seq1	Troponin C type 1b (slow)	-2.42		-8.14	-5.72	-6.48	-13.7	-2.28
comp204865_c0_seq3	Troponin C type 2 (fast)	-4.8	-3.21	-28.02	-8.6	-8.19	-6.26	-3.64
comp194334_c0_seq2	Uncharacterized protein LOC100302470 precursor			6.41	2.97			
comp210834_c1_seq4	Uncharacterized protein LOC393488		-2.54	-5.09	-3.46	-3.82	-5.43	

Comparative subtraction of the channel catfish skin transcriptome from the common carp skin transcriptome

To provide insights into the genomic causes for the scalelessness in catfish, interspecific comparative transcriptome analysis between the scaled common carp and scaleless channel catfish was conducted, allowing identification of 836 genes (Table 14) that were expressed in the skin of common carp but not in the skin of channel catfish. The 836 genes that were only expressed in the skin of common carp but not in the skin of channel catfish were used to perform GO annotation by Blast2GO. Analysis of GO term distribution showed that single-organism process (GO:0044699), binding (GO:0005488), and cell (GO:0005623) was the most common annotation terms in Biological Process, Molecular Function, and Cellular Component GO term categories at level 2, respectively (Figure 6). The GO terms for biological process also include metabolic process, response to stimulus, biological regulation, developmental process, signaling and immune process.

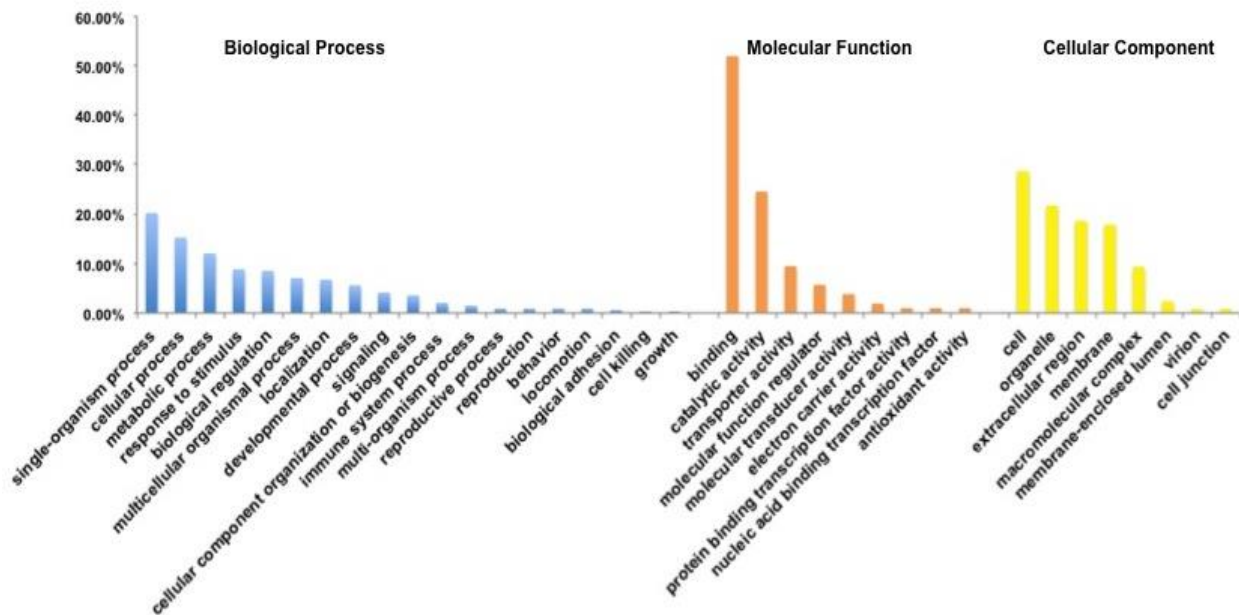


Figure 6. Gene ontology (GO) term categorization and distribution of genes expressed in the skin of common carp but not in the skin of channel catfish. GO-terms were processed using Blast2GO and categorized at level 2 under biological process, molecular function and cellular component GO term categories, respectively.

Table 14. A list of 836 genes that were expressed in the skin of common carp, but not in the skin of channel catfish.

Carp contig ID	Accession	Description
comp10_c0_seq1	XP_009290897.1	Predicted: uncharacterized protein loc103908979 [<i>Danio rerio</i>]
comp100084_c0_seq1	KFM62690.1	Hypothetical protein x975_06733, partial [<i>Stegodyphus mimosarum</i>]
comp101023_c0_seq1	XP_005172295.2	Predicted: arginine and glutamate-rich protein 1-a-like [<i>Danio rerio</i>]
comp1019_c0_seq1	XP_005813273.1	Predicted: uncharacterized protein loc102220987 [<i>xiphophorus maculatus</i>]
comp101978_c0_seq1	NP_001038261.1	Transmembrane protein 244 [<i>Danio rerio</i>]
comp102153_c0_seq1	XP_009301686.1	Predicted: uncharacterized protein loc103911417 [<i>Danio rerio</i>]
comp10267_c1_seq1	XP_002665852.3	Predicted: apolipoprotein b-100 [<i>Danio rerio</i>]
comp10344_c0_seq1	XP_003198174.1	Predicted: uncharacterized protein c17orf104 homolog isoform x2 [<i>Danio rerio</i>]
comp105495_c0_seq2	AAI71355.1	Zgc:65845 [<i>Danio rerio</i>]
comp106436_c0_seq1	XP_001341083.1	Predicted: uncharacterized protein c1orf228 homolog isoform x1 [<i>Danio rerio</i>]
comp10644_c0_seq1	XP_008421469.1	Predicted: uncharacterized protein loc103473208 [<i>poecilia reticulata</i>]
comp106848_c0_seq1	NP_571131.1	Insulin preproprotein [<i>Danio rerio</i>]
comp106849_c0_seq1	XP_004554696.1	Predicted: uncharacterized protein loc101468588 [<i>Maylandia zebra</i>]
comp107723_c0_seq1	KFM73241.1	Cuticle protein 14 isoform b, partial [<i>Stegodyphus mimosarum</i>]
comp10796_c0_seq1	KFM82825.1	Hypothetical protein x975_20306, partial [<i>Stegodyphus mimosarum</i>]
comp107995_c0_seq1	XP_005953050.1	Predicted: uncharacterized protein loc102307237 [<i>Haplochromis burtoni</i>]
comp108316_c0_seq1	CCD82054.1	Smkk7 [<i>schistosoma mansoni</i>]
comp108517_c1_seq1	CDQ79308.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp108598_c0_seq2	NP_001019992.1	Lymphotoxin alpha [<i>Danio rerio</i>]
comp108884_c0_seq1	NP_001104306.1	Uncharacterized protein loc561308 [<i>Danio rerio</i>]
comp109977_c0_seq1	NP_001070076.1	Regulator of g-protein signaling 7-binding protein a [<i>Danio rerio</i>]
comp110673_c0_seq1	DAA35165.1	Tpa_inf: tachykinin 4b [<i>osmerus mordax</i>]
comp112195_c0_seq1	CAD68054.1	Hemocyanin subunit d [<i>Nephila inaurata madagascariensis</i>]
comp112832_c0_seq1	NP_001121242.1	Uncharacterized protein loc100158321 [<i>Xenopus laevis</i>]
comp113631_c1_seq1	CDQ69101.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp114052_c0_seq2	CCV17414.1	Type i interferon f2 [<i>Oncorhynchus mykiss</i>]

comp11477_c0_seq1	XP_005173976.1	Predicted: uncharacterized protein loc100331884 isoform x1 [<i>Danio rerio</i>]
comp115109_c0_seq1	NP_001018364.1	Regulator of g-protein signaling 9-binding protein [<i>Danio rerio</i>]
comp116069_c0_seq1	KFM58514.1	Hypothetical protein x975_26526, partial [<i>Stegodyphus mimosarum</i>]
comp11678_c0_seq1	AAP32203.1	Latency associated nuclear antigen [<i>saimiriine herpesvirus 2</i>]
comp119281_c0_seq1	XP_007246658.1	Predicted: uncharacterized protein loc103045942 isoform x1 [<i>Astyanax mexicanus</i>]
comp121159_c0_seq1	XP_002602904.1	Hypothetical protein brafldraft_198506, partial [<i>Branchiostoma floridae</i>]
comp121327_c0_seq1	XP_002586138.1	Hypothetical protein brafldraft_248486, partial [<i>Branchiostoma floridae</i>]
comp121725_c0_seq1	XP_009045105.1	Hypothetical protein lotgidraft_170991 [<i>Lottia gigantea</i>]
comp123811_c0_seq1	AAW27897.1	Sjchgc09511 protein [<i>schistosoma japonicum</i>]
comp12608_c0_seq1	CDQ63921.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp127378_c0_seq1	ADV40349.1	Hypothetical protein, partial [<i>Latrodectus hesperus</i>]
comp127402_c0_seq1	XP_005168523.1	Predicted: erythroblast nad(p)(+)--arginine adp-ribosyltransferase-like [<i>Danio rerio</i>]
comp127542_c0_seq1	XP_009303403.1	Predicted: pollen-specific leucine-rich repeat extensin-like protein 2 [<i>Danio rerio</i>]
comp128389_c0_seq1	XP_006642931.1	Predicted: uncharacterized protein loc102685222, partial [<i>Lepisosteus oculatus</i>]
comp128958_c0_seq1	EFN85945.1	Hypothetical protein eai_07868, partial [<i>Harpegnathos saltator</i>]
comp129624_c0_seq1	NP_001073414.1	Uncharacterized protein loc322393 precursor [<i>Danio rerio</i>]
comp129649_c0_seq1	XP_001920424.3	Predicted: cyclin n-terminal domain-containing protein 1, partial [<i>Danio rerio</i>]
comp129740_c0_seq1	XP_002120972.3	Predicted: uncharacterized protein loc100182771, partial [<i>Ciona intestinalis</i>]
comp129740_c1_seq1	XP_004913166.1	Predicted: uncharacterized protein loc101734921 [<i>Xenopus (silurana) tropicalis</i>]
comp131099_c0_seq2	ADV40331.1	Hypothetical protein, partial [<i>Latrodectus hesperus</i>]
comp131099_c0_seq3	ADV40339.1	Hypothetical protein, partial [<i>Latrodectus hesperus</i>]
comp13118_c0_seq1	ELU08537.1	Hypothetical protein captedraft_225035 [<i>Capitella teleta</i>]
comp131980_c0_seq1	XP_009304668.1	Predicted: wd repeat-containing protein 49-like [<i>Danio rerio</i>]
comp132691_c0_seq1	XP_003442774.2	Predicted: uncharacterized protein loc100697407 [<i>Oreochromis niloticus</i>]
comp132722_c0_seq1	XP_005169552.1	Predicted: uncharacterized protein loc562542 isoform x1 [<i>Danio rerio</i>]
comp133364_c0_seq2	XP_005164821.1	Predicted: protein casc1 isoform x1 [<i>Danio rerio</i>]
comp133364_c0_seq3	AAI46743.1	Casc1 protein, partial [<i>Danio rerio</i>]
comp13394_c0_seq1	NP_001020665.2	Uncharacterized protein loc557637 [<i>Danio rerio</i>]
comp134142_c0_seq1	KFM73579.1	Hypothetical protein x975_04617, partial [<i>Stegodyphus mimosarum</i>]

comp134728_c0_seq1	XP_009289476.1	Predicted: protein starmaker-like isoform x1 [<i>Danio rerio</i>]
comp134865_c0_seq1	XP_003201483.2	Predicted: ef-hand calcium-binding domain-containing protein 6 isoform x1 [<i>Danio rerio</i>]
comp134888_c0_seq1	XP_819411.1	Hypothetical protein [<i>Trypanosoma cruzi strain cl brener</i>]
comp134998_c0_seq1	XP_007243180.1	Predicted: uncharacterized protein loc103045430 [<i>Astyanax mexicanus</i>]
comp135196_c0_seq1	NP_001103955.1	Meiosis 1 arrest protein [<i>Danio rerio</i>]
comp135506_c0_seq1	XP_005161786.1	Predicted: natterin-3-like [<i>Danio rerio</i>]
comp135969_c0_seq1	NP_001108038.1	Coiled-coil domain-containing protein 89 [<i>Danio rerio</i>]
comp136890_c0_seq1	XP_004076138.1	Predicted: uncharacterized protein loc101160338 [<i>Oryzias latipes</i>]
comp137604_c0_seq1	XP_007247455.1	Predicted: uncharacterized protein loc103027087 [<i>Astyanax mexicanus</i>]
comp137900_c0_seq1	XP_002060684.1	Gk24131 [<i>Drosophila willistoni</i>]
comp139834_c0_seq2	CAD68056.1	Hemocyanin subunit f [<i>Nephila inaurata madagascariensis</i>]
comp140027_c0_seq2	XP_009290051.1	Predicted: uncharacterized protein loc103908849 [<i>Danio rerio</i>]
comp140271_c0_seq1	XP_005162653.1	Predicted: uncharacterized protein si:ch211-146110.7 [<i>Danio rerio</i>]
comp141046_c0_seq1	XP_009293365.1	Predicted: stress response protein nst1-like [<i>Danio rerio</i>]
comp141143_c0_seq1	XP_003199320.1	Predicted: uncharacterized protein loc100535760 [<i>Danio rerio</i>]
comp141198_c0_seq1	XP_005173212.2	Predicted: uncharacterized protein loc100538111 [<i>Danio rerio</i>]
comp141263_c0_seq1	XP_009048256.1	Hypothetical protein lotgidraft_139961, partial [<i>Lottia gigantea</i>]
comp141333_c0_seq1	WP_006679318.1	Hypothetical protein, partial [<i>Paenibacillus dendritiformis</i>]
comp141369_c0_seq1	NP_991140.1	Tuberoinfundibular peptide of 39 residues precursor [<i>Danio rerio</i>]
comp141534_c0_seq1	ABU41022.1	Hypothetical protein [<i>Lepeophtheirus salmonis</i>]
comp141848_c0_seq1	XP_004556595.1	Predicted: uncharacterized protein loc101478783 [<i>Maylandia zebra</i>]
comp142159_c0_seq1	XP_002665172.1	Predicted: uncharacterized protein c8orf74 homolog [<i>Danio rerio</i>]
comp142885_c0_seq1	AAH49447.1	C10orf63 protein, partial [<i>Danio rerio</i>]
comp142919_c0_seq1	NP_001098605.1	Taste receptor, type 2, member 3 [<i>Danio rerio</i>]
comp143004_c0_seq1	XP_706556.2	Predicted: uncharacterized protein cxorf65 homolog [<i>Danio rerio</i>]
comp143770_c0_seq1	NP_001038816.1	E3 ubiquitin-protein ligase rnf182 [<i>Danio rerio</i>]
comp143829_c0_seq1	NP_001245247.1	Urotensin 2 domain containing precursor [<i>Danio rerio</i>]
comp144533_c0_seq1	XP_001333344.3	Predicted: protein maats1 isoform x1 [<i>Danio rerio</i>]
comp144928_c0_seq1	KFM77262.1	Hypothetical protein x975_16027, partial [<i>Stegodyphus mimosarum</i>]

comp145043_c0_seq2	XP_004074007.1	Predicted: uncharacterized protein loc101158037 [<i>Oryzias latipes</i>]
comp145637_c0_seq1	AAZ29482.1	Lengsin [<i>Danio rerio</i>]
comp146524_c0_seq1	XP_009304575.1	Predicted: uncharacterized protein loc798927 isoform x2 [<i>Danio rerio</i>]
comp146883_c0_seq1	XP_007060473.1	Predicted: uncharacterized protein c6orf229 homolog [<i>Chelonia mydas</i>]
comp146922_c0_seq1	KFM80340.1	Hypothetical protein x975_14306, partial [<i>Stegodyphus mimosarum</i>]
comp148055_c1_seq2	XP_007242826.1	Predicted: uncharacterized protein loc103032647 [<i>Astyanax mexicanus</i>]
comp149307_c1_seq1	XP_009031569.1	Hypothetical protein helrodraft_166387 [<i>Helobdella robusta</i>]
comp149652_c0_seq1	XP_005172124.1	Predicted: uncharacterized protein loc100538256 [<i>Danio rerio</i>]
comp149851_c0_seq1	ELR57875.1	Hypothetical protein m91_02367, partial [<i>Bos mutus</i>]
comp149857_c0_seq1	AAI16600.1	Rsph3 protein, partial [<i>Danio rerio</i>]
comp150139_c0_seq2	XP_005169824.1	Predicted: lymphocyte antigen 6f-like [<i>Danio rerio</i>]
comp150854_c0_seq1	NP_001289402.1	Small lysine-rich protein 1 [<i>Danio rerio</i>]
comp151558_c0_seq1	AFN88207.1	Integrase core domain containing protein [<i>Phaseolus vulgaris</i>]
comp151558_c1_seq1	KFD67433.1	Hypothetical protein m514_20469 [<i>Trichuris suis</i>]
comp15262_c0_seq1	KAK59965.1	Hypothetical protein az22_5144 [<i>Bordetella bronchiseptica</i> 980-2]
comp15262_c0_seq2	WP_029550480.1	Hypothetical protein, partial [<i>Staphylococcus aureus</i>]
comp153322_c0_seq1	NP_001129450.1	Uncharacterized protein loc570390 precursor [<i>Danio rerio</i>]
comp153344_c0_seq1	XP_006643111.1	Predicted: mesenteric estrogen-dependent adipogenesis protein-like [<i>Lepisosteus oculatus</i>]
comp153649_c0_seq1	XP_005275737.1	Predicted: uncharacterized protein loc101930397 [<i>Homo sapiens</i>]
comp154775_c0_seq3	EFN76041.1	Hypothetical protein eai_16144, partial [<i>Harpegnathos saltator</i>]
comp154907_c0_seq1	Q1LX29.2	Telomere repeats-binding bouquet formation protein 1 [<i>Danio rerio</i>]
comp155316_c0_seq2	XP_001333900.3	Predicted: uncharacterized protein si:ch211-288g17.4 [<i>Danio rerio</i>]
comp155676_c0_seq1	NP_001138258.1	Si:ch1073-126c3.2 precursor [<i>Danio rerio</i>]
comp156219_c0_seq1	XP_009290112.1	Predicted: osteocrin [<i>Danio rerio</i>]
comp156230_c0_seq1	XP_790627.1	Predicted: uncharacterized protein loc585718 [<i>Strongylocentrotus purpuratus</i>]
comp156867_c0_seq1	NP_001122014.1	Rna-binding protein 44 [<i>Danio rerio</i>]
comp157001_c1_seq1	XP_003199627.1	Predicted: uncharacterized protein loc100537720 [<i>Danio rerio</i>]
comp157403_c0_seq1	KFM72942.1	Hypothetical protein x975_17423, partial [<i>Stegodyphus mimosarum</i>]
comp157781_c0_seq1	XP_009299536.1	Predicted: serine/arginine repetitive matrix protein 4 isoform x1 [<i>Danio rerio</i>]

comp157898_c0_seq1	XP_009291424.1	Predicted: profilin-4 isoform x3 [<i>Danio rerio</i>]
comp158018_c0_seq2	CDQ90839.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp158373_c0_seq1	XP_007907337.1	Predicted: se-cephalotoxin-like [<i>Callorhinchus milii</i>]
comp158731_c0_seq1	BAD69701.1	Olfactory receptor, partial [<i>Cyprinus carpio</i>]
comp158731_c2_seq1	BAA94423.1	Olfactory receptor 47e [<i>Cyprinus carpio</i>]
comp159020_c0_seq1	XP_005461576.1	Predicted: uncharacterized protein loc102079413 [<i>Oreochromis niloticus</i>]
comp159050_c0_seq1	KFM79739.1	Hypothetical protein x975_14832, partial [<i>Stegodyphus mimosarum</i>]
comp159106_c0_seq1	YP_009058796.1	Nadh dehydrogenase subunit 6 (mitochondrion) [<i>Selenops bursarius</i>]
comp159384_c0_seq1	NP_956141.1	Fxyd domain-containing ion transport regulator 6 precursor [<i>Danio rerio</i>]
comp159397_c0_seq2	XP_003766070.1	Predicted: amyotrophic lateral sclerosis 2 chromosomal region candidate gene 11 protein, partial [<i>Sarcophilus harrisi</i>]
comp159453_c0_seq2	XP_009298974.1	Predicted: uncharacterized protein loc103910731 [<i>Danio rerio</i>]
comp159803_c0_seq1	ADV40335.1	Hypothetical protein, partial [<i>Latrodectus hesperus</i>]
comp160173_c0_seq1	KFM58526.1	Tenascin-x, partial [<i>Stegodyphus mimosarum</i>]
comp160312_c1_seq1	XP_009291153.1	Predicted: uncharacterized protein mia2 [<i>Danio rerio</i>]
comp160678_c1_seq1	CDQ97310.1	Unnamed protein product, partial [<i>Oncorhynchus mykiss</i>]
comp161058_c0_seq2	NP_001116715.1	Upf0691 protein c9orf116 homolog [<i>Danio rerio</i>]
comp161211_c0_seq1	NP_001070139.1	Uncharacterized protein c5orf34 homolog [<i>Danio rerio</i>]
comp161622_c1_seq1	XP_009060757.1	Hypothetical protein lotgidraft_165496 [<i>Lottia gigantea</i>]
comp161622_c1_seq2	XP_004912395.1	Predicted: uncharacterized protein loc100485380 [<i>Xenopus (silurana) tropicalis</i>]
comp161913_c0_seq1	XP_002644180.1	Hypothetical protein cbg17156 [<i>Caenorhabditis briggsae</i>]
comp162301_c0_seq2	ACO14214.1	Hras-like suppressor 2 [<i>Esox lucius</i>]
comp162301_c0_seq3	XP_005953223.1	Predicted: uncharacterized protein loc102300622 [<i>Haplochromis burtoni</i>]
comp162447_c0_seq1	XP_001340192.2	Predicted: coiled-coil domain-containing protein 110 [<i>Danio rerio</i>]
comp162669_c1_seq2	XP_003201803.1	Predicted: neuromedin-s isoform x2 [<i>Danio rerio</i>]
comp162880_c0_seq1	XP_002663431.2	Predicted: uncharacterized protein si:ch73-7i4.2 [<i>Danio rerio</i>]
comp163061_c0_seq2	XP_009303042.1	Predicted: uncharacterized protein si:dkey-29d5.1 isoform x2 [<i>Danio rerio</i>]
comp163738_c0_seq1	XP_004337198.1	Serineaspartate repeat-containing protein i, putative [<i>Acanthamoeba castellanii str. Neff</i>]
comp163871_c0_seq2	KFM82860.1	Hypothetical protein x975_12899, partial [<i>Stegodyphus mimosarum</i>]

comp165390_c0_seq1	XP_009298127.1	Predicted: uncharacterized protein loc103910348 [<i>Danio rerio</i>]
comp167478_c0_seq1	NP_922919.1	Siaz-interacting nuclear protein [<i>Danio rerio</i>]
comp167478_c0_seq2	AAI15160.1	Zgc:136513 protein [<i>Danio rerio</i>]
comp167535_c0_seq2	XP_009299053.1	Predicted: uncharacterized protein loc101883270 [<i>Danio rerio</i>]
comp167680_c0_seq1	KFM79856.1	Hemocyanin a chain, partial [<i>Stegodyphus mimosarum</i>]
comp167718_c0_seq2	XP_003200724.1	Predicted: barttin [<i>Danio rerio</i>]
comp168209_c0_seq1	NP_001007761.1	Pdk1-interacting protein 1 precursor [<i>Danio rerio</i>]
comp168433_c0_seq1	NP_001129453.1	Uncharacterized protein loc795037 [<i>Danio rerio</i>]
comp168595_c0_seq1	NP_878302.1	Mesogenin-1 [<i>Danio rerio</i>]
comp168937_c0_seq1	XP_002894773.1	Conserved hypothetical protein [<i>Phytophthora infestans t30-4</i>]
comp169321_c1_seq1	AAX26005.2	Sjchgc07815 protein, partial [<i>Schistosoma japonicum</i>]
comp169557_c0_seq1	NP_001002568.1	Vitamin d-binding protein precursor [<i>Danio rerio</i>]
comp169625_c0_seq2	NP_001108053.1	Uncharacterized protein loc100136863 [<i>Danio rerio</i>]
comp169765_c0_seq2	XP_009302111.1	Predicted: collagen alpha-1(xxvii) chain-like [<i>Danio rerio</i>]
comp170028_c2_seq1	XP_003724272.1	Predicted: uncharacterized protein loc100890760 [<i>Strongylocentrotus purpuratus</i>]
comp170079_c0_seq1	CDQ98453.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp170079_c0_seq2	XP_004548999.1	Predicted: uncharacterized protein loc101487821 [<i>Maylandia zebra</i>]
comp170084_c0_seq1	XP_009302539.1	Predicted: uncharacterized protein loc101884100 [<i>Danio rerio</i>]
comp170430_c0_seq2	XP_002667158.2	Predicted: toll/interleukin-1 receptor domain-containing adapter protein [<i>Danio rerio</i>]
comp170537_c0_seq3	XP_001846843.1	Polyphosphate kinase 2 family [<i>Culex quinquefasciatus</i>]
comp171143_c0_seq1	XP_009291927.1	Predicted: uncharacterized protein loc100330497 [<i>Danio rerio</i>]
comp172132_c0_seq1	KFM61752.1	Hypothetical protein x975_08079, partial [<i>Stegodyphus mimosarum</i>]
comp172165_c1_seq1	XP_004209292.1	Predicted: uncharacterized protein loc101237568, partial [<i>Hydra vulgaris</i>]
comp172639_c0_seq3	XP_006801670.1	Predicted: olfactory receptor 56a4-like [<i>Neolamprologus brichardi</i>]
comp173457_c0_seq1	XP_009292086.1	Predicted: uncharacterized protein loc101886500 isoform x1 [<i>Danio rerio</i>]
comp173852_c0_seq2	CAJ98867.1	Putative interferon gamma 1 [<i>Cyprinus carpio</i>]
comp173881_c1_seq2	BAM11254.1	Preproneuromedin u isoform 3 [<i>Cyprinus carpio</i>]
comp173943_c1_seq2	XP_007230779.1	Predicted: uncharacterized protein c14orf37 homolog isoform x3 [<i>Astyanax mexicanus</i>]
comp173979_c0_seq2	XP_008319875.1	Predicted: uncharacterized protein loc103387133 [<i>Cynoglossus semilaevis</i>]

comp174151_c0_seq1	NP_001289983.1	Small integral membrane protein 18 [<i>Danio rerio</i>]
comp174165_c0_seq1	XP_009295050.1	Predicted: uncharacterized protein si:rp71-17i16.6 isoform x1 [<i>Danio rerio</i>]
comp174298_c0_seq2	KFM65863.1	Hypothetical protein x975_14780, partial [<i>Stegodyphus mimosarum</i>]
comp174395_c0_seq1	CDS10931.1	Hypothetical protein Iramosa1 1417 [<i>Absidia idahoensis</i> var. <i>Thermophila</i>]
comp174448_c1_seq1	XP_008397128.1	Predicted: low quality protein: uncharacterized protein loc103458245 [<i>Poecilia reticulata</i>]
comp174700_c0_seq2	NP_001104703.1	Uncharacterized protein loc100000337 [<i>Danio rerio</i>]
comp174913_c0_seq2	XP_691039.2	Predicted: uncharacterized protein c20orf96 homolog [<i>Danio rerio</i>]
comp175269_c0_seq3	XP_009681285.1	Predicted: serum response factor-binding protein 1-like, partial [<i>Struthio camelus australis</i>]
comp175320_c0_seq1	XP_008261898.1	Predicted: serine/arginine repetitive matrix protein 3-like [<i>Oryctolagus cuniculus</i>]
comp176061_c0_seq2	XP_001341090.2	Predicted: uncharacterized protein c1orf158 homolog isoform x1 [<i>Danio rerio</i>]
comp176124_c0_seq1	XP_005157555.1	Predicted: uncharacterized protein loc100037332 isoform x1 [<i>Danio rerio</i>]
comp176502_c0_seq1	WP_013902671.1	4-diphosphocytidyl-2c-methyl-d-erythritol kinase [<i>Ramlibacter tataouinensis</i>]
comp176505_c0_seq1	KFM71672.1	Hypothetical protein x975_08483, partial [<i>Stegodyphus mimosarum</i>]
comp176625_c0_seq1	XP_007568082.1	Predicted: leucine-rich repeat extensin-like protein 2 [<i>Poecilia formosa</i>]
comp177550_c0_seq1	AAI53525.1	Zgc:153540 protein [<i>Danio rerio</i>]
comp177712_c0_seq1	NP_001289705.1	Lens epithelial protein [<i>Danio rerio</i>]
comp177805_c0_seq2	AAI63018.1	Indoleamine 2,3-dioxygenase 1 [<i>Danio rerio</i>]
comp177886_c0_seq2	NP_001093523.1	Transmembrane protein 232 [<i>Danio rerio</i>]
comp17798_c0_seq1	XP_007250082.1	Predicted: uncharacterized protein loc103041463 [<i>Astyanax mexicanus</i>]
comp178038_c1_seq1	NP_001073426.1	Uncharacterized protein loc555281 precursor [<i>Danio rerio</i>]
comp178144_c0_seq1	KFM71221.1	Hypothetical protein x975_14061, partial [<i>Stegodyphus mimosarum</i>]
comp178912_c0_seq3	BAJ41039.1	Membrane form immunoglobulin heavy chain z2, partial [<i>Cyprinus carpio</i>]
comp179171_c0_seq1	NP_001128617.1	Uncharacterized protein loc100189619 [<i>Danio rerio</i>]
comp179866_c0_seq1	ADV40330.1	Hypothetical protein, partial [<i>Latrodectus hesperus</i>]
comp180489_c0_seq1	AFM08413.1	Kisspeptin1, partial [<i>Cyprinus carpio</i>]
comp180489_c0_seq2	ACK77790.1	Kiss-1a [<i>Carassius auratus</i>]
comp18076_c0_seq1	CDQ86214.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp181242_c0_seq1	NP_001138711.1	Secretory calcium-binding phosphoprotein 7 precursor [<i>Danio rerio</i>]
comp181427_c0_seq1	AFI92846.1	14 kda apolipoprotein [<i>Cyprinus carpio</i>]

comp181468_c0_seq1	XP_700422.2	Predicted: protein phosphatase 1 regulatory subunit 32 isoform x1 [<i>Danio rerio</i>]
comp181956_c0_seq1	XP_705262.1	Predicted: uncharacterized protein si:ch211-214j24.14 [<i>Danio rerio</i>]
comp182167_c0_seq1	ABD51465.1	S11 [<i>Squirrelpox virus</i>]
comp182178_c0_seq1	XP_005162916.1	Predicted: protein phosphatase 1 regulatory subunit 17 [<i>Danio rerio</i>]
comp182195_c0_seq1	KGS79403.1	Hypothetical protein x942_4100 [<i>Burkholderia pseudomallei mshr5596</i>]
comp182988_c0_seq1	XP_009299438.1	Predicted: uncharacterized protein pb18e9.04c-like isoform x2 [<i>Danio rerio</i>]
comp183008_c0_seq1	CDQ80337.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp183149_c0_seq3	XP_006629387.1	Predicted: single-pass membrane and coiled-coil domain-containing protein 3-like [<i>Lepisosteus oculatus</i>]
comp183229_c0_seq1	BAL43179.1	Interleukin 4/13b [<i>Cyprinus carpio</i>]
comp184094_c0_seq1	XP_007577509.1	Predicted: uncharacterized protein loc103155683 [<i>Poecilia formosa</i>]
comp184564_c0_seq2	KFM68763.1	Hypothetical protein x975_13339, partial [<i>Stegodyphus mimosarum</i>]
comp184639_c0_seq3	CDQ94798.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp184834_c0_seq2	XP_009295796.1	Predicted: uncharacterized protein loc103909761 isoform x1 [<i>Danio rerio</i>]
comp184834_c0_seq3	XP_009295798.1	Predicted: uncharacterized protein loc100537243 isoform x1 [<i>Danio rerio</i>]
comp184954_c3_seq1	XP_003198771.1	Predicted: uncharacterized protein c3orf30 homolog [<i>Danio rerio</i>]
comp185062_c0_seq2	NP_001120804.1	Secreted immunoglobulin domain 1 precursor [<i>Danio rerio</i>]
comp185191_c0_seq3	XP_009296330.1	Predicted: uncharacterized protein loc103909857 [<i>Danio rerio</i>]
comp185513_c0_seq1	ADV40120.1	Methyltransferase-like protein, partial [<i>Latrodectus hesperus</i>]
comp185696_c0_seq1	XP_005172950.1	Predicted: uncharacterized protein loc101882091 [<i>Danio rerio</i>]
comp185704_c0_seq3	XP_001619855.1	Hypothetical protein nemvedraft_v1g150057, partial [<i>Nematostella vectensis</i>]
comp185704_c0_seq7	KFH67136.1	Hypothetical protein mveg_07659 [<i>Mortierella verticillata nr1 6337</i>]
comp185724_c0_seq1	ELU14242.1	Hypothetical protein captedraft_191711 [<i>Capitella teleta</i>]
comp186118_c0_seq2	AEG74406.1	Nanog, partial [<i>Carassius auratus</i>]
comp186299_c0_seq1	XP_002665930.1	Predicted: ghrelin [<i>Danio rerio</i>]
comp186311_c1_seq1	XP_004917710.1	Predicted: uncharacterized protein loc101730916 [<i>Xenopus (silurana) tropicalis</i>]
comp186519_c0_seq2	XP_003727901.1	Predicted: uncharacterized protein loc100889156 [<i>Strongylocentrotus purpuratus</i>]
comp18675_c0_seq1	ADV40139.1	Hemocyanin subunit g, partial [<i>Latrodectus hesperus</i>]
comp186984_c0_seq1	NP_001013484.1	Protein fam221a [<i>Danio rerio</i>]

comp187013_c0_seq1	P25421.2	Protachykinin [<i>Carassius auratus</i>]
comp187109_c0_seq1	XP_009291256.1	Predicted: uncharacterized protein loc100001321 isoform x1 [<i>Danio rerio</i>]
comp187216_c0_seq2	NP_001104553.1	Interferon phi 3 precursor [<i>Danio rerio</i>]
comp187417_c0_seq1	XP_005451862.1	Predicted: uncharacterized protein loc102076121 [<i>Oreochromis niloticus</i>]
comp187438_c0_seq2	XP_007230085.1	Predicted: complexin-4-like [<i>Astyanax mexicanus</i>]
comp187712_c0_seq1	XP_005170214.2	Predicted: uncharacterized protein loc101883030 [<i>Danio rerio</i>]
comp188196_c0_seq1	XP_002586683.1	Hypothetical protein brafldraft_105490 [<i>Branchiostoma floridae</i>]
comp188301_c0_seq1	KFM60587.1	U24-ctenitoxin-pn1a, partial [<i>Stegodyphus mimosarum</i>]
comp188430_c0_seq1	KFM78498.1	Hypothetical protein x975_19975, partial [<i>Stegodyphus mimosarum</i>]
comp188605_c0_seq1	ADV40307.1	Hypothetical protein, partial [<i>Latrodectus hesperus</i>]
comp188908_c0_seq1	NP_001002332.1	Uricase [<i>Danio rerio</i>]
comp189119_c0_seq2	XP_009292740.1	Predicted: protein phosphatase 1 regulatory subunit 36 isoform x2 [<i>Danio rerio</i>]
comp189202_c0_seq1	XP_009294987.1	Predicted: uncharacterized protein si:ch211-208g1.1 [<i>Danio rerio</i>]
comp189204_c0_seq5	BAM09181.1	Interferon gamma [<i>Carassius auratus langsdorfii</i>]
comp189377_c0_seq3	XP_002662049.3	Predicted: uncharacterized protein si:dkey-229d2.7 [<i>Danio rerio</i>]
comp189379_c1_seq1	XP_005173936.1	Predicted: proline-rich protein 7 isoform x2 [<i>Danio rerio</i>]
comp189391_c0_seq1	XP_009298825.1	Predicted: uncharacterized protein loc101884446 [<i>Danio rerio</i>]
comp189412_c1_seq1	XP_007544728.1	Predicted: uncharacterized protein loc103132897 [<i>Poecilia formosa</i>]
comp189440_c0_seq3	AFU81784.1	Hypothetical protein [<i>Ctenopharyngodon idella</i>]
comp189450_c0_seq1	AAH96937.1	Zgc:174622 protein, partial [<i>Danio rerio</i>]
comp189477_c0_seq1	KFM72188.1	Hypothetical protein x975_13453, partial [<i>Stegodyphus mimosarum</i>]
comp189687_c0_seq1	XP_002942726.1	Predicted: uncharacterized protein loc100495238 [<i>Xenopus (silurana) tropicalis</i>]
comp190075_c0_seq2	XP_005173816.1	Predicted: uncharacterized protein wu:fb12c09 [<i>Danio rerio</i>]
comp190357_c0_seq3	CDQ66896.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp190443_c0_seq1	WP_022124091.1	Type i restriction enzyme [<i>Clostridium sp. Cag:510</i>]
comp190468_c0_seq1	XP_002660836.2	Predicted: uncharacterized protein loc100331996 [<i>Danio rerio</i>]
comp190617_c0_seq1	EMC82497.1	Hypothetical protein a306_09513, partial [<i>Columba livia</i>]
comp190628_c0_seq1	XP_009301430.1	Predicted: mucin-3a [<i>Danio rerio</i>]
comp191179_c0_seq1	NP_001289677.1	Uncharacterized protein loc100535653 precursor [<i>Danio rerio</i>]

comp191179_c0_seq3	NP_001139884.1	Cq067 protein precursor [<i>Salmo salar</i>]
comp191181_c0_seq1	XP_005172061.1	Predicted: uroplakin 2-like isoform x1 [<i>Danio rerio</i>]
comp191193_c0_seq1	NP_001138712.1	Secretory calcium-binding phosphoprotein 1 precursor [<i>Danio rerio</i>]
comp191297_c0_seq2	XP_005174744.2	Predicted: uncharacterized protein loc101885332 [<i>Danio rerio</i>]
comp191335_c0_seq3	XP_003200228.2	Predicted: uncharacterized protein loc100537037 [<i>Danio rerio</i>]
comp191397_c0_seq3	XP_009304491.1	Predicted: uncharacterized protein loc103911857 isoform x1 [<i>Danio rerio</i>]
comp191785_c2_seq2	XP_005466225.1	Predicted: uncharacterized protein loc102081204 [<i>Oreochromis niloticus</i>]
comp191792_c0_seq1	XP_691110.1	Predicted: apolipoprotein c-ii [<i>Danio rerio</i>]
comp191858_c0_seq2	BAJ09457.1	Gsp-37 [<i>Carassius auratus</i>]
comp191913_c0_seq4	XP_005164165.1	Predicted: phosphoinositide-interacting protein [<i>Danio rerio</i>]
comp192115_c0_seq1	XP_005170692.1	Predicted: uncharacterized protein loc101883546 [<i>Danio rerio</i>]
comp192196_c0_seq3	XP_008654099.1	Predicted: uncharacterized protein loc103634322 [<i>Zea mays</i>]
comp192259_c0_seq1	EFZ10292.1	Hypothetical protein sinv_11569, partial [<i>Solenopsis invicta</i>]
comp192259_c0_seq2	EFZ15394.1	Hypothetical protein sinv_05588, partial [<i>Solenopsis invicta</i>]
comp192259_c0_seq3	XP_002404583.1	Hypothetical protein iscw_iscw008006 [<i>Ixodes scapularis</i>]
comp192454_c0_seq2	XP_001338145.2	Predicted: uncharacterized protein wu:fk35f04 [<i>Danio rerio</i>]
comp192634_c0_seq4	XP_001919813.3	Predicted: coiled-coil domain-containing protein 78 isoform x1 [<i>Danio rerio</i>]
comp192692_c0_seq2	NP_001138708.1	Secretory calcium-binding phosphoprotein 5 precursor [<i>Danio rerio</i>]
comp192728_c0_seq2	XP_005169245.1	Predicted: uncharacterized protein loc101886660 [<i>Danio rerio</i>]
comp192799_c1_seq1	NP_001038282.2	Coiled-coil domain-containing protein 87 [<i>Danio rerio</i>]
comp192954_c0_seq2	NP_001268717.1	A-kinase anchor protein 12 isoform 1 [<i>Danio rerio</i>]
comp192992_c0_seq1	AHJ79064.1	Apolipoprotein A-I [<i>Cyprinus carpio</i>]
comp192992_c0_seq2	AII80532.1	Apolipoprotein AIB1 [<i>Cyprinus carpio</i>]
comp192999_c0_seq3	NP_001004525.1	Uncharacterized protein loc368902 precursor [<i>Danio rerio</i>]
comp193109_c0_seq2	NP_956705.1	Protein slx4ip [<i>Danio rerio</i>]
comp193116_c0_seq1	CDQ87646.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp193133_c0_seq1	AAF82387.1	Af111028_1 gastrin-releasing peptide precursor [<i>Carassius auratus</i>]
comp193424_c2_seq6	CDY06100.1	Bnaa09g36800d [<i>Brassica napus</i>]
comp193428_c0_seq1	CDQ64680.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]

comp193556_c0_seq2	NP_991149.1	Prion protein a precursor [<i>Danio rerio</i>]
comp193745_c1_seq2	XP_007245273.1	Predicted: nuclear gtpase slip-gc-like isoform x1 [<i>Astyanax mexicanus</i>]
comp193818_c0_seq3	AAN71721.1	Reverse transcriptase/ribonuclease h, partial [<i>Danio rerio</i>]
comp193928_c0_seq1	NP_001289712.1	Uncharacterized protein loc100535414 [<i>Danio rerio</i>]
comp193992_c0_seq1	AIT51847.1	Bone gla protein [<i>Cyprinus carpio 'color'</i>]
comp194096_c0_seq1	NP_001018607.1	Interleukin-1 receptor-associated kinase 1-binding protein 1 homolog [<i>Danio rerio</i>]
comp194139_c0_seq1	XP_009295493.1	Predicted: lymphocyte antigen 86 isoform x2 [<i>Danio rerio</i>]
comp194193_c0_seq1	XP_009299482.1	Predicted: uncharacterized protein si:ch211-12e13.1 [<i>Danio rerio</i>]
comp194216_c0_seq1	CAD35749.1	Putative corticotropin releasing hormone binding protein 2 [<i>Cyprinus carpio</i>]
comp194221_c0_seq5	XP_002666714.1	Predicted: uncharacterized protein ccl34b.8 [<i>Danio rerio</i>]
comp194439_c1_seq10	XP_002595816.1	Hypothetical protein brafldraft_96790 [<i>Branchiostoma floridae</i>]
comp194534_c0_seq2	XP_010106798.1	Hypothetical protein 1484_009238 [<i>Morus notabilis</i>]
comp194665_c0_seq2	XP_009304656.1	Predicted: tumor necrosis factor receptor superfamily member 13b [<i>Danio rerio</i>]
comp194687_c0_seq1	XP_007228813.1	Predicted: coiled-coil domain-containing protein 177-like [<i>Astyanax mexicanus</i>]
comp194798_c0_seq1	XP_009300357.1	Predicted: kynureninase [<i>Danio rerio</i>]
comp194800_c2_seq1	XP_005160431.2	Predicted: uncharacterized protein loc101882067 [<i>Danio rerio</i>]
comp194813_c0_seq1	KFM74144.1	Hypothetical protein x975_07331, partial [<i>Stegodyphus mimosarum</i>]
comp194840_c0_seq2	NP_001121769.1	Uncharacterized protein loc100006693 [<i>Danio rerio</i>]
comp194924_c0_seq1	XP_693940.1	Predicted: coiled-coil domain-containing protein 105 [<i>Danio rerio</i>]
comp194998_c0_seq10	AGO65037.1	Agouti signaling protein [<i>Cyprinus carpio 'color'</i>]
comp195293_c0_seq1	EJY57353.1	Ael017334-pa [<i>Aedes aegypti</i>]
comp195312_c0_seq1	XP_009291080.1	Predicted: transmembrane protein flj37396 isoform x1 [<i>Danio rerio</i>]
comp195481_c1_seq1	XP_009294834.1	Predicted: pentafunctional arom polypeptide-like isoform x1 [<i>Danio rerio</i>]
comp195532_c0_seq1	XP_005462325.1	Predicted: uncharacterized protein loc102075860 [<i>Oreochromis niloticus</i>]
comp195902_c0_seq2	XP_683798.3	Predicted: dc-stamp domain-containing protein 1 isoform x1 [<i>Danio rerio</i>]
comp195981_c1_seq10	XP_699953.2	Predicted: uncharacterized protein loc571286 [<i>Danio rerio</i>]
comp195981_c1_seq6	XP_009294326.1	Predicted: uncharacterized protein loc103909539 isoform x3 [<i>Danio rerio</i>]
comp196127_c0_seq1	XP_004917708.1	Predicted: uncharacterized protein loc101730796 isoform x1 [<i>Xenopus (silurana) tropicalis</i>]
comp196200_c2_seq2	XP_008558869.1	Predicted: uncharacterized protein loc103579287, partial [<i>Microplitis demolitor</i>]

comp196249_c0_seq2	AEO18266.1	Hypothetical protein [<i>Pseudocowpox virus</i>]
comp196529_c0_seq3	NP_001177311.1	Selenoprotein I [<i>Danio rerio</i>]
comp196623_c1_seq3	AHK80892.1	Il-12 p35 [<i>Ctenopharyngodon idella</i>]
comp196747_c2_seq5	NP_001116090.1	Uncharacterized protein loc100142641 precursor [<i>Danio rerio</i>]
comp196805_c0_seq1	XP_009299519.1	Predicted: uncharacterized protein loc100334678 [<i>Danio rerio</i>]
comp196844_c0_seq1	ADF97613.1	D-type binding-protein 1, partial [<i>Hypophthalmichthys molitrix</i>]
comp196844_c0_seq5	AAH93146.1	Zgc:162082 protein, partial [<i>Danio rerio</i>]
comp196869_c3_seq4	AAI53391.1	Si:ch211-219i10.1 protein, partial [<i>Danio rerio</i>]
comp196944_c0_seq2	XP_005158083.2	Predicted: uncharacterized protein loc101882791 [<i>Danio rerio</i>]
comp196944_c0_seq3	CAK04174.1	Novel immune type receptor protein, partial [<i>Danio rerio</i>]
comp196976_c0_seq1	NP_001076366.1	Potassium voltage-gated channel subfamily e member 4 [<i>Danio rerio</i>]
comp197289_c2_seq4	ELU15299.1	Hypothetical protein captedraft_193779 [<i>Capitella teleta</i>]
comp197404_c4_seq1	XP_002664363.2	Predicted: uncharacterized protein si:dkey-92i17.2 [<i>Danio rerio</i>]
comp197410_c0_seq3	KFV70020.1	Foot protein 1 variant 1, partial [<i>Picoides pubescens</i>]
comp197552_c0_seq3	XP_002168906.2	Predicted: uncharacterized protein loc100199418 [<i>Hydra vulgaris</i>]
comp197576_c0_seq1	XP_009303756.1	Predicted: uncharacterized protein loc103911757 [<i>Danio rerio</i>]
comp197706_c0_seq13	XP_005164624.2	Predicted: uncharacterized protein loc101885044 [<i>Danio rerio</i>]
comp197706_c0_seq14	XP_005172218.2	Predicted: uncharacterized protein loc101884836 [<i>Danio rerio</i>]
comp197712_c1_seq2	XP_001335256.1	Predicted: uncharacterized protein si:dkey-30j10.5 [<i>Danio rerio</i>]
comp197728_c0_seq11	NP_998612.1	E3 ubiquitin-protein ligase tm129 [<i>Danio rerio</i>]
comp197728_c0_seq5	XP_001520916.2	Predicted: transmembrane protein 129 [<i>Ornithorhynchus anatinus</i>]
comp197755_c3_seq15	XP_009289321.1	Predicted: uncharacterized protein loc100334705 isoform x2 [<i>Danio rerio</i>]
comp197797_c0_seq3	XP_001344057.1	Predicted: natural cytotoxicity triggering receptor 3 [<i>Danio rerio</i>]
comp197806_c0_seq1	XP_001919860.2	Predicted: uncharacterized protein si:dkey-27h10.2 isoform x1 [<i>Danio rerio</i>]
comp197816_c2_seq4	WP_000022428.1	Lipopolysaccharide 1,2-glucosyltransferase [<i>Helicobacter pylori</i>]
comp197862_c1_seq1	XP_009302597.1	Predicted: uncharacterized protein loc101882297 [<i>Danio rerio</i>]
comp197902_c0_seq4	XP_003198166.1	Predicted: equinatoxin-4-like [<i>Danio rerio</i>]
comp197922_c0_seq1	XP_003199077.1	Predicted: purkinje cell protein 4-like protein 1 [<i>Danio rerio</i>]
comp197925_c0_seq4	XP_004067342.1	Predicted: uncharacterized protein loc101169027 [<i>Oryzias latipes</i>]

comp197988_c0_seq10	NP_001139028.1	Si:dkey-22i16.3 precursor [<i>Danio rerio</i>]
comp198048_c0_seq3	EFN76852.1	Hypothetical protein eai_16499, partial [<i>Harpegnathos saltator</i>]
comp198063_c0_seq20	XP_009300031.1	Predicted: uncharacterized protein loc103911110 [<i>Danio rerio</i>]
comp198122_c4_seq12	AAH45517.2	Zgc:56719 protein, partial [<i>Danio rerio</i>]
comp198228_c0_seq1	XP_003722449.1	Proteophosphoglycan 5 [<i>Leishmania major strain friedlin</i>]
comp198253_c0_seq5	XP_002662199.3	Predicted: uncharacterized protein loc100334140 [<i>Danio rerio</i>]
comp198255_c0_seq4	XP_005164954.1	Predicted: n-acyl-phosphatidylethanolamine-hydrolyzing phospholipase d [<i>Danio rerio</i>]
comp198265_c0_seq13	XP_009294473.1	Predicted: uncharacterized protein loc103909563 [<i>Danio rerio</i>]
comp198383_c0_seq2	NP_998047.1	Cysteine and tyrosine-rich protein 1 precursor [<i>Danio rerio</i>]
comp198411_c0_seq1	XP_009017602.1	Hypothetical protein helrodraft_172694 [<i>Helobdella robusta</i>]
comp198496_c0_seq4	XP_009296335.1	Predicted: leucine-rich repeat-containing protein 56 isoform x1 [<i>Danio rerio</i>]
comp198587_c0_seq2	XP_001336486.1	Predicted: uncharacterized protein loc100000098 isoform x1 [<i>Danio rerio</i>]
comp198810_c1_seq3	XP_009294180.1	Predicted: uncharacterized protein loc103909516 [<i>Danio rerio</i>]
comp198882_c0_seq2	ACL97385.1	Gag-pol polyprotein [<i>Medicago truncatula</i>]
comp198903_c5_seq1	AGV39404.1	Neurokinin b [<i>Carassius auratus</i>]
comp198964_c1_seq1	XP_009293660.1	Predicted: uncharacterized protein loc103909452 [<i>Danio rerio</i>]
comp199065_c0_seq5	AAI52493.1	Zgc:165381 protein [<i>Danio rerio</i>]
comp199158_c0_seq6	NP_001268390.1	Valacyclovir hydrolase [<i>Danio rerio</i>]
comp199158_c0_seq7	AAH45899.1	Loc402804 protein, partial [<i>Danio rerio</i>]
comp199235_c0_seq4	XP_002169937.2	Predicted: uncharacterized protein loc100198666 [<i>Hydra vulgaris</i>]
comp199235_c0_seq5	ELT87812.1	Hypothetical protein captedraft_212881 [<i>Capitella teleta</i>]
comp199320_c1_seq2	NP_001164414.1	Uncharacterized protein loc100327252 precursor [<i>Danio rerio</i>]
comp199364_c2_seq3	XP_009294784.1	Predicted: uncharacterized protein loc101884811 [<i>Danio rerio</i>]
comp199394_c1_seq5	XP_008280311.1	Predicted: protein shisa-5-like [<i>Stegastes partitus</i>]
comp199398_c0_seq1	EKC41203.1	Hypothetical protein cgi_10026796 [<i>Crassostrea gigas</i>]
comp199398_c0_seq8	EKC32747.1	Hypothetical protein cgi_10017156 [<i>Crassostrea gigas</i>]
comp199406_c0_seq2	XP_001333661.1	Predicted: igf-like family receptor 1 [<i>Danio rerio</i>]
comp199543_c0_seq4	XP_005165815.2	Predicted: immunoglobulin superfamily member 1-like [<i>Danio rerio</i>]
comp199620_c1_seq1	XP_009296323.1	Predicted: uncharacterized protein loc101882374 isoform x1 [<i>Danio rerio</i>]

comp199620_c1_seq2	XP_005160504.2	Predicted: uncharacterized protein loc101887144 [<i>Danio rerio</i>]
comp199659_c0_seq1	NP_001038347.2	Uncharacterized protein loc559078 [<i>Danio rerio</i>]
comp199782_c0_seq2	XP_005170607.1	Predicted: kininogen-1 isoform x1 [<i>Danio rerio</i>]
comp199854_c0_seq4	CAI61347.1	Tpa: interleukin-11b [<i>Danio rerio</i>]
comp200074_c2_seq3	XP_002662623.1	Predicted: uncharacterized protein si:dkey-125e8.3 [<i>Danio rerio</i>]
comp200079_c0_seq2	NP_001094418.2	Uncharacterized protein c8orf88 homolog [<i>Danio rerio</i>]
comp200079_c0_seq5	ABR24799.1	Granulito [<i>Danio rerio</i>]
comp200130_c1_seq7	XP_005159452.1	Predicted: uncharacterized protein loc100034474 isoform x2 [<i>Danio rerio</i>]
comp200248_c0_seq2	XP_009300583.1	Predicted: uncharacterized protein loc100330485 [<i>Danio rerio</i>]
comp200259_c0_seq2	XP_005162157.1	Predicted: uncharacterized protein at5g50100, mitochondrial-like isoform x1 [<i>Danio rerio</i>]
comp200276_c0_seq4	XP_005159212.1	Predicted: tetratricopeptide repeat protein 23 isoform x2 [<i>Danio rerio</i>]
comp200303_c0_seq4	XP_004070666.1	Predicted: uncharacterized protein loc101171081 [<i>Oryzias latipes</i>]
comp200304_c0_seq2	XP_002666923.1	Predicted: iq and ubiquitin-like domain-containing protein [<i>Danio rerio</i>]
comp200336_c0_seq2	NP_001082912.1	Transmembrane protein 119b precursor [<i>Danio rerio</i>]
comp200337_c1_seq15	NP_001083005.1	Centrosomal protein of 128 kda [<i>Danio rerio</i>]
comp200563_c0_seq5	XP_002666584.3	Predicted: uncharacterized protein si:rp71-80o10.4 [<i>Danio rerio</i>]
comp200705_c4_seq6	XP_007569453.1	Predicted: uncharacterized protein loc103149984 [<i>Poecilia formosa</i>]
comp200726_c3_seq14	ELT96182.1	Hypothetical protein captedraft_185978 [<i>Capitella teleta</i>]
comp200781_c0_seq4	XP_694540.3	Predicted: upf0764 protein c16orf89 homolog [<i>Danio rerio</i>]
comp200781_c0_seq5	AAI52291.1	Loc566173 protein, partial [<i>Danio rerio</i>]
comp200784_c0_seq2	XP_009024094.1	Hypothetical protein helrodraft_85188, partial [<i>Helobdella robusta</i>]
comp200786_c0_seq1	WP_022410511.1	Hypothetical protein [<i>Ruminococcus sp. Cag:330</i>]
comp200786_c0_seq10	EFX65385.1	Hypothetical protein dappudraft_117311 [<i>Daphnia pulex</i>]
comp200814_c3_seq2	XP_001334154.3	Predicted: uncharacterized protein loc796591 [<i>Danio rerio</i>]
comp200814_c3_seq3	XP_009305917.1	Predicted: uncharacterized protein loc103912048 [<i>Danio rerio</i>]
comp200936_c1_seq2	NP_001124081.1	Uncharacterized protein loc794982 precursor [<i>Danio rerio</i>]
comp200937_c0_seq1	XP_009055516.1	Hypothetical protein lotgidraft_153369 [<i>Lottia gigantea</i>]
comp200937_c0_seq4	XP_009055621.1	Hypothetical protein lotgidraft_153490 [<i>Lottia gigantea</i>]
comp200995_c0_seq1	XP_005466545.1	Predicted: gata zinc finger domain-containing protein 14-like, partial [<i>Oreochromis niloticus</i>]

comp201115_c1_seq1	XP_003443670.1	Predicted: zymogen granule membrane protein 16-like [<i>Oreochromis niloticus</i>]
comp201365_c0_seq1	XP_009305922.1	Predicted: uncharacterized protein loc103912050 [<i>Danio rerio</i>]
comp201365_c2_seq3	XP_009298457.1	Predicted: thymidine phosphorylase [<i>Danio rerio</i>]
comp201398_c1_seq1	XP_003723412.1	Predicted: uncharacterized protein loc100888373 [<i>Strongylocentrotus purpuratus</i>]
comp201451_c0_seq1	NP_001018331.1	Uncharacterized protein loc552925 [<i>Danio rerio</i>]
comp201484_c0_seq4	XP_009301078.1	Predicted: uncharacterized protein loc103911308 [<i>Danio rerio</i>]
comp201561_c0_seq6	XP_002161223.2	Predicted: uncharacterized protein loc100197965 [<i>Hydra vulgaris</i>]
comp201575_c0_seq2	XP_001345005.1	Predicted: serine-rich and transmembrane domain-containing protein 1 [<i>Danio rerio</i>]
comp201601_c0_seq2	XP_009289654.1	Predicted: myosin heavy chain, clone 203 [<i>Danio rerio</i>]
comp201782_c0_seq2	ELT93329.1	Hypothetical protein captdraft_202674 [<i>Capitella teleta</i>]
comp201919_c0_seq5	XP_689447.1	Predicted: uncharacterized protein c2orf42 homolog isoform x2 [<i>Danio rerio</i>]
comp202029_c0_seq2	ELR50663.1	Hypothetical protein m91_01402, partial [<i>Bos mutus</i>]
comp202179_c0_seq2	NP_001028925.1	Uncharacterized protein loc619272 [<i>Danio rerio</i>]
comp202423_c0_seq2	XP_007573868.1	Predicted: protein gapt-like [<i>Poecilia formosa</i>]
comp202444_c0_seq12	XP_009290554.1	Predicted: uncharacterized protein loc100536667 [<i>Danio rerio</i>]
comp202613_c0_seq7	XP_001969987.1	Gg25361, isoform b [<i>Drosophila erecta</i>]
comp202686_c0_seq4	XP_009303764.1	Predicted: uncharacterized protein loc101884045 isoform x2 [<i>Danio rerio</i>]
comp202801_c0_seq5	XP_009289802.1	Predicted: uncharacterized protein loc100002197 isoform x3 [<i>Danio rerio</i>]
comp202852_c0_seq9	BAA36700.1	Serum amyloid a protein [<i>Cyprinus carpio</i>]
comp202875_c1_seq4	XP_003198998.1	Predicted: neurosecretory protein vgf [<i>Danio rerio</i>]
comp202876_c0_seq2	XP_005166055.1	Predicted: fibronectin type iii domain-containing protein 9 [<i>Danio rerio</i>]
comp202901_c0_seq1	XP_002612248.1	Hypothetical protein brafldraft_100078 [<i>Branchiostoma floridae</i>]
comp202901_c0_seq10	XP_002602442.1	Hypothetical protein brafldraft_117021 [<i>Branchiostoma floridae</i>]
comp202901_c0_seq9	XP_009858063.1	Predicted: uncharacterized protein loc104265534 [<i>Ciona intestinalis</i>]
comp202966_c0_seq3	XP_009300581.1	Predicted: uncharacterized protein loc101883268 [<i>Danio rerio</i>]
comp202981_c3_seq8	XP_005160448.1	Predicted: uncharacterized protein loc101884661 [<i>Danio rerio</i>]
comp203019_c0_seq1	NP_001076517.1	Uncharacterized protein loc100034401 [<i>Danio rerio</i>]
comp203024_c0_seq5	XP_006818369.1	Predicted: flocculation protein flo1-like [<i>Saccoglossus kowalevskii</i>]
comp203076_c0_seq5	XP_005172178.1	Predicted: uncharacterized protein loc101883672 [<i>Danio rerio</i>]

comp203283_c0_seq1	ELT99683.1	Hypothetical protein captedraft_216329, partial [<i>Capitella teleta</i>]
comp203357_c0_seq3	AAI55802.1	C14orf159 protein, partial [<i>Danio rerio</i>]
comp203357_c0_seq4	XP_001923890.1	Predicted: upf0317 protein c14orf159 homolog, mitochondrial [<i>Danio rerio</i>]
comp203444_c0_seq1	AAI71722.1	Zgc:165600 protein [<i>Danio rerio</i>]
comp203444_c0_seq4	XP_007232909.1	Predicted: lysosome-associated membrane glycoprotein 5 [<i>Astyanax mexicanus</i>]
comp203464_c2_seq1	AGO58864.1	Immunoglobulin v-set domain, partial [<i>Carassius auratus</i>]
comp203486_c0_seq2	NP_001070103.2	Putative transferase caf17 homolog, mitochondrial [<i>Danio rerio</i>]
comp203648_c0_seq2	XP_005161441.1	Predicted: uncharacterized protein si:dkey-29d5.2 [<i>Danio rerio</i>]
comp203650_c0_seq11	XP_009296868.1	Predicted: uncharacterized protein si:ch211-106h4.12 isoform x1 [<i>Danio rerio</i>]
comp203699_c0_seq8	XP_005464488.1	Predicted: uncharacterized protein loc100712513 isoform x2 [<i>Oreochromis niloticus</i>]
comp203783_c2_seq10	XP_006810778.1	Predicted: uncharacterized protein loc102775714, partial [<i>Neolamprologus brichardi</i>]
comp203863_c3_seq3	AAI07829.1	Wu:fc47e12 protein, partial [<i>Danio rerio</i>]
comp203863_c3_seq4	XP_692475.3	Predicted: uncharacterized protein si:ch211-210c8.6 [<i>Danio rerio</i>]
comp203972_c1_seq1	NP_001245156.1	Diverse immunoglobulin domain-containing protein 3.1 precursor [<i>Danio rerio</i>]
comp204037_c0_seq3	AAQ93680.1	Starmaker [<i>Danio rerio</i>]
comp204112_c1_seq1	NP_001180398.1	Selenoprotein j [<i>Danio rerio</i>]
comp204250_c0_seq1	NP_001129024.1	Uncharacterized protein loc557050 precursor [<i>Danio rerio</i>]
comp204250_c0_seq4	XP_009294323.1	Predicted: uncharacterized protein loc100329262 isoform x2 [<i>Danio rerio</i>]
comp204310_c1_seq9	XP_009062597.1	Hypothetical protein lotgidraft_166878 [<i>Lottia gigantea</i>]
comp204321_c0_seq4	WP_030294476.1	Nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase [<i>streptomyces katrae</i>]
comp204326_c0_seq3	XP_004573586.1	Predicted: uncharacterized protein loc101484042 [<i>Maylandia zebra</i>]
comp204690_c3_seq2	XP_002666721.1	Predicted: uncharacterized protein loc100332522 [<i>Danio rerio</i>]
comp204697_c0_seq1	XP_009063784.1	Hypothetical protein lotgidraft_155025 [<i>Lottia gigantea</i>]
comp204701_c0_seq8	XP_007535441.1	Predicted: ly6/plaur domain-containing protein 2-like [<i>Erinaceus europaeus</i>]
comp204725_c0_seq1	XP_001343422.1	Predicted: 3-dehydroquinase synthase, chloroplastic-like isoform x1 [<i>Danio rerio</i>]
comp205043_c0_seq4	XP_002666816.1	Predicted: transmembrane protein 138 [<i>Danio rerio</i>]
comp205160_c0_seq4	XP_009298807.1	Predicted: uncharacterized protein loc103910565 [<i>Danio rerio</i>]
comp205160_c0_seq5	XP_009298918.1	Predicted: uncharacterized protein loc103910680 [<i>Danio rerio</i>]
comp205416_c1_seq1	NP_956488.1	Small acidic protein [<i>Danio rerio</i>]

comp205428_c0_seq6	XP_005171742.2	Predicted: uncharacterized protein loc560730 isoform x1 [<i>Danio rerio</i>]
comp205479_c0_seq2	AAN71722.1	Putative tyrosine recombinase, partial [<i>Danio rerio</i>]
comp205671_c1_seq1	XP_004539690.1	Predicted: uncharacterized protein loc101488055 [<i>Maylandia zebra</i>]
comp205671_c1_seq3	XP_005912968.1	Predicted: uncharacterized protein loc102302972 [<i>Haplochromis burtoni</i>]
comp205744_c1_seq4	EYC03632.1	Hypothetical protein y032_0092g2532 [<i>Ancylostoma ceylanicum</i>]
comp205806_c3_seq12	XP_003201059.1	Predicted: uncharacterized protein loc100535646 isoform x1 [<i>Danio rerio</i>]
comp205806_c3_seq4	XP_005161701.1	Predicted: uncharacterized protein loc101883120 [<i>Danio rerio</i>]
comp205865_c0_seq1	XP_001866950.1	Conserved hypothetical protein [<i>Culex quinquefasciatus</i>]
comp205865_c0_seq8	ELT88501.1	Hypothetical protein captedraft_129409 [<i>Capitella teleta</i>]
comp206040_c2_seq10	XP_004576295.1	Predicted: uncharacterized protein loc101469477 isoform x1 [<i>Maylandia zebra</i>]
comp206363_c0_seq14	XP_001348153.1	Maubl, putative [<i>Plasmodium falciparum 3d7</i>]
comp206366_c2_seq2	XP_005157804.1	Predicted: lymphocyte antigen 6d [<i>Danio rerio</i>]
comp206424_c0_seq12	XP_001866083.1	Conserved hypothetical protein [<i>Culex quinquefasciatus</i>]
comp206441_c1_seq3	XP_005159349.1	Predicted: uncharacterized protein loc100034462 isoform x1 [<i>Danio rerio</i>]
comp206557_c0_seq4	AAC94994.1	Secretogranin ii precursor [<i>Carassius auratus</i>]
comp206577_c0_seq6	XP_005163742.1	Predicted: protein mnn4-like [<i>Danio rerio</i>]
comp206631_c0_seq5	XP_003460529.2	Predicted: uncharacterized protein loc100710209, partial [<i>Oreochromis niloticus</i>]
comp206631_c0_seq7	XP_009301083.1	Predicted: uncharacterized protein loc103911311 [<i>Danio rerio</i>]
comp206632_c0_seq11	XP_009301001.1	Predicted: high mobility group protein hmgi-c-like [<i>Danio rerio</i>]
comp206702_c0_seq6	XP_005159312.1	Predicted: protein fam118b isoform x2 [<i>Danio rerio</i>]
comp206726_c0_seq2	XP_009289661.1	Predicted: uncharacterized protein loc100537259 isoform x2 [<i>Danio rerio</i>]
comp206793_c0_seq17	XP_005168629.1	Predicted: uncharacterized protein loc101886761 [<i>Danio rerio</i>]
comp206979_c0_seq3	NP_001280600.1	Actinoporin-like protein [<i>Danio rerio</i>]
comp207121_c0_seq8	XP_009294652.1	Predicted: complement factor h-related protein 1-like isoform x2 [<i>Danio rerio</i>]
comp207132_c0_seq9	XP_005167555.1	Predicted: calcium-responsive transcription factor isoform x2 [<i>Danio rerio</i>]
comp207183_c0_seq15	XP_005167196.1	Predicted: uncharacterized protein loc101883002, partial [<i>Danio rerio</i>]
comp207184_c3_seq19	NP_001107058.1	Uncharacterized protein loc558926 precursor [<i>Danio rerio</i>]
comp207596_c0_seq12	XP_009296345.1	Predicted: uncharacterized protein loc101882425 [<i>Danio rerio</i>]
comp207696_c0_seq5	XP_009054533.1	Hypothetical protein lotgidraft_161044 [<i>Lottia gigantea</i>]

comp207792_c0_seq13	XP_005486042.1	Predicted: ben domain-containing protein 6 isoform x2 [<i>Zonotrichia albicollis</i>]
comp207792_c0_seq8	XP_002661275.1	Predicted: uncharacterized protein si:ch73-167f10.1 isoform x1 [<i>Danio rerio</i>]
comp207824_c0_seq4	XP_009301063.1	Predicted: cell wall protein rbr3-like [<i>Danio rerio</i>]
comp207947_c1_seq12	NP_942119.3	Fish-egg lectin-like precursor [<i>Danio rerio</i>]
comp207979_c1_seq13	XP_001922067.4	Predicted: slam family member 9-like [<i>Danio rerio</i>]
comp208153_c1_seq1	XP_009297318.1	Predicted: eukaryotic translation initiation factor 3 subunit a-like [<i>Danio rerio</i>]
comp20829_c0_seq1	NP_001122228.1	Neuromedin bb [<i>Danio rerio</i>]
comp208387_c1_seq1	XP_001336187.3	Predicted: trehalase [<i>Danio rerio</i>]
comp208438_c0_seq4	CDQ92604.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp208447_c0_seq4	AGK24957.1	Leptin b [<i>Cyprinus carpio 'jian'</i>]
comp208554_c0_seq3	NP_956785.1	Tbc1 domain family member 19 [<i>Danio rerio</i>]
comp208614_c0_seq1	ESA42264.1	Hypothetical protein ncu12107 [<i>Neurospora crassa or74a</i>]
comp208761_c2_seq5	XP_009299177.1	Predicted: uncharacterized protein loc100332323 [<i>Danio rerio</i>]
comp208783_c0_seq10	AAN71720.1	Putative gag protein [<i>Danio rerio</i>]
comp208783_c0_seq6	XP_009304434.1	Predicted: proline and serine-rich protein 1-like [<i>Danio rerio</i>]
comp208783_c0_seq9	XP_009297963.1	Predicted: uncharacterized protein loc101883550 [<i>Danio rerio</i>]
comp208976_c0_seq6	NP_001185496.1	Wd repeat-containing protein 76 [<i>Danio rerio</i>]
comp208987_c2_seq5	NP_001119930.1	Uncharacterized protein loc796378 [<i>Danio rerio</i>]
comp208995_c0_seq4	XP_005172492.1	Predicted: uncharacterized protein loc101883630 [<i>Danio rerio</i>]
comp209059_c0_seq14	XP_002004017.1	Gi19857 [<i>Drosophila mojavensis</i>]
comp20909_c0_seq1	ELU09913.1	Hypothetical protein captedraft_201962 [<i>Capitella teleta</i>]
comp209180_c4_seq4	NP_001006061.1	Coiled-coil domain-containing protein 113 [<i>Danio rerio</i>]
comp209209_c0_seq1	XP_009302348.1	Predicted: uncharacterized protein wu:fk66f10 isoform x4 [<i>Danio rerio</i>]
comp209209_c0_seq6	XP_005167250.1	Predicted: calphotin isoform x1 [<i>Danio rerio</i>]
comp209495_c0_seq4	XP_005100969.1	Predicted: uncharacterized protein loc101860022 [<i>Aplysia californica</i>]
comp209539_c0_seq4	ELU15108.1	Hypothetical protein captedraft_215954 [<i>Capitella teleta</i>]
comp209646_c0_seq4	NP_001070846.1	Nuclear-interacting partner of alk [<i>Danio rerio</i>]
comp209681_c1_seq7	XP_009297871.1	Predicted: uncharacterized protein loc103910281 [<i>Danio rerio</i>]
comp209684_c1_seq4	KFO62223.1	Hypothetical protein n302_02520, partial [<i>Corvus brachyrhynchos</i>]

comp209704_c0_seq4	NP_001076565.1	Membrane protein fam174b isoform 1 precursor [<i>Danio rerio</i>]
comp209848_c0_seq1	NP_001128614.1	Uncharacterized protein loc100189615 [<i>Danio rerio</i>]
comp209988_c3_seq4	CDO62370.1	Conserved plasmodium protein, unknown function [<i>Plasmodium reichenowi</i>]
comp210064_c2_seq3	XP_005470062.1	Predicted: uncharacterized protein loc102077948 [<i>Oreochromis niloticus</i>]
comp210064_c5_seq3	XP_008178660.1	Predicted: uncharacterized protein loc100574477 [<i>Acyrtosiphon pisum</i>]
comp210175_c0_seq1	XP_009298902.1	Predicted: uncharacterized protein loc101883144 [<i>Danio rerio</i>]
comp210175_c0_seq10	XP_009302544.1	Predicted: uncharacterized protein loc101885951 [<i>Danio rerio</i>]
comp210175_c0_seq3	XP_005168232.2	Predicted: uncharacterized protein loc100329771 [<i>Danio rerio</i>]
comp210175_c0_seq6	XP_008281292.1	Predicted: uncharacterized protein loc103358197 [<i>Stegastes partitus</i>]
comp210310_c1_seq12	XP_706537.4	Predicted: uncharacterized protein loc570021 [<i>Danio rerio</i>]
comp210343_c0_seq3	ETO34933.1	Cell surface glycoprotein, partial [<i>Reticulomyxa filosa</i>]
comp210350_c0_seq11	XP_002632808.1	Hypothetical protein cbg22662 [<i>Caenorhabditis briggsae</i>]
comp210488_c2_seq5	XP_686623.6	Predicted: ladinin-1, partial [<i>Danio rerio</i>]
comp210567_c0_seq2	NP_001138293.1	Uncharacterized protein loc100151367 precursor [<i>Danio rerio</i>]
comp210697_c0_seq13	YP_001096166.2	Unnamed protein product [<i>Cyprinid herpesvirus 3</i>]
comp210762_c0_seq7	XP_003199551.2	Predicted: uncharacterized protein sb:cb207 isoform x1 [<i>Danio rerio</i>]
comp210803_c0_seq6	WP_028459123.1	Hypothetical protein [<i>Chloroflexus sp. Y-396-1</i>]
comp210820_c1_seq9	XP_009292552.1	Predicted: uncharacterized protein loc103909240 [<i>Danio rerio</i>]
comp210840_c0_seq4	XP_009301418.1	Predicted: uncharacterized protein loc103911378 [<i>Danio rerio</i>]
comp211190_c0_seq12	XP_001913238.1	Hypothetical protein 003-32 [<i>Oikopleura dioica</i>]
comp211190_c0_seq7	XP_008434340.1	Predicted: probable serine/threonine-protein kinase clka, partial [<i>Poecilia reticulata</i>]
comp211212_c0_seq2	XP_009291168.1	Predicted: uncharacterized protein loc101886346 [<i>Danio rerio</i>]
comp211339_c0_seq2	XP_009295306.1	Predicted: zinc finger protein 761-like [<i>Danio rerio</i>]
comp21142_c0_seq1	KFM79180.1	Hypothetical protein x975_06366, partial [<i>Stegodyphus mimosarum</i>]
comp211563_c2_seq11	XP_009291448.1	Predicted: cell growth regulator with ef hand domain protein 1 [<i>Danio rerio</i>]
comp212006_c0_seq4	NP_001135735.1	Disrupted in schizophrenia 1 protein [<i>Danio rerio</i>]
comp212272_c0_seq4	XP_002660426.1	Predicted: iq domain-containing protein e isoform x1 [<i>Danio rerio</i>]
comp212542_c0_seq7	CDQ63792.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp212598_c0_seq4	XP_005168369.1	Predicted: uncharacterized protein si:ch211-132g1.3 isoform x1 [<i>Danio rerio</i>]

comp212870_c1_seq3	XP_009299522.1	Predicted: uncharacterized protein loc101885516 [<i>Danio rerio</i>]
comp213142_c2_seq5	XP_005457739.1	Predicted: uncharacterized protein loc102082910 [<i>Oreochromis niloticus</i>]
comp213237_c1_seq3	NP_001121746.1	Uncharacterized protein loc797544 [<i>Danio rerio</i>]
comp213306_c0_seq2	NP_001164503.1	Si:rp71-36n21.1 [<i>Danio rerio</i>]
comp213306_c0_seq5	NP_001139173.1	Si:ch211-168h21.3 [<i>Danio rerio</i>]
comp213524_c2_seq11	Q5RGS7.2	Spermatogenesis-associated protein 45 [<i>Danio rerio</i>]
comp213596_c2_seq5	AAH74098.1	Zgc:136474 protein, partial [<i>Danio rerio</i>]
comp213731_c0_seq10	A3KNI7.2	cyclin-d1-binding protein 1 homolog [<i>Danio rerio</i>]
comp213799_c0_seq11	XP_009299209.1	Predicted: uncharacterized protein loc103910949 isoform x1 [<i>Danio rerio</i>]
comp213799_c0_seq12	XP_009299178.1	Predicted: uncharacterized protein loc103910913 isoform x1 [<i>Danio rerio</i>]
comp214_c0_seq1	XP_007251900.1	Predicted: uncharacterized protein loc103037946 [<i>Astyanax mexicanus</i>]
comp214098_c0_seq13	XP_007249304.1	Predicted: uncharacterized protein loc103026616 isoform x1 [<i>Astyanax mexicanus</i>]
comp214408_c6_seq7	ELR52589.1	Hypothetical protein m91_18887, partial [<i>Bos mutus</i>]
comp214445_c0_seq10	XP_003971694.1	Predicted: uncharacterized protein loc101073018 [<i>Takifugu rubripes</i>]
comp214544_c0_seq1	XP_693881.4	Predicted: vitellogenin [<i>Danio rerio</i>]
comp214544_c0_seq4	AAI22326.1	Vtg3 protein, partial [<i>Danio rerio</i>]
comp214655_c2_seq5	XP_009294302.1	Predicted: uncharacterized protein loc100006656 isoform x1 [<i>Danio rerio</i>]
comp215471_c0_seq10	XP_005163007.1	Predicted: galaxin-like isoform x2 [<i>Danio rerio</i>]
comp215612_c1_seq10	XP_007238050.1	Predicted: uncharacterized protein loc103042335 [<i>Astyanax mexicanus</i>]
comp215947_c1_seq1	NP_001017892.1	Uncharacterized protein loc550591 [<i>Danio rerio</i>]
comp215958_c2_seq4	CDQ71004.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp216067_c0_seq1	XP_005158355.2	Predicted: uncharacterized protein loc101886405 [<i>Danio rerio</i>]
comp216115_c0_seq2	XP_005174258.1	Predicted: uncharacterized protein loc101882885 [<i>Danio rerio</i>]
comp216210_c0_seq2	XP_005158666.1	Predicted: uncharacterized threonine-rich gpi-anchored glycoprotein pj4664.02 [<i>Danio rerio</i>]
comp217689_c0_seq1	EKC28297.1	Hypothetical protein cgi_10021602 [<i>Crassostrea gigas</i>]
comp217706_c0_seq1	KDQ49121.1	Hypothetical protein jaaardraft_91570, partial [<i>Jaapia argillacea mucl 33604</i>]
comp218561_c0_seq1	XP_001032881.1	Hypothetical protein ttherm_00486690 [<i>Tetrahymena thermophila</i>]
comp21946_c0_seq1	NP_001002753.1	Coiled-coil domain-containing protein 103 [<i>Danio rerio</i>]
comp219669_c0_seq1	EFB22179.1	Hypothetical protein panda_015010, partial [<i>Ailuropoda melanoleuca</i>]

comp221320_c0_seq1	ELU10237.1	Hypothetical protein captedraft_206270, partial [<i>Capitella teleta</i>]
comp223309_c0_seq1	XP_005451942.1	Predicted: uncharacterized protein loc102078155 [<i>Oreochromis niloticus</i>]
comp224394_c0_seq1	KFM76332.1	Hypothetical protein x975_12329, partial [<i>Stegodyphus mimosarum</i>]
comp224775_c0_seq1	ELU00530.1	Hypothetical protein captedraft_192795 [<i>Capitella teleta</i>]
comp224811_c0_seq1	KFM75964.1	Hemocyanin c chain, partial [<i>Stegodyphus mimosarum</i>]
comp224982_c0_seq1	XP_008305108.1	Predicted: tmf-regulated nuclear protein 1 [<i>Stegastes partitus</i>]
comp225360_c0_seq1	BAJ91197.1	Predicted protein [<i>Hordeum vulgare subsp. Vulgare</i>]
comp226204_c0_seq1	XP_006790128.1	Predicted: uncharacterized protein loc102788911 [<i>Neolamprologus brichardi</i>]
comp227579_c0_seq1	XP_001341563.1	Predicted: thyrotropin subunit beta-like [<i>Danio rerio</i>]
comp227677_c0_seq1	AAI09414.1	Hypothetical protein mgc35261-like (h. Sapiens) [<i>Danio rerio</i>]
comp227938_c0_seq1	KFM71670.1	Hypothetical protein x975_08481, partial [<i>Stegodyphus mimosarum</i>]
comp228173_c0_seq1	XP_001645325.1	Hypothetical protein kpol_1058p4 [<i>Vanderwaltozyma polyspora dsm 70294</i>]
comp228218_c0_seq1	XP_009045810.1	Hypothetical protein lotgidraft_176384 [<i>Lottia gigantea</i>]
comp228482_c0_seq1	XP_009305953.1	Predicted: uncharacterized protein loc103912058, partial [<i>Danio rerio</i>]
comp229098_c0_seq1	XP_001342957.4	Predicted: uncharacterized protein si:ch73-303b9.1 [<i>Danio rerio</i>]
comp230193_c0_seq1	EMC81657.1	Hypothetical protein a306_10486, partial [<i>Columba livia</i>]
comp230235_c0_seq1	KFM80759.1	Hypothetical protein x975_07644, partial [<i>Stegodyphus mimosarum</i>]
comp230552_c0_seq1	XP_009295336.1	Predicted: ubx domain-containing protein 10 isoform x2 [<i>Danio rerio</i>]
comp230973_c0_seq1	ACB69423.1	Preprosomatostatin i [<i>Ctenopharyngodon idella</i>]
comp232648_c0_seq1	CDQ83808.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp232650_c0_seq1	XP_007256255.1	Predicted: forkhead-associated domain-containing protein 1-like [<i>Astyanax mexicanus</i>]
comp232911_c0_seq1	XP_005915870.1	Predicted: uncharacterized protein loc102298471 [<i>Haplochromis burtoni</i>]
comp233093_c0_seq1	CAP09281.1	Novel protein, partial [<i>Danio rerio</i>]
comp233606_c0_seq1	NP_956821.1	Sperm flagellar protein 1 [<i>Danio rerio</i>]
comp234350_c0_seq1	XP_005170335.1	Predicted: tumor suppressor candidate 5 homolog [<i>Danio rerio</i>]
comp234742_c0_seq1	XP_008867523.1	Hypothetical protein h310_04797 [<i>Aphanomyces invadans</i>]
comp234908_c0_seq1	XP_005927171.1	Predicted: uncharacterized protein c10orf105-like isoform x1 [<i>Haplochromis burtoni</i>]
comp235274_c0_seq1	BAJ97111.1	Predicted protein [<i>Hordeum vulgare subsp. Vulgare</i>]
comp235795_c0_seq1	KFU88813.1	Hypothetical protein m959_11072, partial [<i>Chaetura pelagica</i>]

comp236149_c0_seq1	XP_005159357.1	Predicted: sodium/nucleoside cotransporter 1 isoform x1 [<i>Danio rerio</i>]
comp237270_c0_seq1	XP_008302474.1	Predicted: uncharacterized protein loc103374206, partial [<i>Stegastes partitus</i>]
comp237944_c0_seq1	XP_009306085.1	Predicted: uncharacterized protein loc103912075 [<i>Danio rerio</i>]
comp238565_c0_seq1	XP_009297746.1	Predicted: transmembrane protein 114 [<i>Danio rerio</i>]
comp238610_c0_seq1	NP_001104680.1	Coiled-coil domain-containing protein 153 [<i>Danio rerio</i>]
comp240311_c0_seq1	BAC98932.1	Chitinase [<i>Plasmodium yoelii</i>]
comp240803_c0_seq1	XP_005159359.1	Predicted: neuromedin-b isoform x1 [<i>Danio rerio</i>]
comp242384_c0_seq1	KFM68761.1	Hypothetical protein x975_13337, partial [<i>Stegodyphus mimosarum</i>]
comp242450_c0_seq1	NP_001154965.1	Overexpressed in colon carcinoma 1 protein homolog [<i>Danio rerio</i>]
comp242773_c0_seq1	KFZ62107.1	Hypothetical protein n338_07810, partial [<i>podiceps cristatus</i>]
comp243346_c0_seq1	XP_009293513.1	Predicted: uncharacterized protein loc103909430 isoform x2 [<i>Danio rerio</i>]
comp243449_c0_seq1	KFM70847.1	Hypothetical protein x975_27234, partial [<i>Stegodyphus mimosarum</i>]
comp243597_c0_seq1	CAI12014.1	Novel protein [<i>Danio rerio</i>]
comp244128_c0_seq1	ABC47326.1	Tegumental protein 20.8 kda [<i>Clonorchis sinensis</i>]
comp244160_c0_seq1	XP_006125984.1	Predicted: uncharacterized protein c12orf60 homolog [<i>Pelodiscus sinensis</i>]
comp244613_c0_seq1	XP_009298152.1	Predicted: putative uncharacterized protein c16orf96 homolog isoform x1 [<i>Danio rerio</i>]
comp245177_c0_seq1	EOB08063.1	Hypothetical protein anapl_00298, partial [<i>Anas platyrhynchos</i>]
comp245274_c0_seq1	XP_687326.2	Predicted: protein kiaa1045 homolog isoform x1 [<i>Danio rerio</i>]
comp245743_c0_seq1	CDQ92329.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp245820_c0_seq1	CDQ81533.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp246054_c0_seq1	XP_007259767.1	Predicted: uncharacterized protein loc103046630 [<i>Astyanax mexicanus</i>]
comp246421_c0_seq1	CDI96865.1	Expressed conserved protein [<i>Echinococcus multilocularis</i>]
comp247078_c0_seq1	XP_005167195.1	Predicted: vegetative cell wall protein gp1-like [<i>Danio rerio</i>]
comp247131_c0_seq1	XP_009299262.1	Predicted: myelin regulatory factor-like protein isoform x2 [<i>Danio rerio</i>]
comp247917_c0_seq1	CDQ71087.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp248115_c0_seq1	ABU41102.1	Zeelin1-like protein [<i>Lepeophtheirus salmonis</i>]
comp248240_c0_seq1	ADD73552.1	Vitellogenin 2 [<i>Paracyclopsina nana</i>]
comp248312_c0_seq1	KFM59933.1	Mfs-type transporter, partial [<i>Stegodyphus mimosarum</i>]
comp248731_c0_seq1	AEB60990.1	Replication associated protein [<i>Barbel circovirus</i>]

comp250345_c0_seq1	KFM60372.1	Hypothetical protein x975_20530, partial [<i>Stegodyphus mimosarum</i>]
comp250700_c0_seq1	ELU14781.1	Hypothetical protein captedraft_210007 [<i>Capitella teleta</i>]
comp250749_c0_seq1	KFM80341.1	Hypothetical protein x975_14307, partial [<i>Stegodyphus mimosarum</i>]
comp250935_c0_seq1	XP_009021495.1	Hypothetical protein helrodraft_175877 [<i>Helobdella robusta</i>]
comp251012_c0_seq1	KFM73502.1	Hemocyanin d chain, partial [<i>Stegodyphus mimosarum</i>]
comp252015_c0_seq1	CDQ86624.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp252934_c0_seq1	XP_706387.3	Predicted: uncharacterized protein si:dkeyp-69c1.9 [<i>Danio rerio</i>]
comp25379_c0_seq1	ADM14321.1	Minor ampullate spidroin, partial [<i>Latrodectus hesperus</i>]
comp255089_c0_seq1	CDQ90848.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp255237_c0_seq1	XP_002665640.3	Predicted: stathmin domain-containing protein 1 [<i>Danio rerio</i>]
comp255319_c0_seq1	XP_001625190.1	Predicted protein [<i>Nematostella vectensis</i>]
comp255457_c0_seq1	XP_007578992.1	Predicted: putative uncharacterized protein c14orf132 homolog [<i>Poecilia formosa</i>]
comp255876_c0_seq1	NP_001138717.1	Secretory calcium-binding phosphoprotein 9 precursor [<i>Danio rerio</i>]
comp256380_c0_seq1	XP_008188260.1	Predicted: uncharacterized protein loc103310749 [<i>Acyrtosiphon pisum</i>]
comp256491_c0_seq1	XP_004346643.1	Hypothetical protein caog_04958 [<i>Capsaspora owczarzewski</i> atcc 30864]
comp256514_c0_seq1	BAJ95050.1	Predicted protein [<i>Hordeum vulgare subsp. Vulgare</i>]
comp256796_c0_seq1	XP_009301231.1	Predicted: mesoderm posterior ba isoform x1 [<i>Danio rerio</i>]
comp257695_c0_seq1	XP_005155405.1	Predicted: uncharacterized protein loc101882400 isoform x2 [<i>Danio rerio</i>]
comp258237_c0_seq1	KFM67624.1	Hypothetical protein x975_22945, partial [<i>Stegodyphus mimosarum</i>]
comp258903_c0_seq1	AAH83463.1	H1m protein, partial [<i>Danio rerio</i>]
comp259100_c0_seq1	XP_001624051.1	Predicted protein [<i>Nematostella vectensis</i>]
comp259289_c0_seq1	XP_006814831.1	Predicted: uncharacterized protein loc100372610 [<i>Saccoglossus kowalevskii</i>]
comp259302_c0_seq1	XP_005170775.1	Predicted: upf0583 protein c15orf59 homolog [<i>Danio rerio</i>]
comp260086_c0_seq1	XP_005729388.1	Predicted: uncharacterized protein loc102210508 [<i>Pundamilia nyererei</i>]
comp260814_c0_seq1	NP_001077328.1	Uncharacterized protein loc793007 precursor [<i>Danio rerio</i>]
comp260856_c0_seq1	KFM80577.1	Hypothetical protein x975_05627, partial [<i>Stegodyphus mimosarum</i>]
comp260975_c0_seq1	CAD68052.1	Hemocyanin subunit a [<i>Nephila inaurata madagascariensis</i>]
comp261633_c0_seq1	KFM77447.1	Hypothetical protein x975_19248, partial [<i>Stegodyphus mimosarum</i>]
comp261786_c0_seq1	KFM68743.1	Hypothetical protein x975_12832, partial [<i>Stegodyphus mimosarum</i>]

comp262155_c0_seq1	XP_001338098.4	Predicted: adhesive plaque matrix protein [<i>Danio rerio</i>]
comp262620_c0_seq1	CDS26988.1	Mrna binding protein [<i>Hymenolepis microstoma</i>]
comp263371_c0_seq1	NP_001165870.1	Prokineticin-2 precursor [<i>Danio rerio</i>]
comp263443_c0_seq1	XP_005174537.1	Predicted: linker histone h1m isoform x1 [<i>Danio rerio</i>]
comp263453_c0_seq1	NP_001121872.1	Uncharacterized protein loc100150067 [<i>Danio rerio</i>]
comp263757_c0_seq1	YP_004376332.1	Putative replication protein [<i>Barbel circovirus</i>]
comp263861_c0_seq1	XP_002606696.1	Hypothetical protein brafldraft_72537 [<i>Branchiostoma floridae</i>]
comp265453_c0_seq1	XP_001627046.1	Predicted protein [<i>Nematostella vectensis</i>]
comp265548_c0_seq1	ABU41024.1	Hypothetical protein [<i>Lepeophtheirus salmonis</i>]
comp265646_c0_seq1	GAA49983.1	Tegument antigen [<i>Clonorchis sinensis</i>]
comp265765_c0_seq1	XP_002592393.1	Hypothetical protein brafldraft_67258 [<i>Branchiostoma floridae</i>]
comp266020_c0_seq1	CDQ73355.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp266202_c0_seq1	XP_009301836.1	Predicted: uncharacterized protein loc101886861 [<i>Danio rerio</i>]
comp266649_c0_seq1	NP_001108596.1	Uncharacterized protein c4orf22 homolog [<i>Danio rerio</i>]
comp266783_c0_seq1	CDQ74554.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp267277_c0_seq1	AHZ63110.1	Conserved hypothetical protein [<i>Mesobuthus gibbosus</i>]
comp2677_c0_seq1	XP_002404039.1	Conserved hypothetical protein [<i>Ixodes scapularis</i>]
comp2677_c1_seq1	ELU12663.1	Hypothetical protein captedraft_39725, partial [<i>Capitella teleta</i>]
comp267992_c0_seq1	AAI29396.1	Zgc:158703 [<i>Danio rerio</i>]
comp268606_c0_seq1	NP_001098409.1	Membrane protein mlc1 [<i>Danio rerio</i>]
comp269420_c0_seq1	NP_955465.1	Meiosis-specific nuclear structural protein 1 [<i>Danio rerio</i>]
comp269690_c0_seq1	XP_002666785.2	Predicted: uncharacterized protein loc100330348 [<i>Danio rerio</i>]
comp270619_c0_seq1	NP_001038594.1	Iq motif-containing protein h [<i>Danio rerio</i>]
comp27077_c0_seq1	NP_001017684.1	Cytochrome b5 domain-containing protein 1 [<i>Danio rerio</i>]
comp271468_c0_seq1	XP_009821559.1	Hypothetical protein h257_00527 [<i>Aphanomyces astaci</i>]
comp272196_c0_seq1	XP_001626044.1	Predicted protein [<i>Nematostella vectensis</i>]
comp272286_c0_seq1	ACI41238.1	Aggregate spider glue 1 [<i>Nephila clavipes</i>]
comp272415_c0_seq1	XP_009826732.1	Hypothetical protein h257_04058 [<i>Aphanomyces astaci</i>]
comp272844_c0_seq1	XP_009299020.1	Predicted: interleukin-15-like [<i>Danio rerio</i>]

comp272896_c0_seq1	XP_005173123.1	Predicted: protein fam196b-like [<i>Danio rerio</i>]
comp275390_c0_seq1	XP_008279400.1	Predicted: uncharacterized protein loc103356868 [<i>Stegastes partitus</i>]
comp275668_c0_seq1	XP_009300445.1	Predicted: cyclic nucleotide-binding domain-containing protein 1 isoform x2 [<i>Danio rerio</i>]
comp275787_c0_seq1	EFN89541.1	Hypothetical protein eai_00719, partial [<i>Harpegnathos saltator</i>]
comp275827_c0_seq1	NP_001008629.1	Pih1 domain-containing protein 2 [<i>Danio rerio</i>]
comp276081_c0_seq1	CDQ92388.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp276852_c0_seq1	CAQ15258.1	Novel protein similar to human chromosome 14 open reading frame 115 [<i>Danio rerio</i>]
comp277003_c0_seq1	KDO23594.1	Hypothetical protein sprg_10789 [<i>Saprolegnia parasitica cbs 223.65</i>]
comp277232_c0_seq1	NP_001038597.1	Uncharacterized protein loc567437 [<i>Danio rerio</i>]
comp277801_c0_seq1	KFM61135.1	Hypothetical protein x975_04514, partial [<i>Stegodyphus mimosarum</i>]
comp277953_c0_seq1	NP_001121849.1	Kinocilin [<i>Danio rerio</i>]
comp278780_c0_seq1	XP_002600389.1	Hypothetical protein brafldraft_99580 [<i>Branchiostoma floridae</i>]
comp278992_c0_seq1	BAP68993.1	Rxlr effector candidate protein [<i>Hyaloperonospora arabidopsidis emoy2</i>]
comp279700_c0_seq1	XP_706476.2	Predicted: uncharacterized protein loc569091 isoform x2 [<i>Danio rerio</i>]
comp280344_c0_seq1	XP_005159810.1	Predicted: sushi domain-containing protein 5 [<i>Danio rerio</i>]
comp280834_c0_seq1	ABX75487.1	Unkown [<i>Lycosa singoriensis</i>]
comp280849_c0_seq1	KFQ77770.1	Hypothetical protein n335_13478, partial [<i>Phaethon lepturus</i>]
comp281066_c0_seq1	NP_001037800.1	Uncharacterized protein loc555550 [<i>Danio rerio</i>]
comp281196_c0_seq1	XP_001622016.1	Hypothetical protein nemvedraft_v1g221280 [<i>Nematostella vectensis</i>]
comp281695_c0_seq1	XP_008617247.1	Hypothetical protein sdrg_12941 [<i>Saprolegnia diclina vs20</i>]
comp281784_c0_seq1	KFM62296.1	Hypothetical protein x975_04030, partial [<i>Stegodyphus mimosarum</i>]
comp282118_c0_seq1	KFM71246.1	Hypothetical protein x975_24110, partial [<i>Stegodyphus mimosarum</i>]
comp282207_c0_seq1	KFM73035.1	Hypothetical protein x975_00114, partial [<i>Stegodyphus mimosarum</i>]
comp282248_c0_seq1	KGB35999.1	Hypothetical protein ms3_04262 [<i>schistosoma haematobium</i>]
comp282344_c0_seq1	XP_009290056.1	Predicted: uncharacterized protein loc103908851 [<i>Danio rerio</i>]
comp282612_c0_seq1	XP_009025828.1	Hypothetical protein helrodraft_163805 [<i>Helobdella robusta</i>]
comp282656_c0_seq1	WP_000607649.1	Hypothetical protein, partial [<i>Bacillus cereus</i>]
comp283669_c0_seq1	XP_009290547.1	Predicted: serine/threonine-protein kinase 31-like [<i>Danio rerio</i>]
comp285079_c0_seq1	XP_009305958.1	Predicted: uncharacterized protein loc103912060 [<i>Danio rerio</i>]

comp285103_c0_seq1	YP_007003779.1	Protein orf121 [<i>Cyprinid herpesvirus 1</i>]
comp285134_c0_seq1	KFM64041.1	Mite allergen lep d 7, partial [<i>Stegodyphus mimosarum</i>]
comp285212_c0_seq1	CCD79478.1	Hypothetical protein smp_171780 [<i>Schistosoma mansoni</i>]
comp285398_c0_seq1	ADE06666.1	Insulin-like growth factor binding protein-related protein 6 short [<i>Amblyomma americanum</i>]
comp286316_c0_seq1	CDQ76813.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp286554_c0_seq1	KFM58788.1	Hypothetical protein x975_14350, partial [<i>Stegodyphus mimosarum</i>]
comp286691_c0_seq1	CAD68055.1	Hemocyanin subunit e [<i>Nephila inaurata madagascariensis</i>]
comp286843_c0_seq1	XP_001641912.1	Predicted protein [<i>Nematostella vectensis</i>]
comp287535_c0_seq1	XP_003199288.1	Predicted: uncharacterized protein si:ch211-218d20.15 isoform x1 [<i>Danio rerio</i>]
comp287783_c0_seq1	XP_001021414.1	Hypothetical protein ttherm_00317330 [<i>Tetrahymena thermophila</i>]
comp287859_c0_seq1	ACH47562.1	Photosystem ii protein d1 (chloroplast) [<i>Geranium macrorrhizum</i>]
comp288352_c0_seq1	CAE45012.1	Interleukin-12 p35 subunit [<i>Cyprinus carpio</i>]
comp288927_c0_seq1	XP_001921262.3	Predicted: wd repeat-containing protein 93 isoform x2 [<i>Danio rerio</i>]
comp288977_c0_seq1	XP_008879461.1	Fructose-bisphosphate aldolase, class ii [<i>Aphanomyces invadans</i>]
comp289907_c0_seq1	ELT87537.1	Hypothetical protein captedraft_192823 [<i>Capitella teleta</i>]
comp289932_c0_seq1	KFM83237.1	Hypothetical protein x975_25266, partial [<i>Stegodyphus mimosarum</i>]
comp290656_c0_seq1	NP_001116743.1	Small integral membrane protein 24 precursor [<i>Danio rerio</i>]
comp290796_c0_seq1	NP_001121859.1	Odorant receptor, family h, subfamily 137, member 2 [<i>Danio rerio</i>]
comp291438_c0_seq1	XP_007237349.1	Predicted: probable g-protein coupled receptor 149, partial [<i>Astyanax mexicanus</i>]
comp291792_c0_seq1	XP_004212018.1	Predicted: uncharacterized protein loc101239309 [<i>Hydra vulgaris</i>]
comp291977_c0_seq1	XP_001458073.1	Hypothetical protein [<i>Paramecium tetraurelia strain d4-2</i>]
comp292240_c0_seq1	CDI96482.1	Hypothetical protein emuj_000005500 [<i>Echinococcus multilocularis</i>]
comp292333_c0_seq1	CDQ92909.1	Unnamed protein product, partial [<i>Oncorhynchus mykiss</i>]
comp292642_c0_seq1	EMD31869.1	Hypothetical protein cersudraft_162701, partial [<i>Ceriporiopsis subvermispora b</i>]
comp292663_c0_seq1	NP_001094099.1	Taste receptor, type 2, member 202 [<i>Danio rerio</i>]
comp292701_c0_seq1	XP_008189844.1	Predicted: uncharacterized protein loc103311853 [<i>Acyrtosiphon pisum</i>]
comp292797_c0_seq1	XP_001681641.1	Conserved hypothetical protein [<i>Leishmania major strain friedlin</i>]
comp293455_c0_seq1	XP_007229455.1	Predicted: uncharacterized protein loc103036065 [<i>Astyanax mexicanus</i>]
comp293872_c0_seq1	YP_007003729.1	Putative capsid triplex subunit 1 [<i>Cyprinid herpesvirus 1</i>]

comp293981_c0_seq1	XP_008861965.1	Hypothetical protein h310_00821 [<i>Aphanomyces invadans</i>]
comp294025_c0_seq1	KFM82414.1	Hypothetical protein x975_23666, partial [<i>Stegodyphus mimosarum</i>]
comp294362_c0_seq1	XP_009298477.1	Predicted: tetratricopeptide repeat protein gnn-like isoform x3 [<i>Danio rerio</i>]
comp294940_c0_seq1	KFM74767.1	Hypothetical protein x975_10850, partial [<i>Stegodyphus mimosarum</i>]
comp295235_c0_seq1	XP_009304755.1	Predicted: calymmin isoform x6 [<i>Danio rerio</i>]
comp295273_c0_seq1	ACF19412.1	Major ampullate spidroin 1b precursor, partial [<i>Nephila clavipes</i>]
comp295993_c0_seq1	NP_001119861.1	Uncharacterized protein loc557261 [<i>Danio rerio</i>]
comp30380_c0_seq1	KFM79465.1	Cuticle protein 10.9, partial [<i>Stegodyphus mimosarum</i>]
comp31775_c0_seq1	XP_005459208.1	Predicted: uncharacterized protein loc100703023 [<i>Oreochromis niloticus</i>]
comp3201_c0_seq1	XP_007229630.1	Predicted: transmembrane protein 25-like [<i>Astyanax mexicanus</i>]
comp32874_c0_seq1	XP_005810454.1	Predicted: uncharacterized protein loc102218311 [<i>Xiphophorus maculatus</i>]
comp33098_c0_seq1	XP_003199543.1	Predicted: uncharacterized protein wu:fi42e03 [<i>Danio rerio</i>]
comp33172_c0_seq1	CAN80145.1	Hypothetical protein vitisv_015720 [<i>Vitis vinifera</i>]
comp33187_c0_seq1	WP_010893150.1	Hypothetical protein [<i>Xylella fastidiosa</i>]
comp33659_c0_seq1	EDN65133.1	Hypothetical protein bgp_6678 [<i>Beggiatoa sp. Ps</i>]
comp37897_c0_seq1	XP_006799421.1	Predicted: uncharacterized protein loc102778332 [<i>Neolamprologus brichardi</i>]
comp38924_c0_seq1	XP_009299917.1	Predicted: uncharacterized protein loc541433 isoform x1 [<i>Danio rerio</i>]
comp38950_c0_seq1	XP_009291427.1	Predicted: heat repeat-containing protein 4 isoform x1 [<i>Danio rerio</i>]
comp39305_c0_seq1	XP_002516402.1	Conserved hypothetical protein [<i>Ricinus communis</i>]
comp39893_c0_seq1	XP_007250673.1	Predicted: uncharacterized protein loc103037290 isoform x2 [<i>Astyanax mexicanus</i>]
comp41726_c0_seq1	ADD73551.1	Vitellogenin 1 [<i>Paracyclopsina nana</i>]
comp42586_c0_seq1	KFM78497.1	Hypothetical protein x975_19974, partial [<i>Stegodyphus mimosarum</i>]
comp43208_c0_seq1	CDQ60959.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp47526_c0_seq1	XP_009305952.1	Predicted: centrosomal protein of 83 kda-like [<i>Danio rerio</i>]
comp48087_c0_seq1	AAH95007.1	Oep protein [<i>Danio rerio</i>]
comp48370_c0_seq1	XP_002935199.2	Predicted: uncharacterized protein loc100490020 [<i>Xenopus (silurana) tropicalis</i>]
comp50546_c0_seq1	ACO51129.1	Unknown, partial [<i>Hypophthalmichthys nobilis</i>]
comp50793_c0_seq1	KFM73807.1	Hypothetical protein x975_26840, partial [<i>Stegodyphus mimosarum</i>]
comp52750_c0_seq1	XP_005159700.1	Predicted: brain and acute leukemia cytoplasmic protein [<i>Danio rerio</i>]

comp52933_c0_seq1	ABU41045.1	Hypothetical protein [<i>Lepeophtheirus salmonis</i>]
comp52951_c1_seq1	XP_001342546.1	Predicted: uncharacterized protein si:ch211-167j6.3 isoform x1 [<i>Danio rerio</i>]
comp53337_c0_seq1	XP_006628839.1	Predicted: uncharacterized protein c15orf65-like [<i>Lepisosteus oculatus</i>]
comp5592_c0_seq1	EFT09945.1	Hypothetical protein hmpref9619_01598 [<i>Propionibacterium acnes hl082pa2</i>]
comp57565_c0_seq1	XP_009290103.1	Predicted: sialic acid-binding ig-like lectin 12 isoform x1 [<i>Danio rerio</i>]
comp58016_c0_seq1	AAH91806.1	Loc553336 protein, partial [<i>Danio rerio</i>]
comp58598_c0_seq1	XP_009027343.1	Hypothetical protein helrodraft_164044 [<i>Helobdella robusta</i>]
comp58638_c0_seq2	KGE11290.1	Copper oxidase [<i>Burkholderia gladioli</i>]
comp60079_c0_seq1	XP_005952964.1	Predicted: uncharacterized protein loc102300618 [<i>Haplochromis burtoni</i>]
comp60405_c0_seq1	XP_003199616.1	Predicted: sterile alpha motif domain-containing protein 14 [<i>Danio rerio</i>]
comp61222_c0_seq1	XP_005169615.1	Predicted: microtubule-associated protein 10 [<i>Danio rerio</i>]
comp62011_c0_seq1	KFD67411.1	Low quality protein: hypothetical protein m514_20447 [<i>Trichuris suis</i>]
comp651_c0_seq1	KGB42270.1	Cercarial protease, partial [<i>Schistosoma haematobium</i>]
comp66313_c0_seq1	AHX03164.1	Interleukin-23 p19 [<i>Ctenopharyngodon idella</i>]
comp66614_c0_seq1	XP_005163268.1	Predicted: transmembrane emp24 domain-containing protein 6 [<i>Danio rerio</i>]
comp7091_c0_seq1	NP_001153485.1	Meiosis-specific protein mei4-like [<i>Danio rerio</i>]
comp71134_c0_seq1	WP_011078084.1	Hypothetical protein [<i>Erwinia amylovora</i>]
comp74233_c0_seq1	CDQ74374.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp75381_c1_seq1	XP_009293624.1	Predicted: uncharacterized protein loc103909444 isoform x1 [<i>Danio rerio</i>]
comp79492_c0_seq1	XP_003197712.2	Predicted: secreted phosphoprotein 24 [<i>Danio rerio</i>]
comp79739_c0_seq1	CDR00723.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp80321_c0_seq1	ADF97604.1	Zonadhesin, partial [<i>Hypophthalmichthys molitrix</i>]
comp80663_c0_seq1	XP_706281.1	Predicted: uncharacterized protein si:dkey-30e9.6 isoform x1 [<i>Danio rerio</i>]
comp80859_c0_seq1	XP_001345297.1	Predicted: uncharacterized protein loc100006601 [<i>Danio rerio</i>]
comp81574_c0_seq1	XP_001344930.3	Predicted: uncharacterized protein c18orf63 homolog [<i>Danio rerio</i>]
comp8287_c0_seq1	XP_005164735.1	Predicted: a-kinase anchor protein 14 [<i>Danio rerio</i>]
comp83147_c0_seq1	XP_003728146.1	Predicted: uncharacterized protein loc100890688 [<i>Strongylocentrotus purpuratus</i>]
comp8349_c0_seq1	AFV60022.1	Anti-mullerian hormone [<i>Carassius auratus ssp. 'Pengze'</i>]
comp85966_c1_seq1	XP_001334391.5	Predicted: protein fam133-like [<i>Danio rerio</i>]

comp86128_c0_seq1	XP_003647814.1	Hypothetical protein ecym_7148 [<i>Erethothecium cymbalariae</i> dbvpg#7215]
comp86988_c0_seq1	XP_005171600.1	Predicted: g-protein coupled receptor family c group 5 member b-like [<i>Danio rerio</i>]
comp87498_c0_seq1	NP_001124245.1	Uncharacterized protein loc559801 [<i>Danio rerio</i>]
comp87978_c0_seq1	NP_001108367.1	Uncharacterized protein loc100141330 [<i>Danio rerio</i>]
comp8801_c0_seq1	XP_008183038.1	Predicted: uncharacterized protein loc103309427 [<i>Acyrtosiphon pisum</i>]
comp88981_c0_seq1	GAA38700.1	Hypothetical protein clf_101970 [<i>Clonorchis sinensis</i>]
comp9021_c0_seq1	XP_009296140.1	Predicted: uncharacterized protein loc568519 isoform x2 [<i>Danio rerio</i>]
comp90648_c0_seq1	KFM60578.1	Hypothetical protein x975_21288, partial [<i>Stegodyphus mimosarum</i>]
comp91139_c0_seq2	XP_006815409.1	Predicted: uncharacterized protein loc102808709 [<i>Saccoglossus kowalevskii</i>]
comp91334_c0_seq1	KFM72943.1	Hypothetical protein x975_17424, partial [<i>Stegodyphus mimosarum</i>]
comp91626_c0_seq1	KFM78496.1	Hypothetical protein x975_19973, partial [<i>Stegodyphus mimosarum</i>]
comp9168_c0_seq2	CDR07108.1	Lysophospholipase, putative [<i>Plasmodium chabaudi chabaudi</i>]
comp92594_c0_seq1	NP_001138709.1	Secretory calcium-binding phosphoprotein 6 precursor [<i>Danio rerio</i>]
comp92659_c0_seq1	XP_009289505.1	Predicted: testis-expressed sequence 11 protein [<i>Danio rerio</i>]
comp92839_c0_seq1	XP_009303509.1	Predicted: rna polymerase-associated protein leo1-like isoform x2 [<i>Danio rerio</i>]
comp93032_c0_seq1	AAO39752.1	Salmon-type gonadotropin-releasing hormone precursor [<i>Cyprinus carpio</i>]
comp93093_c1_seq1	CDQ96703.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
comp93260_c0_seq1	KFM79892.1	Hypothetical protein x975_26638, partial [<i>Stegodyphus mimosarum</i>]
comp93526_c0_seq1	XP_683018.3	Predicted: uncharacterized protein si:ch211-242e8.1 isoform x1 [<i>Danio rerio</i>]
comp93863_c0_seq1	XP_005452781.1	Predicted: uncharacterized protein loc102077212 isoform x1 [<i>Oreochromis niloticus</i>]
comp94387_c0_seq1	XP_006819733.1	Predicted: sporulation-specific protein 15-like [<i>Saccoglossus kowalevskii</i>]
comp94820_c0_seq1	XP_002663575.3	Predicted: mucin-5ac [<i>Danio rerio</i>]
comp94827_c0_seq1	XP_009293501.1	Predicted: nose resistant to fluoxetine protein 6-like [<i>Danio rerio</i>]
comp96758_c0_seq1	CAA31963.1	Unnamed protein product [<i>Cyprinus carpio</i>]
comp97382_c0_seq1	AAI22420.1	Loc100006494 protein, partial [<i>Danio rerio</i>]
comp97421_c0_seq1	XP_001340218.3	Predicted: uncharacterized protein loc799914 isoform x1 [<i>Danio rerio</i>]
comp99365_c0_seq1	XP_006801328.1	Predicted: uncharacterized protein loc102778642 [<i>Neolamprologus brichardi</i>]
comp99443_c0_seq1	AAH95239.1	Zgc:92753 protein [<i>Danio rerio</i>]
comp99788_c0_seq1	XP_009302287.1	Predicted: potassium voltage-gated channel subfamily kqt member 2-like [<i>Danio rerio</i>]

The cross intersection of differentially expressed genes during carp scale regeneration and genes only expressed in common carp skin, but not in channel catfish skin

The scale regeneration experiments were conducted in common carp and 1,173 differentially expressed genes were identified during scale regeneration. Further, the interspecific transcriptome comparison between scaled carp and scaleless channel catfish that revealed 836 genes expressed in carp but not in channel catfish (Table 14). The cross intersection between the interspecific skin transcriptomes and the differentially expressed genes during scale regeneration revealed 13 known genes and five unknown genes as overlap genes (Table 15). Among these 18 genes, 10 known genes were up-regulated, including 14 kDa apolipoprotein, actinoporin-like protein, apolipoprotein A-I, apolipoprotein AIb1, lymphocyte antigen 6D, protein FAM133-like, protein GAPT-like, secretory calcium-binding phosphoprotein 7, si:dkey-22i16.3 (fa93e10), and uroplakin 2-like; while three known gene was down-regulated, including fish-egg lectin-like, myosin heavy chain, clone 203, and zymogen granule membrane protein 16-like.

Of the shared 18 genes, three genes were determined to be significantly up-regulated after scale removal 12 h, nine genes differentially up-regulated after scale removal 3 days or 5 days and three genes differentially expressed after scale removal 14 days and three genes were determined to be down-regulated during scale regeneration (Figure 7). The up-regulated genes detected at early stages after scale removal included GAPT, si:dkey-27h10.2 and UPK21, while up-regulated genes detected at 3 days or 5 days after scale removal included APNL, FAM133, Fa93e10, LOC100330485, LOC101883268, LY6D, SCPP7, Si:dkey-30j10.5, and Wu:fk35f04.

Of these up-regulated genes, some are only detected at one or two time points with significant up-regulation after scale removal, such as APNL, LOC100330485, LOC101883268, and LY6D, while several genes were more stably up-regulated during scale regeneration such as FAM133, Fa93e10, SCPP7, Si:dkey-30j10.5, and Wu:fk35f04. Interestingly, all three significant up-regulated genes at 14 days after scale removal are apolipoprotein, including 14 kDa apolipoprotein, apolipoprotein A-I, and apolipoprotein AIb1 (Figure 7).

Table 15. The cross intersection of 18 genes including 13 known genes and five uncharacterized genes that were differentially expressed during carp scale regeneration and only expressed in common carp skin, but not in channel catfish skin.

Up-regulated genes
14 kDa apolipoprotein
Actinoporin-like protein
Apolipoprotein A-I
Apolipoprotein AIb1
Lymphocyte antigen 6D
Protein FAM133-like
Protein GAPT-like
Secretory calcium-binding phosphoprotein 7
Si:dkey-22i16.3 (fa93e10)
Uncharacterized protein LOC100330485
Uncharacterized protein LOC101883268
Uncharacterized protein si:dkey-27h10.2 isoform X1
Uncharacterized protein si:dkey-30j10.5
Uncharacterized protein wu:fk35f04
Uroplakin 2-like isoform X1

Down-regulated genes
Fish-egg lectin-like
Myosin heavy chain, clone 203
Zymogen granule membrane protein 16-like

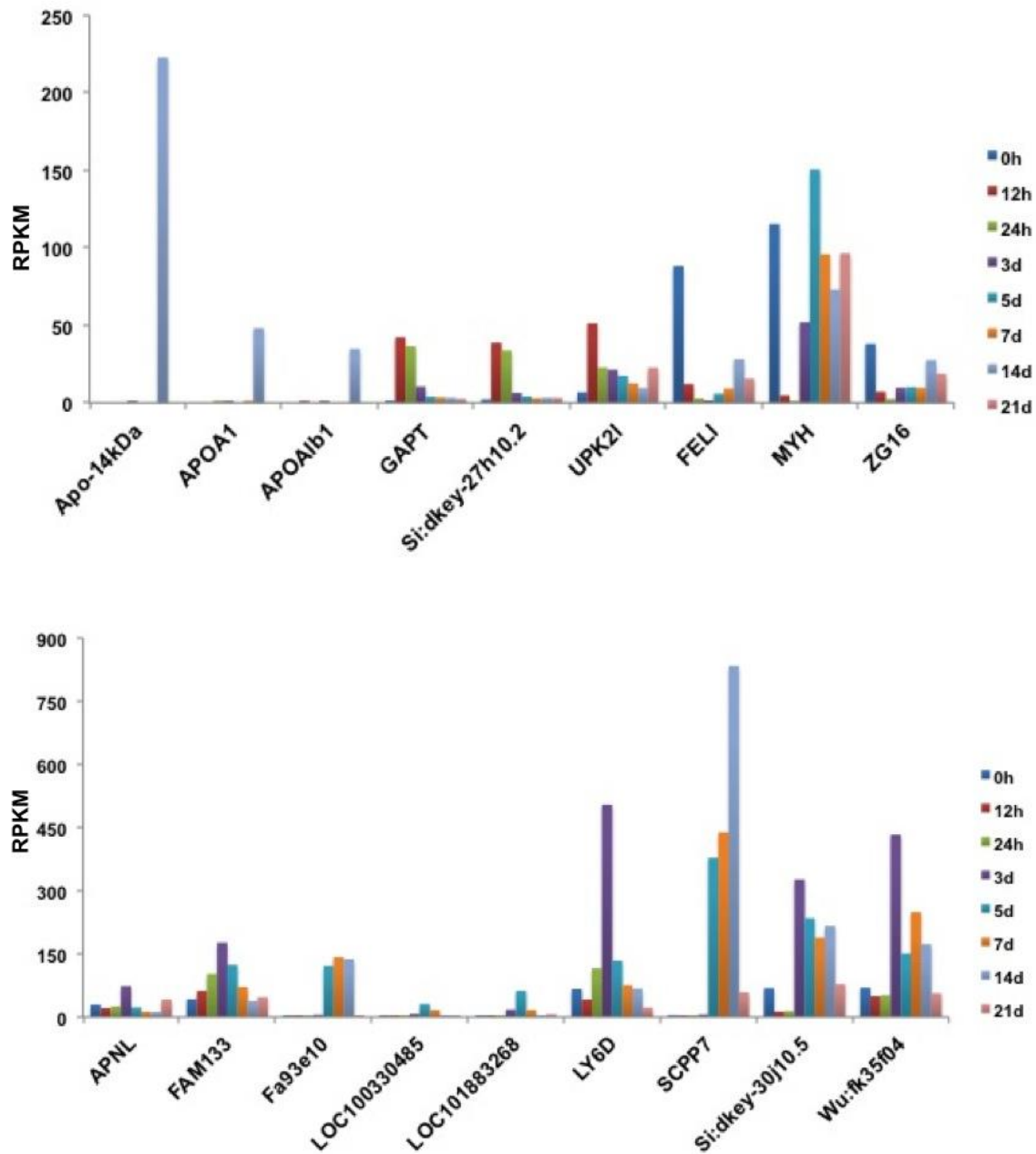


Figure 7. The cross intersection of 18 genes expression during the course of scale regeneration, with their expression levels expressed as reads per kilobase exon per million (RPKM) on the Y-axis, genes indicated on the X-axis, and time points expressed in bars of different colors. Gene abbreviations are the following: Apo-14kDa, 14 kDa apolipoprotein; APOA1, apolipoprotein A-I;

APOA1b1, apolipoprotein A1b1; APNL, actinoporin-like protein; LOC100330485, Uncharacterized protein LOC100330485; LOC101883268, Uncharacterized protein LOC101883268; LY6D, lymphocyte antigen 6D; FAM133, protein FAM133-like; GAPT, protein GAPT-like; Si:dkey-30j10.5, Uncharacterized protein si:dkey-30j10.5; Si:dkey-27h10.2, Uncharacterized protein si:dkey-27h10.2 isoform X1; SCPP7, secretory calcium-binding phosphoprotein 7; Fa93e10, si:dkey-22i16.3; UPK2l, uroplakin 2-like; FEL1, fish-egg lectin-like; MYH, myosin heavy chain, clone 203; Wu:fk35f04, Uncharacterized protein wu:fk35f04; and ZG16, zymogen granule membrane protein 16-like.

Comparative subtraction of the channel catfish skin transcriptome from the pleco skin transcriptome

Similarly, to further confirm the genomic caused for the scalelessness in catfish, the subtractive analysis of the scaled pleco skin and scaleless channel catfish skin transcriptome was conducted, which allowed identification of 704 genes that were expressed in the skin of scaled pleco but not in scaleless channel catfish (Table 16).

The 704 genes that were only expressed in the skin of pleco but not in the skin of channel catfish then were used to perform GO annotation by Blast2GO. As shown in Figure 8, GO term distribution revealed that cellular process (GO:0009987), binding (GO:0005488), and cell

(GO:0005623) was the most common annotation terms in Biological Process, Molecular Function, and Cellular Component GO term categories at level 2, respectively.

Of the shared 18 genes, that were only expressed in scaled carp skin but not in channel catfish skin and were differentially expressed genes during scale regeneration, three genes were expressed in scaled pleco skin as well, including lymphocyte antigen 6D, secretory calcium-binding phosphoprotein and uncharacterized protein si:dkey-30j10.5.

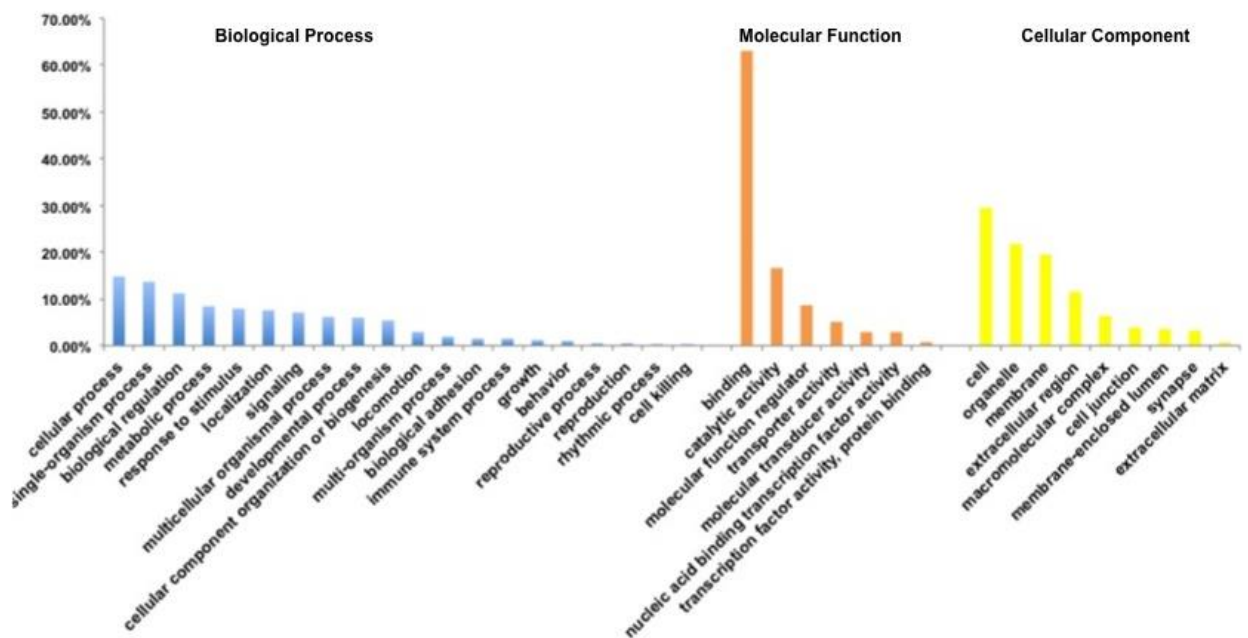


Figure 8. Gene ontology (GO) term categorization and distribution of genes expressed in the skin of scaled pleco but not in the skin of scaleless channel catfish. GO-terms were processed using Blast2GO and categorized at level 2 under biological process, molecular function and cellular component GO term categories, respectively.

Table 16. A list of 704 genes that are expressed in the skin of scaled pleco but not in scaleless channel catfish.

Pleco contig ID	Accession number	Gene description
c100842_g1_i1	BAA92751.1	Unnamed protein product [<i>Mus musculus</i>]
c101345_g1_i1	XP_006524143.1	Predicted: FLYWCH-type zinc finger-containing protein 1 isoform X3 [<i>Mus musculus</i>]
c101851_g1_i1	EMS47167.1	Hypothetical protein TRIUR3_03039 [<i>Triticum urartu</i>]
c101949_g1_i1	EDL22259.1	Host cell factor C1 regulator 1 (XPO1-dependent), isoform CRA_b [<i>Mus musculus</i>]
c102100_g1_i1	XP_006626032.1	Predicted: radial spoke head protein 3 homolog [<i>Lepisosteus oculatus</i>]
c102155_g1_i1	NP_001158131.1	TPR and ankyrin repeat-containing protein 1 [<i>Mus musculus</i>]
c102335_g1_i1	KFQ20252.1	Hypothetical protein N332_05522, partial [<i>Mesitornis unicolor</i>]
c10273_g1_i1	NP_001093106.1	PNMA-like protein 2 [<i>Mus musculus</i>]
c102937_g1_i1	S31035	Retrovirus-related gag polyprotein - mouse intracisternal A-particle MIAD8 (fragment)
c102963_g1_i1	XP_008980859.1	Predicted: uncharacterized protein LOC103788446 [<i>Callithrix jacchus</i>]
c103047_g1_i1	XP_007239478.1	Predicted: uncharacterized protein LOC103025867 [<i>Astyanax mexicanus</i>]
c103153_g1_i1	EDL09379.1	Mcg144592, partial [<i>Mus musculus</i>]
c104292_g1_i1	XP_002948511.1	Hypothetical protein VOLCADRAFT_88942 [<i>Volvox carteri f. Nagariensis</i>]
c104389_g1_i1	XP_006997987.1	Predicted: uncharacterized protein LOC284861-like [<i>Peromyscus maniculatus bairdii</i>]
c10439_g1_i2	EHB06074.1	Hypothetical protein GW7_15101, partial [<i>Heterocephalus glaber</i>]
c104410_g1_i1	XP_008424746.1	Predicted: uncharacterized protein LOC103475117 [<i>Poecilia reticulata</i>]
c104892_g1_i1	EDL24331.1	Immunity-related gtpase family, Q, isoform CRA_b, partial [<i>Mus musculus</i>]
c105124_g1_i1	EDM05893.1	Rcg33587 [<i>Rattus norvegicus</i>]
c105617_g1_i1	EAW61777.1	Hcg1986491 [<i>Homo sapiens</i>]
c105804_g1_i1	XP_002368828.1	RNA recognition motif-containing protein [<i>Toxoplasma gondii ME49</i>]
c10652_g1_i1	ETJ00816.1	Hypothetical protein Q611_LSC00001G0001, partial [<i>Leuconostoc sp. DORA_2</i>]
c107309_g1_i1	EDL27984.1	Mcg1040065 [<i>Mus musculus</i>]
c109020_g1_i1	EGW13716.1	Hypothetical protein I79_021101 [<i>Cricetulus griseus</i>]
c109343_g1_i1	AHB63507.1	Replicase polyprotein [<i>Bottlenose dolphin coronavirus HKU22</i>]
c109393_g1_i1	ERE87282.1	Putative overexpressed in colon carcinoma 1 protein like protein [<i>Cricetulus griseus</i>]
c109632_g1_i1	XP_002105689.1	GD16637 [<i>Drosophila simulans</i>]

c10967_g1_i1	CDQ67788.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c110344_g1_i1	ERE72751.1	A-kinase anchor protein 5 [<i>Cricetulus griseus</i>]
c11090_g1_i1	EGV91811.1	Hypothetical protein I79_020551 [<i>Cricetulus griseus</i>]
c111049_g1_i1	XP_001090400.2	Predicted: hypothetical protein LOC702119 [<i>Macaca mulatta</i>]
c111291_g1_i1	XP_007242826.1	Predicted: uncharacterized protein LOC103032647 [<i>Astyanax mexicanus</i>]
c111729_g1_i1	XP_007251440.1	Predicted: uncharacterized protein LOC103035242 [<i>Astyanax mexicanus</i>]
c112489_g1_i1	AAB31958.1	Antigen LEC-A [<i>Mus sp.</i>]
c112962_g1_i1	BAA95040.1	Unnamed protein product [<i>Mus musculus</i>]
c112978_g1_i1	EGW01874.1	Hypothetical protein I79_014262 [<i>Cricetulus griseus</i>]
c113195_g1_i1	CDQ78997.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c113211_g1_i1	XP_009050641.1	Hypothetical protein LOTGIDRAFT_81020, partial [<i>Lottia gigantea</i>]
c113240_g1_i1	GAA38700.1	Hypothetical protein CLF_101970 [<i>Clonorchis sinensis</i>]
c113896_g1_i1	BAB25056.1	Unnamed protein product [<i>Mus musculus</i>]
c113899_g1_i1	XP_006712960.1	Predicted: hyphal wall protein 2-like isoform X1 [<i>Homo sapiens</i>]
c114284_g1_i1	EDM00813.1	Rcg62717 [<i>Rattus norvegicus</i>]
c114674_g1_i1	EDL07130.1	Mcg113526, isoform CRA_c, partial [<i>Mus musculus</i>]
c114684_g1_i1	AAH51481.1	Smap2 protein [<i>Mus musculus</i>]
c114734_g1_i1	EGV91142.1	Hypothetical protein I79_026273 [<i>Cricetulus griseus</i>]
c115108_g1_i1	CAX69972.1	Hypothetical protein [<i>Schistosoma japonicum</i>]
c115207_g1_i1	EGL84524.1	LPXTG-motif cell wall anchor domain protein [<i>Streptococcus infantis SK1076</i>]
c115210_g1_i1	BAE23346.1	Unnamed protein product [<i>Mus musculus</i>]
c115252_g1_i1	EDM15703.1	Rcg59541 [<i>Rattus norvegicus</i>]
c115802_g1_i1	XP_007236727.1	Predicted: sperm-associated antigen 8 isoform X1 [<i>Astyanax mexicanus</i>]
c11587_g1_i1	EDL22605.1	Mcg21131, isoform CRA_c, partial [<i>Mus musculus</i>]
c116396_g1_i1	EDL85920.1	Rcg37312 [<i>Rattus norvegicus</i>]
c117375_g1_i1	EDL16410.1	Mcg144664, partial [<i>Mus musculus</i>]
c117946_g1_i1	XP_007236723.1	Predicted: uncharacterized protein LOC103043569 [<i>Astyanax mexicanus</i>]
c118724_g1_i1	XP_008606783.1	Hypothetical protein SDRG_02950 [<i>Saprolegnia diclina VS20</i>]
c11876_g2_i1	BAC33378.1	Unnamed protein product, partial [<i>Mus musculus</i>]

c119295_g1_i1	AAK12835.1	AF241244_1 cell adhesion protein trophinin [<i>Mus musculus</i>]
c119363_g1_i1	EGV91157.1	Hypothetical protein I79_026242 [<i>Cricetulus griseus</i>]
c119823_g1_i1	XP_009204301.1	Predicted: uncharacterized protein LOC103884011 [<i>Papio anubis</i>]
c119836_g1_i1	EDL31271.1	Cdna sequence BC047219, isoform CRA_b, partial [<i>Mus musculus</i>]
c120059_g1_i1	NP_061260.1	Activity-regulated cytoskeleton-associated protein [<i>Mus musculus</i>]
c121068_g1_i1	BAC29352.1	Unnamed protein product [<i>Mus musculus</i>]
c121212_g1_i1	EDL09503.1	Cdna sequence BC031181, isoform CRA_b, partial [<i>Mus musculus</i>]
c121363_g1_i1	XP_008591043.1	Predicted: uncharacterized protein LOC103608382 [<i>Galeopterus variegatus</i>]
c121444_g1_i1	CDQ75705.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c121963_g1_i1	XP_002564949.1	Pc22g09370 [<i>Penicillium rubens Wisconsin 54-1255</i>]
c122005_g1_i1	XP_002667467.3	Predicted: uncharacterized protein LOC100330690 [<i>Danio rerio</i>]
c122207_g1_i1	XP_001956624.1	GF24494 [<i>Drosophila ananassae</i>]
c122996_g1_i1	AAH13739.2	Slc7a5 protein [<i>Mus musculus</i>]
c123300_g1_i1	XP_002802861.1	Predicted: hypothetical protein LOC100424049 [<i>Macaca mulatta</i>]
c123663_g1_i1	XP_007254257.1	Predicted: olfactory receptor 2AK2-like [<i>Astyanax mexicanus</i>]
c123753_g1_i1	BAB27444.1	Unnamed protein product, partial [<i>Mus musculus</i>]
c123800_g1_i1	A0P9L2.1	Interleukin-1 receptor-associated kinase 1-binding protein 1 homolog
c124980_g1_i1	XP_670342.1	Hypothetical protein [<i>Plasmodium berghei</i> ANKA]
c125191_g1_i1	XP_007229630.1	Predicted: transmembrane protein 25-like [<i>Astyanax mexicanus</i>]
c125806_g1_i1	XP_007256000.1	Predicted: uncharacterized protein LOC103024271 [<i>Astyanax mexicanus</i>]
c126021_g1_i1	EDL34974.1	Mcg148190 [<i>Mus musculus</i>]
c126209_g1_i1	XP_007578992.1	Predicted: putative uncharacterized protein c14orf132 homolog [<i>Poecilia formosa</i>]
c126271_g1_i1	EDL76875.1	Rcg25468, isoform CRA_d [<i>Rattus norvegicus</i>]
c126742_g1_i1	XP_001625722.1	Predicted protein [<i>Nematostella vectensis</i>]
c126924_g1_i1	EDM16142.1	Rcg60275, isoform CRA_a [<i>Rattus norvegicus</i>]
c127208_g1_i1	WP_018397736.1	Hypothetical protein [<i>filamentous cyanobacterium ESFC-1</i>]
c127338_g1_i1	XP_001629387.1	Predicted protein [<i>Nematostella vectensis</i>]
c127355_g1_i1	XP_001307665.1	Hypothetical protein [<i>Trichomonas vaginalis</i> G3]
c127598_g1_i1	EGV99448.1	Hypothetical protein I79_014622 [<i>Cricetulus griseus</i>]

c127736_g1_i1	ELU12301.1	Hypothetical protein CAPTEDRAFT_185782, partial [<i>Capitella teleta</i>]
c127978_g1_i1	YP_007003918.1	Membrane protein Allo46 [<i>Cyprinid herpesvirus 2</i>]
c128056_g1_i1	CDQ99818.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c128076_g1_i1	BAB30652.1	Unnamed protein product [<i>Mus musculus</i>]
c128247_g1_i1	XP_007230779.1	Predicted: uncharacterized protein c14orf37 homolog isoform X3 [<i>Astyanax mexicanus</i>]
c128563_g1_i1	EFN82476.1	Hypothetical protein EAI_06335, partial [<i>Harpegnathos saltator</i>]
c128630_g1_i1	KEZ16769.1	Hypothetical protein CP98_03998 [<i>Sphingobium yanoikuyae</i>]
c12882_g1_i1	XP_009296749.1	Predicted: kinocilin isoform X1 [<i>Danio rerio</i>]
c129182_g1_i1	XP_002011705.1	GI11176 [<i>Drosophila mojavensis</i>]
c129359_g1_i1	NP_001073661.1	Otoconin-90 precursor [<i>Danio rerio</i>]
c129717_g1_i1	CDI87586.1	Hypothetical protein, conserved [<i>Eimeria praecox</i>]
c130145_g1_i1	WP_007448311.1	Aminoglycoside 3'-phosphotransferase [<i>Streptomyces coelicoflavus</i>]
c130229_g1_i1	XP_009296533.1	Predicted: uncharacterized protein LOC101885503 [<i>Danio rerio</i>]
c13038_g2_i1	EDL18693.1	GTL2, imprinted maternally expressed untranslated mrna [<i>Mus musculus</i>]
c130714_g1_i1	XP_004047720.1	Predicted: uncharacterized protein LOC101137755 [<i>Gorilla gorilla gorilla</i>]
c130876_g1_i1	XP_009298960.1	Predicted: uncharacterized protein c10orf107 homolog isoform X3 [<i>Danio rerio</i>]
c131640_g1_i1	BAE29142.1	Unnamed protein product, partial [<i>Mus musculus</i>]
c131762_g1_i1	XP_004048024.1	Predicted: uncharacterized protein LOC101132959 [<i>Gorilla gorilla gorilla</i>]
c131801_g1_i1	EDL31340.1	Mcg1034285 [<i>Mus musculus</i>]
c132217_g1_i1	EGW00777.1	Vomeronasal type-1 receptor 4 [<i>Cricetulus griseus</i>]
c132387_g1_i1	BAE27224.1	Unnamed protein product [<i>Mus musculus</i>]
c132468_g1_i1	AAA39469.1	Ly-6C [<i>Mus musculus domesticus</i>]
c133497_g1_i1	EDL40614.1	Mcg1041138 [<i>Mus musculus</i>]
c134241_g1_i1	EDL18721.1	Mcg18349 [<i>Mus musculus</i>]
c134353_g1_i1	NP_058664.2	Oligodendrocyte transcription factor 1 [<i>Mus musculus</i>]
c13463_g1_i1	XP_006511677.1	Predicted: uncharacterized protein LOC102639656 [<i>Mus musculus</i>]
c134810_g1_i1	EDL32729.1	Mcg140854, isoform CRA_b, partial [<i>Mus musculus</i>]
c134836_g1_i1	NP_001009264.1	MORF4 family-associated protein 1 [<i>Rattus norvegicus</i>]
c134911_g1_i1	XP_006522095.1	Predicted: transmembrane protein 191C isoform X2 [<i>Mus musculus</i>]

c134972_g1_i1	BAE90677.1	Unnamed protein product [<i>Macaca fascicularis</i>]
c135121_g1_i1	WP_019126563.1	Hypothetical protein [<i>Actinomyces sp.</i> Ph3]
c135130_g1_i1	YP_008378359.1	IE-2 [<i>Choristoneura rosaceana alphabaculovirus</i>]
c135468_g1_i1	EDL06275.1	Solute carrier family 32 (GABA vesicular transporter), member 1 [<i>Mus musculus</i>]
c135825_g1_i1	Q64253.2	Lymphocyte antigen 6E
c135862_g1_i1	XP_002806084.1	Predicted: hypothetical protein LOC100429674, partial [<i>Macaca mulatta</i>]
c135915_g1_i1	XP_008174928.1	Predicted: uncharacterized protein LOC101952564 [<i>Chrysemys picta bellii</i>]
c136547_g1_i1	BAC25675.1	Unnamed protein product [<i>Mus musculus</i>]
c136865_g1_i1	AAN28928.1	Mexneurin precursor [<i>Mus musculus</i>]
c137081_g1_i1	EDL05847.1	Ubiquitin-like 3, isoform CRA_a, partial [<i>Mus musculus</i>]
c137842_g1_i1	CDQ81692.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c138157_g1_i1	EDL20429.1	Mcg1033179 [<i>Mus musculus</i>]
c138366_g1_i1	EDL86253.1	Calcium/calmodulin-dependent protein kinase II gamma, isoform CRA_d [<i>Rattus norvegicus</i>]
c138960_g1_i1	BAC32997.1	Unnamed protein product, partial [<i>Mus musculus</i>]
c13899_g1_i1	XP_006847508.1	Hypothetical protein AMTR_s00014p00045960, partial [<i>Amborella trichopoda</i>]
c139013_g1_i1	XP_001620040.1	Hypothetical protein NEMVEDRAFT_v1g63584, partial [<i>Nematostella vectensis</i>]
c13906_g2_i1	EDM08879.1	Rcg43116 [<i>Rattus norvegicus</i>]
c139235_g1_i1	BAA87041.1	EIG 180 [<i>Mus musculus</i>]
c139480_g1_i1	AAR37049.1	Hypothetical polypeptide 22A [<i>Mesocricetus auratus</i>]
c1395_g1_i1	EDM08292.1	Calmodulin 3, isoform CRA_a [<i>Rattus norvegicus</i>]
c139504_g1_i1	CDQ79416.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c139514_g1_i1	BAE32365.1	Unnamed protein product [<i>Mus musculus</i>]
c14039_g1_i1	Q9UHZ2.1	Metastasis-associated lung adenocarcinoma transcript 1
c140502_g1_i1	EDK97438.1	Mcg142434 [<i>Mus musculus</i>]
c141223_g1_i1	XP_005164821.1	Predicted: protein CASC1 isoform X1 [<i>Danio rerio</i>]
c141755_g1_i1	BAB22836.1	Unnamed protein product [<i>Mus musculus</i>]
c141834_g1_i1	GAA31668.2	Innexin unc-9 [<i>Clonorchis sinensis</i>]
c141930_g1_i1	KFO33847.1	Hypothetical protein H920_04841 [<i>Fukomys damarensis</i>]
c142121_g1_i1	KFO62244.1	Uncharacterized protein c2orf16, partial [<i>Corvus brachyrhynchos</i>]

c142474_g1_i1	XP_006244946.1	Predicted: uncharacterized protein LOC100359890 [<i>Rattus norvegicus</i>]
c143261_g1_i1	AHB73987.1	1ab polyprotein [<i>Infectious bronchitis virus</i>]
c1435_g1_i2	AAI15083.1	Zgc:136320 protein [<i>Danio rerio</i>]
c143590_g1_i1	GAA52233.1	Hypothetical protein CLF_107652 [<i>Clonorchis sinensis</i>]
c143632_g1_i1	XP_007236181.1	Predicted: uncharacterized protein c1orf228 homolog isoform X2 [<i>Astyanax mexicanus</i>]
c143684_g1_i1	EDL21723.1	Oxysterol binding protein-like 8, isoform CRA_c [<i>Mus musculus</i>]
c143765_g1_i1	CAH93106.1	Hypothetical protein [<i>Pongo abelii</i>]
c145309_g1_i1	KFO26890.1	Hypothetical protein H920_11712 [<i>Fukomys damarensis</i>]
c145641_g1_i1	XP_007236491.1	Predicted: probable G-protein coupled receptor 151-like [<i>Astyanax mexicanus</i>]
c145990_g1_i1	BAE22222.1	Unnamed protein product, partial [<i>Mus musculus</i>]
c146086_g1_i1	EDL20374.1	Mcg140094, partial [<i>Mus musculus</i>]
c146148_g1_i1	EDL22314.1	Mcg1051026 [<i>Mus musculus</i>]
c146536_g1_i1	BAC31167.1	Unnamed protein product [<i>Mus musculus</i>]
c146541_g1_i1	EDM05482.1	Rcg33577 [<i>Rattus norvegicus</i>]
c146558_g1_i1	BAE37404.1	Unnamed protein product [<i>Mus musculus</i>]
c146644_g1_i1	EDL87674.1	Rcg41979 [<i>Rattus norvegicus</i>]
c14666_g2_i1	BAC38006.1	Unnamed protein product [<i>Mus musculus</i>]
c146895_g1_i1	XP_009304755.1	Predicted: calymmin isoform X6 [<i>Danio rerio</i>]
c147053_g1_i1	AFN88207.1	Integrase core domain containing protein [<i>Phaseolus vulgaris</i>]
c147467_g1_i1	CDQ62963.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c147548_g1_i1	BAC27069.1	Unnamed protein product [<i>Mus musculus</i>]
c147781_g1_i1	AAF22641.1	AF181560_1 prosaas [<i>Mus musculus</i>]
c147900_g1_i1	EST42909.1	Hypothetical protein SS50377_17443 [<i>Spironucleus salmonicida</i>]
c147952_g1_i1	WP_017318568.1	Hypothetical protein [<i>Mastigocladopsis repens</i>]
c148144_g1_i1	EDL94281.1	Rcg57788 [<i>Rattus norvegicus</i>]
c148402_g1_i1	CDQ98453.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c148722_g1_i1	WP_002969180.1	MULTISPECIES: hypothetical protein [<i>Brucella</i>]
c148789_g1_i1	EDL29164.1	Mcg1035439, partial [<i>Mus musculus</i>]
c149221_g1_i1	AID54952.1	Glyco-gag polyprotein [<i>Mus musculus</i>]

c149546_g1_i1	AAR87485.1	UGS148 protein [<i>Mus musculus</i>]
c149847_g1_i1	XP_007234500.1	Predicted: Ig-like V-type domain-containing protein FAM187A-like [<i>Astyanax mexicanus</i>]
c15023_g1_i1	XP_001343684.3	Predicted: interleukin-12 subunit alpha-like [<i>Danio rerio</i>]
c150298_g1_i1	NP_991140.1	Tuberoinfundibular peptide of 39 residues precursor [<i>Danio rerio</i>]
c150635_g1_i1	AAH58970.1	Gprasp1 protein [<i>Mus musculus</i>]
c151767_g1_i1	XP_008297596.1	Predicted: tetratricopeptide repeat protein 40 [<i>Stegastes partitus</i>]
c15194_g1_i1	CDQ57559.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c151962_g1_i1	CEF52190.1	Gram positive anchor [<i>Lactococcus garvieae</i>]
c152121_g1_i1	WP_012677560.1	Hypothetical protein [<i>Streptococcus equi</i>]
c152495_g1_i1	XP_009043616.1	Hypothetical protein LOTGIDRAFT_103053, partial [<i>Lottia gigantea</i>]
c152766_g1_i1	XP_008301310.1	Predicted: transmembrane protein 235 [<i>Stegastes partitus</i>]
c152795_g1_i1	AAI47313.1	Predicted gene, 547127 [<i>Mus musculus</i>]
c153871_g1_i1	EDM05076.1	Rcg33916 [<i>Rattus norvegicus</i>]
c154038_g1_i1	KDN85788.1	Hypothetical protein KCH_24460 [<i>Kitasatospora cheerisanensis KCTC 2395</i>]
c154143_g1_i1	XP_007525766.1	Predicted: uncharacterized protein LOC103115848 [<i>Erinaceus europaeus</i>]
c154150_g1_i1	EGV91155.1	Hypothetical protein I79_026246 [<i>Cricetulus griseus</i>]
c155109_g1_i1	NP_033241.1	Somatostatin precursor [<i>Mus musculus</i>]
c155420_g1_i1	WP_010925328.1	Lipoprotein VSAG (fragment), partial [<i>Mycoplasma pulmonis</i>]
c155620_g1_i1	WP_014292444.1	Hypothetical protein [<i>Oceanimonas</i> sp. <i>GKI</i>]
c155694_g1_i1	CDQ85331.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c156122_g1_i1	WP_015128893.1	Hypothetical protein [<i>Calothrix</i> sp. <i>PCC 7507</i>]
c156445_g1_i1	EDL93525.1	Rcg45444, partial [<i>Rattus norvegicus</i>]
c156543_g1_i1	NP_848486.1	Uncharacterized protein c8orf46 homolog [<i>Mus musculus</i>]
c156671_g1_i1	CCD59828.1	Hypothetical protein Smp_096370 [<i>Schistosoma mansoni</i>]
c156757_g1_i1	XP_006531633.1	Predicted: uncharacterized protein LOC102636946 [<i>Mus musculus</i>]
c156993_g1_i1	XP_002800796.1	Predicted: hypothetical protein LOC100430074 [<i>Macaca mulatta</i>]
c157055_g1_i1	EDL18070.1	Cell cycle exit and neuronal differentiation 1, isoform CRA_b, partial [<i>Mus musculus</i>]
c157616_g1_i1	CDJ47856.1	Hypothetical protein, conserved [<i>Eimeria brunetti</i>]
c157839_g1_i1	EDL89706.1	Rcg42628, isoform CRA_a [<i>Rattus norvegicus</i>]

c158076_g1_i1	EDL10557.1	Mcg1027127 [<i>Mus musculus</i>]
c158474_g1_i1	BAE30274.1	Unnamed protein product, partial [<i>Mus musculus</i>]
c158797_g1_i1	XP_004281246.1	Predicted: uncharacterized protein LOC101283539 [<i>Orcinus orca</i>]
c158812_g1_i1	XP_007258388.1	Predicted: nucleolin 2-like [<i>Astyanax mexicanus</i>]
c159198_g1_i1	EDL35214.1	Mcg126774 [<i>Mus musculus</i>]
c160009_g1_i1	AAL86907.1	AF481964_1 brain stress early protein [<i>Mus musculus</i>]
c160263_g1_i1	EDL08425.1	Mcg125297, partial [<i>Mus musculus</i>]
c160815_g1_i1	BAC35400.1	Unnamed protein product [<i>Mus musculus</i>]
c160926_g1_i1	BAE36617.1	Unnamed protein product, partial [<i>Mus musculus</i>]
c161262_g1_i1	KFK24463.1	Hypothetical protein AALP_aas61418u000100 [<i>Arabis alpina</i>]
c161276_g1_i1	XP_008995768.1	Predicted: uncharacterized protein LOC100407912 [<i>Callithrix jacchus</i>]
c16214_g1_i1	ACO13700.1	C9orf116 [<i>Esox lucius</i>]
c162422_g1_i1	EDL14934.1	Isoprenylcysteine carboxyl methyltransferase, isoform CRA_b, partial [<i>Mus musculus</i>]
c16267_g1_i1	XP_007228075.1	Predicted: bromodomain-containing protein DDB_G0270170-like [<i>Astyanax mexicanus</i>]
c163074_g1_i1	AAI17846.1	Usp54 protein [<i>Mus musculus</i>]
c163212_g1_i1	EDL41159.1	Mcg148421, partial [<i>Mus musculus</i>]
c16338_g1_i1	NP_001082963.1	2-amino-3-carboxymuconate-6-semialdehyde decarboxylase [<i>Danio rerio</i>]
c163513_g1_i1	CCD75416.1	Hypothetical protein Smp_161790.2 [<i>Schistosoma mansoni</i>]
c163671_g1_i1	XP_005101243.1	Predicted: protein rtoa-like [<i>Aplysia californica</i>]
c164373_g1_i1	ACM69338.1	Regulated endocrine-specific protein 18 isoform c [<i>Mus musculus</i>]
c165022_g1_i1	EDL32278.1	Mcg1034377, partial [<i>Mus musculus</i>]
c165186_g1_i1	WP_005934374.1	Hypothetical protein [<i>Xanthomonas axonopodis</i>]
c165363_g1_i1	BAB26293.1	Unnamed protein product [<i>Mus musculus</i>]
c165385_g1_i1	XP_008210265.1	Predicted: uncharacterized protein LOC103316743 [<i>Nasonia vitripennis</i>]
c165433_g1_i1	XP_002804589.1	Predicted: hypothetical protein LOC100423150 [<i>Macaca mulatta</i>]
c165759_g1_i1	XP_002739137.1	Predicted: uncharacterized protein LOC100373634 [<i>Saccoglossus kowalevskii</i>]
c16586_g1_i1	EDL17491.1	Mcg142724 [<i>Mus musculus</i>]
c166424_g1_i1	XP_007237819.1	Predicted: cysteine and tyrosine-rich protein 1 [<i>Astyanax mexicanus</i>]
c166878_g1_i1	EGW02204.1	Hypothetical protein I79_011772 [<i>Cricetulus griseus</i>]

c166973_g1_i1	XP_005470062.1	Predicted: uncharacterized protein LOC102077948 [<i>Oreochromis niloticus</i>]
c167077_g1_i1	NP_071872.1	Probable G-protein coupled receptor 88 [<i>Mus musculus</i>]
c167086_g1_i1	EAX08464.1	Hcg2019948, isoform CRA_b, partial [<i>Homo sapiens</i>]
c167274_g1_i1	XP_006815409.1	Predicted: uncharacterized protein LOC102808709 [<i>Saccoglossus kowalevskii</i>]
c167287_g1_i1	EDL22199.1	Mcg1048587 [<i>Mus musculus</i>]
c16752_g1_i2	XP_007253848.1	Predicted: testis-expressed sequence 36 protein [<i>Astyanax mexicanus</i>]
c167719_g1_i1	XP_007240067.1	Predicted: uncharacterized protein LOC103031164 [<i>Astyanax mexicanus</i>]
c167740_g1_i1	AAI09414.1	Hypothetical protein MGC35261-like (H. Sapiens) [<i>Danio rerio</i>]
c167799_g1_i1	XP_009305958.1	Predicted: uncharacterized protein LOC103912060 [<i>Danio rerio</i>]
c167846_g1_i1	BAB32273.1	Unnamed protein product [<i>Mus musculus</i>]
c168260_g1_i1	BAC35739.1	Unnamed protein product, partial [<i>Mus musculus</i>]
c168580_g1_i1	BAE25920.1	Unnamed protein product [<i>Mus musculus</i>]
c168646_g1_i1	NP_031720.1	Secretogranin-1 precursor [<i>Mus musculus</i>]
c169079_g1_i1	EDL13124.1	Cornichon homolog 3 (Drosophila), isoform CRA_b, partial [<i>Mus musculus</i>]
c169441_g1_i1	XP_004080414.1	Predicted: uncharacterized protein LOC101168157 [<i>Oryzias latipes</i>]
c169998_g1_i1	XP_008708269.1	Predicted: uncharacterized protein LOC103680539 [<i>Ursus maritimus</i>]
c170242_g1_i1	EDL76149.1	Bruno-like 4, RNA binding protein (Drosophila) (predicted) [<i>Rattus norvegicus</i>]
c170299_g1_i1	EDL41641.1	Mcg1045345 [<i>Mus musculus</i>]
c170645_g1_i1	BAB22608.1	Unnamed protein product [<i>Mus musculus</i>]
c170897_g1_i1	XP_001008643.1	Bifunctional endo-1,4-beta-xylanase xyla precursor, putative [<i>Tetrahymena thermophila</i>]
c170983_g1_i1	EGW07563.1	Hypothetical protein I79_016962 [<i>Cricetulus griseus</i>]
c171275_g1_i1	NP_001034474.1	VGF nerve growth factor inducible precursor [<i>Mus musculus</i>]
c171453_g1_i1	BAC38488.1	Unnamed protein product [<i>Mus musculus</i>]
c171718_g1_i1	WP_012420809.1	Von Willebrand factor type A [<i>Akkermansia muciniphila</i>]
c171773_g1_i1	XP_007238619.1	Predicted: uncharacterized protein c6orf118 homolog [<i>Astyanax mexicanus</i>]
c172142_g1_i1	XP_005526297.1	Predicted: histone-lysine N-methyltransferase set1-like [<i>Pseudopodoces humilis</i>]
c172212_g1_i1	BAE38023.1	Unnamed protein product [<i>Mus musculus</i>]
c172342_g1_i1	BAC86065.1	Unnamed protein product [<i>Homo sapiens</i>]
c172398_g1_i1	KGB36068.1	Hypothetical protein MS3_04346, partial [<i>Schistosoma haematobium</i>]

c173608_g1_i1	EDL15312.1	Mcg145956, partial [<i>Mus musculus</i>]
c174156_g1_i1	XP_007236442.1	Predicted: probable G-protein coupled receptor 148-like [<i>Astyanax mexicanus</i>]
c175264_g1_i1	EDL01298.1	Mcg142063, isoform CRA_a [<i>Mus musculus</i>]
c176055_g1_i1	EDL14526.1	Mcg126952, partial [<i>Mus musculus</i>]
c176730_g1_i1	EDL83939.1	Rcg40826 [<i>Rattus norvegicus</i>]
c176741_g1_i1	ADQ12919.1	Integrin-associated protein [<i>Mus musculus</i>]
c176978_g1_i1	XP_006542824.1	Predicted: protein FAM205A-like [<i>Mus musculus</i>]
c177000_g1_i1	EDL06499.1	Mcg141821, partial [<i>Mus musculus</i>]
c177863_g1_i1	XP_008328636.1	Predicted: pollen-specific leucine-rich repeat extensin-like protein 2 [<i>Cynoglossus semilaevis</i>]
c17830_g1_i1	EDL15560.1	Aldolase 3, C isoform, isoform CRA_a [<i>Mus musculus</i>]
c17857_g1_i1	XP_007237416.1	Predicted: lengsin [<i>Astyanax mexicanus</i>]
c178666_g1_i1	XP_003090393.1	Hypothetical protein CRE_29256, partial [<i>Caenorhabditis remanei</i>]
c178796_g1_i1	XP_005289250.2	Predicted: tumor necrosis factor receptor superfamily member 13B [<i>Chrysemys picta bellii</i>]
c178849_g1_i1	ETE63729.1	Octapeptide-repeat protein T2, partial [<i>Ophiophagus hannah</i>]
c179384_g1_i1	EDL40765.1	Mcg146145, partial [<i>Mus musculus</i>]
c180050_g1_i1	NP_001272069.1	Uncharacterized protein LOC101865950 [<i>Macaca fascicularis</i>]
c18015_g1_i1	EDM08300.1	Similar to MGC15476 protein (predicted) [<i>Rattus norvegicus</i>]
c180215_g1_i1	EGW14852.1	Hypothetical protein I79_021390 [<i>Cricetulus griseus</i>]
c180248_g1_i1	EDL78478.1	Rcg31638 [<i>Rattus norvegicus</i>]
c180455_g1_i1	EDM08398.1	Rcg24570 [<i>Rattus norvegicus</i>]
c18164_g1_i1	XP_668908.1	Hypothetical protein, partial [<i>Plasmodium berghei</i> ANKA]
c18420_g1_i1	NP_001164119.1	Transmembrane inner ear expressed protein isoform b [<i>Danio rerio</i>]
c18438_g1_i1	XP_001620448.1	Hypothetical protein NEMVEDRAFT_v1g223105 [<i>Nematostella vectensis</i>]
c18549_g1_i1	XP_007249304.1	Predicted: uncharacterized protein LOC103026616 isoform X1 [<i>Astyanax mexicanus</i>]
c18604_g1_i1	AAH16125.1	Napb protein [<i>Mus musculus</i>]
c19125_g2_i1	BAE33644.1	Unnamed protein product, partial [<i>Mus musculus</i>]
c19138_g1_i1	NP_001138708.1	Secretory calcium-binding phosphoprotein 5 precursor [<i>Danio rerio</i>]
c19813_g1_i1	WP_016089861.1	Hypothetical protein [<i>Bacillus cereus</i>]
c20217_g1_i1	CAA36968.1	Unnamed protein product (mitochondrion) [<i>Rattus norvegicus</i>]

c20515_g1_i1	XP_006249199.1	Predicted: neurosecretory protein VGF isoform X2 [<i>Rattus norvegicus</i>]
c20968_g1_i1	NP_808403.2	Transmembrane protein 130 precursor [<i>Mus musculus</i>]
c21321_g1_i1	EDL13269.1	Mcg147451 [<i>Mus musculus</i>]
c21321_g2_i1	BAB24979.3	Unnamed protein product, partial [<i>Mus musculus</i>]
c21326_g1_i1	CDQ57558.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c21326_g2_i1	ETE59542.1	Coiled-coil domain-containing protein, partial [<i>Ophiophagus hannah</i>]
c21355_g1_i1	XP_673201.1	Hypothetical protein [<i>Plasmodium berghei</i> ANKA]
c21590_g1_i2	CDQ85125.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c21653_g1_i1	XP_007233501.1	Predicted: serine/threonine-protein kinase 31 [<i>Astyanax mexicanus</i>]
c21662_g1_i1	CDQ74562.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c21662_g2_i1	XP_007241390.1	Predicted: mesogenin-1 [<i>Astyanax mexicanus</i>]
c21664_g1_i1	XP_007252550.1	Predicted: UPF0258 protein KIAA1024-like [<i>Astyanax mexicanus</i>]
c22188_g1_i1	NP_982973.2	ABR027Cp [<i>Ashbya gossypii</i> ATCC 10895]
c22942_g1_i1	KFO57522.1	Hypothetical protein N302_11013, partial [<i>Corvus brachyrhynchos</i>]
c22953_g2_i1	XP_001617828.1	Hypothetical protein NEMVEDRAFT_v1g156598 [<i>Nematostella vectensis</i>]
c23118_g1_i1	EDM06714.1	Rcg33941, partial [<i>Rattus norvegicus</i>]
c23259_g1_i1	BAC41425.1	Mkiaa0443 protein, partial [<i>Mus musculus</i>]
c23751_g1_i1	XP_008407658.1	Predicted: uncharacterized protein LOC103464984 isoform X1 [<i>Poecilia reticulata</i>]
c23784_g1_i1	XP_009048646.1	Hypothetical protein LOTGIDRAFT_112517, partial [<i>Lottia gigantea</i>]
c23790_g2_i1	EDL97533.1	Rcg27788, isoform CRA_a [<i>Rattus norvegicus</i>]
c2381_g1_i1	EDL89206.1	Guanine nucleotide binding protein, gamma 7, isoform CRA_a [<i>Rattus norvegicus</i>]
c24507_g1_i1	EDM13418.1	Williams-Beuren syndrome chromosome region 1 homolog (human) [<i>Rattus norvegicus</i>]
c25035_g1_i1	XP_009063978.1	Hypothetical protein LOTGIDRAFT_130569, partial [<i>Lottia gigantea</i>]
c25321_g1_i1	EMC81091.1	Hypothetical protein A306_11119, partial [<i>Columba livia</i>]
c25402_g1_i1	KFM05684.1	Hypothetical protein AS27_00155, partial [<i>Aptenodytes forsteri</i>]
c25577_g2_i1	XP_006630821.1	Predicted: coiled-coil domain-containing protein 147-like [<i>Lepisosteus oculatus</i>]
c25980_g2_i1	EDL21165.1	Mcg19512 [<i>Mus musculus</i>]
c26132_g1_i1	EGW09556.1	Hypothetical protein I79_011373 [<i>Cricetulus griseus</i>]
c26838_g1_i1	WP_020736887.1	Hypothetical protein [<i>Sorangium cellulosum</i>]

c27237_g1_i1	XP_005455846.1	Predicted: uncharacterized protein LOC102078211 [<i>Oreochromis niloticus</i>]
c2801_g1_i1	XP_002667158.2	Predicted: toll/interleukin-1 receptor domain-containing adapter protein [<i>Danio rerio</i>]
c28089_g1_i1	XP_007650256.1	Predicted: ly-6/neurotoxin-like protein 1 [<i>Cricetulus griseus</i>]
c28218_g1_i1	XP_007240066.1	Predicted: secreted phosphoprotein 24-like [<i>Astyanax mexicanus</i>]
c28486_g1_i1	BAE91017.1	Unnamed protein product [<i>Macaca fascicularis</i>]
c28940_g1_i1	EGW03010.1	Hypothetical protein I79_004323 [<i>Cricetulus griseus</i>]
c2943_g1_i1	BAC97929.1	Mkiaa0342 protein, partial [<i>Mus musculus</i>]
c29481_g1_i1	KEC60235.1	Hypothetical protein O95_00438 [<i>Bartonella henselae JK 53</i>]
c3028_g1_i1	BAA95107.1	Unnamed protein product [<i>Mus musculus</i>]
c30761_g1_i1	XP_007245300.1	Predicted: UPF0722 protein c11orf88 homolog isoform X2 [<i>Astyanax mexicanus</i>]
c31016_g1_i1	AAB71763.1	Fractalkine [<i>Mus musculus</i>]
c31016_g2_i1	BAC30544.1	Unnamed protein product [<i>Mus musculus</i>]
c31272_g1_i1	EPR63577.1	Hypothetical protein TGGT1_246190 [<i>Toxoplasma gondii</i> GT1]
c31421_g1_i1	XP_006638254.1	Predicted: coiled-coil domain-containing protein 103-like [<i>Lepisosteus oculatus</i>]
c31664_g1_i1	XP_007255219.1	Predicted: sterile alpha motif domain-containing protein 14-like [<i>Astyanax mexicanus</i>]
c31807_g1_i1	AIG55234.1	Prion protein, partial [<i>Mus fragilicauda</i>]
c32207_g1_i1	ELU08271.1	Hypothetical protein CAPTEDRAFT_220090 [<i>Capitella teleta</i>]
c32765_g1_i1	EHJ76642.1	Hypothetical protein KGM_07544 [<i>Danaus plexippus</i>]
c32972_g1_i1	XP_002804197.1	Predicted: hypothetical protein LOC100425763 [<i>Macaca mulatta</i>]
c33226_g1_i1	XP_001629001.1	Predicted protein, partial [<i>Nematostella vectensis</i>]
c33529_g2_i1	NP_573504.1	Membrane protein MLC1 [<i>Mus musculus</i>]
c34019_g1_i1	CAA43714.1	I54 protein [<i>Mus musculus</i>]
c34100_g1_i1	XP_007251617.1	Predicted: uncharacterized protein LOC103036982 [<i>Astyanax mexicanus</i>]
c34154_g2_i1	EMC89600.1	Hypothetical protein A306_01396, partial [<i>Columba livia</i>]
c34303_g1_i1	EGW11361.1	Hypothetical protein I79_012196 [<i>Cricetulus griseus</i>]
c34401_g1_i1	XP_007249303.1	Predicted: uncharacterized abhydrolase domain-containing protein DDB_G0269086-like isoform X2 [<i>Astyanax mexicanus</i>]
c34876_g1_i1	CAM76919.1	Hypothetical protein MGR_0775 [<i>Magnetospirillum gryphiswaldense MSR-1</i>]
c3492_g1_i1	CDI79912.1	Hypothetical protein EPH_0008640 [<i>Eimeria praecox</i>]

c3534_g1_i1	XP_007249308.1	Predicted: neurotensin/neuromedin N [<i>Astyanax mexicanus</i>]
c35518_g1_i1	XP_009052523.1	Hypothetical protein LOTGIDRAFT_143612, partial [<i>Lottia gigantea</i>]
c35701_g1_i1	EDM06855.1	Rho GDP dissociation inhibitor (GDI) alpha, isoform CRA_a [<i>Rattus norvegicus</i>]
c35840_g1_i1	BAB23350.1	Unnamed protein product [<i>Mus musculus</i>]
c35840_g2_i1	XP_003507946.2	Predicted: protein BEX3 isoform X1 [<i>Cricetulus griseus</i>]
c36026_g1_i1	NP_495202.1	Protein ZK84.1 [<i>Caenorhabditis elegans</i>]
c36131_g3_i1	EDL88262.1	Atpase, H ⁺ transporting, V0 subunit E isoform 2, isoform CRA_a [<i>Rattus norvegicus</i>]
c3665_g1_i1	EDL18692.1	Mcg147613 [<i>Mus musculus</i>]
c37691_g1_i1	CDX37445.1	Major spike protein G [<i>Mesorhizobium sp. ORS3359</i>]
c38032_g1_i1	NP_033078.2	Protein BEX1 [<i>Mus musculus</i>]
c38863_g1_i1	EDL25436.1	Neurogranin [<i>Mus musculus</i>]
c39140_g3_i1	NP_033408.1	Thy-1 membrane glycoprotein preproprotein [<i>Mus musculus</i>]
c39173_g2_i1	XP_005929403.1	Predicted: uncharacterized protein LOC102303347 [<i>Haplochromis burtoni</i>]
c39859_g1_i1	XP_001634591.1	Predicted protein, partial [<i>Nematostella vectensis</i>]
c39892_g3_i1	BAE37881.1	Unnamed protein product [<i>Mus musculus</i>]
c39896_g1_i1	XP_006225720.1	Predicted: putative uncharacterized protein YIL169C [<i>Rattus norvegicus</i>]
c40198_g1_i2	WP_029771141.1	Hypothetical protein, partial [<i>Neisseria meningitidis</i>]
c40198_g2_i1	WP_014106332.1	Hypothetical protein [<i>Komagataeibacter medellinensis</i>]
c40459_g1_i1	XP_001640838.1	Predicted protein, partial [<i>Nematostella vectensis</i>]
c40499_g1_i1	XP_678762.1	Pb-fam-2 protein, partial [<i>Plasmodium berghei ANKA</i>]
c40691_g2_i1	XP_004918284.1	Predicted: uncharacterized protein L138-like [<i>Xenopus (Silurana) tropicalis</i>]
c40929_g2_i1	EOA93074.1	Hypothetical protein Anapl_17963 [<i>Anas platyrhynchos</i>]
c41499_g2_i1	KGN92281.1	Hypothetical protein HQ43_09755 [<i>Porphyromonas canoris</i>]
c41507_g1_i1	BAE43189.1	Unnamed protein product, partial [<i>Mus musculus</i>]
c41633_g1_i1	EDL09486.1	Mcg147332 [<i>Mus musculus</i>]
c41856_g1_i1	XP_007891694.1	Predicted: uncharacterized protein c8orf48 homolog [<i>Callorhinchus milii</i>]
c41856_g1_i3	CDR01031.1	Unnamed protein product, partial [<i>Oncorhynchus mykiss</i>]
c42211_g1_i1	XP_006811164.1	Predicted: regenerating islet-derived protein 3-gamma-like [<i>Saccoglossus kowalevskii</i>]
c42543_g2_i1	WP_011961160.1	Hypothetical protein [<i>Psychrobacter sp. Prwf-1</i>]

c42881_g1_i1	EDL95834.1	Rcg58156, isoform CRA_a [<i>Rattus norvegicus</i>]
c43462_g1_i1	XP_009062279.1	Hypothetical protein LOTGIDRAFT_107337, partial [<i>Lottia gigantea</i>]
c44069_g1_i1	XP_001846843.1	Polyphosphate kinase 2 family [<i>Culex quinquefasciatus</i>]
c44644_g1_i1	XP_005463683.1	Predicted: uncharacterized protein LOC102077904 [<i>Oreochromis niloticus</i>]
c44673_g2_i1	CBH15423.1	Hypothetical protein, conserved [<i>Trypanosoma brucei gambiense DAL972</i>]
c45312_g1_i1	XP_001923919.1	Predicted: protein FAM216A [<i>Danio rerio</i>]
c45580_g1_i1	EDL23826.1	Mcg114170, isoform CRA_a, partial [<i>Mus musculus</i>]
c46090_g1_i1	Q6ZQX7.2	CQ097_HUMAN rename: Full=Uncharacterized protein c17orf97
c46331_g2_i1	XP_004212838.1	Predicted: uncharacterized protein LOC101234603 [<i>Hydra vulgaris</i>]
c46331_g3_i1	XP_008179116.1	Predicted: uncharacterized protein LOC100164825 [<i>Acyrtosiphon pisum</i>]
c46495_g2_i1	EDL18689.1	Mcg11475, isoform CRA_b, partial [<i>Mus musculus</i>]
c46627_g2_i2	XP_007259611.1	Predicted: trehalase, partial [<i>Astyanax mexicanus</i>]
c46848_g4_i1	XP_007235165.1	Predicted: myelin regulatory factor-like isoform X1 [<i>Astyanax mexicanus</i>]
c47100_g1_i1	XP_007260062.1	Predicted: taste receptor type 2 member 1-like [<i>Astyanax mexicanus</i>]
c47632_g1_i1	XP_001870401.1	Glycosyltransferase pgle [<i>Culex quinquefasciatus</i>]
c4809_g1_i1	XP_005168429.1	Predicted: uncharacterized protein si:dkeyp-118a3.2 [<i>Danio rerio</i>]
c48228_g1_i1	EDL89793.1	Rcg42708, isoform CRA_c [<i>Rattus norvegicus</i>]
c48276_g1_i1	KFO71552.1	Extracellular matrix protein A, partial [<i>Cuculus canorus</i>]
c48348_g1_i1	XP_002798603.1	Predicted: hypothetical protein LOC100425053 [<i>Macaca mulatta</i>]
c4870_g1_i1	XP_001446697.1	Hypothetical protein, partial [<i>Paramecium tetraurelia strain d4-2</i>]
c48961_g2_i1	XP_001635285.1	Predicted protein, partial [<i>Nematostella vectensis</i>]
c49051_g2_i1	WP_006070668.1	Type I secretion protein [<i>Vibrio shilonii</i>]
c49077_g2_i1	NP_035842.1	WW domain-binding protein 5 [<i>Mus musculus</i>]
c4921_g1_i1	XP_004035818.1	Hypothetical protein IMG5_087780, partial [<i>Ichthyophthirius multifiliis</i>]
c49662_g1_i1	EDL33905.1	Mcg146411 [<i>Mus musculus</i>]
c500_g1_i1	WP_024383107.1	Cell surface protein [<i>Streptococcus suis</i>]
c50706_g1_i1	XP_008281292.1	Predicted: uncharacterized protein LOC103358197 [<i>Stegastes partitus</i>]
c50706_g2_i1	XP_009299012.1	Predicted: uncharacterized protein LOC101883188 [<i>Danio rerio</i>]
c51235_g2_i1	XP_007250548.1	Predicted: uncharacterized protein c4orf22 homolog [<i>Astyanax mexicanus</i>]

c51307_g1_i1	XP_009262637.1	Hypothetical protein FPSE_11245 [<i>Fusarium pseudograminearum</i> CS3096]
c51319_g1_i1	XP_006306006.1	Hypothetical protein CARUB_v10011286mg [<i>Capsella rubella</i>]
c51430_g1_i1	XP_009293660.1	Predicted: uncharacterized protein LOC103909452 [<i>Danio rerio</i>]
c51618_g1_i2	XP_005020376.1	Predicted: UBX domain-containing protein 11-like isoform X3 [<i>Anas platyrhynchos</i>]
c51784_g1_i1	KFO62245.1	Hypothetical protein N302_04833, partial [<i>Corvus brachyrhynchos</i>]
c51920_g2_i1	EDL18901.1	Mcg147631 [<i>Mus musculus</i>]
c52208_g1_i1	XP_002999585.1	Hypothetical protein [<i>Candida glabrata</i> CBS 138]
c52866_g1_i1	CDI75002.1	Hypothetical protein EPH_0037840 [<i>Eimeria praecox</i>]
c52969_g1_i2	KDP25648.1	Hypothetical protein JCGZ_20804 [<i>Jatropha curcas</i>]
c52990_g3_i2	YP_007003691.1	Membrane protein ORF25 [<i>Cyprinid herpesvirus 1</i>]
c53157_g1_i1	XP_002609483.1	Hypothetical protein BRAFLDRAFT_132945 [<i>Branchiostoma floridae</i>]
c53264_g1_i2	CDQ61117.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c53298_g1_i1	XP_005470310.1	Predicted: uncharacterized protein LOC102077433 [<i>Oreochromis niloticus</i>]
c53731_g1_i1	CDR96068.1	Hypothetical protein, conserved [<i>Babesia bigemina</i>]
c53796_g1_i1	NP_082651.1	CAAX box 1 homolog C [<i>Mus musculus</i>]
c53796_g1_i2	NP_077132.1	Mammalian retrotransposon derived 8b [<i>Mus musculus</i>]
c53801_g1_i1	XP_008825423.1	Predicted: uncharacterized protein LOC103729211 [<i>Nannospalax galili</i>]
c54326_g1_i1	AFD52977.1	Proline and lysine rich protein [<i>Zea mays</i>]
c54368_g1_i2	XP_007253647.1	Predicted: uncharacterized protein LOC103026932 [<i>Astyanax mexicanus</i>]
c54404_g2_i2	XP_007243318.1	Predicted: synaptonemal complex central element protein 3-like [<i>Astyanax mexicanus</i>]
c54674_g1_i3	XP_009059173.1	Hypothetical protein LOTGIDRAFT_123884, partial [<i>Lottia gigantea</i>]
c54955_g1_i1	XP_001617493.1	Hypothetical protein NEMVEDRAFT_v1g8860, partial [<i>Nematostella vectensis</i>]
c54981_g2_i1	XP_001754371.1	Predicted protein [<i>Physcomitrella patens</i>]
c55194_g3_i2	XP_006808936.1	Predicted: uncharacterized protein LOC102799956 [<i>Neolamprologus brichardi</i>]
c55206_g1_i1	XP_005827594.1	Hypothetical protein GUIHDRAFT_113400 [<i>Guillardia theta</i> CCMP2712]
c55810_g1_i2	XP_007252884.1	Predicted: cholesterol desaturase daf-36-like [<i>Astyanax mexicanus</i>]
c55810_g2_i1	CDQ69765.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c55908_g1_i3	XP_007242938.1	Predicted: small acidic protein-like isoform X2 [<i>Astyanax mexicanus</i>]
c56921_g1_i1	EFN75656.1	Hypothetical protein EAI_04958, partial [<i>Harpegnathos saltator</i>]

c57151_g1_i1	CDJ33359.1	GD23040, related [<i>Eimeria mitis</i>]
c573_g1_i1	BAC25833.1	Unnamed protein product [<i>Mus musculus</i>]
c57509_g2_i2	CBH13412.1	Hypothetical protein, conserved [<i>Trypanosoma brucei gambiense DAL972</i>]
c57519_g1_i2	XP_005166153.2	Predicted: uncharacterized protein LOC101883682 isoform X2 [<i>Danio rerio</i>]
c57709_g2_i1	XP_009002532.1	Predicted: uncharacterized protein LOC103795335 isoform X2 [<i>Callithrix jacchus</i>]
c57712_g1_i2	NP_001138712.1	Secretory calcium-binding phosphoprotein 1 precursor [<i>Danio rerio</i>]
c57734_g1_i1	XP_007246986.1	Predicted: meiosis-specific nuclear structural protein 1-like [<i>Astyanax mexicanus</i>]
c581_g1_i1	XP_007244490.1	Predicted: proline-rich protein 7 [<i>Astyanax mexicanus</i>]
c58406_g1_i1	XP_007254534.1	Membrane protein FAM174B [<i>Astyanax mexicanus</i>]
c58494_g1_i2	XP_008772940.1	Uncharacterized protein LOC102557163 isoform X2 [<i>Rattus norvegicus</i>]
c58494_g3_i1	KFV92746.1	Hypothetical protein N326_06003, partial [<i>Eurypyga helias</i>]
c58655_g1_i1	WP_040060181.1	Hypothetical protein n2 [<i>Thalassospira xiamenensis</i>]
c58801_g1_i2	XP_007260272.1	Uncharacterized protein LOC103043683 [<i>Astyanax mexicanus</i>]
c59111_g2_i3	XP_009059264.1	Hypothetical protein LOTGIDRAFT_124069, partial [<i>Lottia gigantea</i>]
c59349_g1_i1	AAY17284.1	Osteocalcin [<i>Diplodus puntazzo</i>]
c59349_g2_i1	NP_001278818.1	Bone gamma-carboxyglutamate protein, like precursor [<i>Danio rerio</i>]
c59561_g1_i1	XP_008988974.1	Melanoma-associated antigen C1-like [<i>Callithrix jacchus</i>]
c59597_g1_i1	XP_728215.1	Hypothetical protein n1 [<i>Plasmodium yoelii yoelii 17XNL</i>]
c59633_g1_i2	EGD53615.1	Pentapeptide repeat-containing protein [<i>Gordonia neofelifaecis NRRL B-59395</i>]
c59633_g1_i4	EHK87819.1	Hypothetical protein SZMC14600_08544 [<i>Saccharomonospora azurea SZMC 14600</i>]
c59719_g1_i1	XP_005536257.1	Hypothetical protein CYME_CMI068C [<i>Cyanidioschyzon merolae strain 10D</i>]
c59833_g1_i1	XP_008298563.1	Kynureninase [<i>Stegastes partitus</i>]
c59886_g2_i1	XP_007235146.1	Transmembrane protein 138-like isoform X1 [<i>Astyanax mexicanus</i>]
c59889_g1_i1	WP_044375321.1	Hypothetical protein n5 [<i>Streptomyces ahygroscopicus</i>]
c6003_g1_i1	XP_006669639.1	MYB DNA binding protein (Tbf1), putative [<i>Cordyceps militaris CM01</i>]
c60220_g1_i1	KFY67239.1	Hypothetical protein V496_01678 [<i>Pseudogymnoascus pannorum VKM F-4515 (FW-2607)</i>]
c60331_g1_i1	XP_007256566.1	Uncharacterized protein c8orf88-like isoform X1 [<i>Astyanax mexicanus</i>]
c60581_g2_i4	XP_009065501.1	Hypothetical protein LOTGIDRAFT_168955, partial [<i>Lottia gigantea</i>]
c60619_g4_i1	XP_005185890.1	Extensin [<i>Musca domestica</i>]

c60987_g2_i1	WP_041895475.1	Hypothetical protein, partial [<i>Myroides profundus</i>]
c61447_g1_i3	KFW61670.1	Lwamide neuropeptides, partial [<i>Pygoscelis adeliae</i>]
c61919_g2_i1	KFD46202.1	Hypothetical protein M513_12902 [<i>Trichuris suis</i>]
c62070_g1_i1	XP_007240146.1	PDZK1-interacting protein 1 isoform X1 [<i>Astyanax mexicanus</i>]
c62166_g1_i1	XP_007231465.1	Transmembrane protein 119 [<i>Astyanax mexicanus</i>]
c62283_g1_i1	ACL34851.1	Hypothetical protein BGAPBR_D0006 (plasmid) [<i>Borrelia garinii PBr</i>]
c62507_g1_i2	XP_002942726.1	Uncharacterized protein LOC100495238 [<i>Xenopus (Silurana) tropicalis</i>]
c62734_g1_i1	XP_001683321.1	Hypothetical protein LMJF_22_1700 [<i>Leishmania major strain Friedlin</i>]
c62973_g1_i1	KFO60377.1	Hypothetical protein N302_03680, partial [<i>Corvus brachyrhynchos</i>]
c63055_g2_i1	XP_003174682.1	Stress protein DDR48 [<i>Microsporium gypseum CBS 118893</i>]
c63077_g1_i2	EFN73065.1	Hypothetical protein EAG_00196, partial [<i>Camponotus floridanus</i>]
c63203_g1_i1	XP_002664475.3	Uncharacterized protein urp1 [<i>Danio rerio</i>]
c63606_g1_i1	GAA47606.1	Hypothetical protein CLF_100574 [<i>Clonorchis sinensis</i>]
c63731_g1_i1	XP_001335256.1	Uncharacterized protein si:dkey-30j10.5 [<i>Danio rerio</i>]
c63851_g3_i3	XP_692475.3	Uncharacterized protein si:ch211-210c8.6 [<i>Danio rerio</i>]
c63851_g3_i4	AAI07829.1	Wu:fc47e12 protein, partial [<i>Danio rerio</i>]
c63935_g1_i2	XP_009293513.1	Uncharacterized protein LOC103909430 isoform X2 [<i>Danio rerio</i>]
c64103_g3_i1	WP_040524350.1	Hypothetical protein, partial [<i>Legionella shakespearei</i>]
c64103_g4_i1	XP_010737823.1	Elastin [<i>Larimichthys crocea</i>]
c64127_g1_i1	NP_001140037.1	Serum amyloid A-5 protein precursor [<i>Salmo salar</i>]
c64219_g1_i1	YP_005352862.1	Orf1ab gene product [<i>Night-heron coronavirus HKU19</i>]
c64321_g1_i1	YP_009118462.1	Nuclear antigen LANA-1 [<i>Equid herpesvirus 5</i>]
c64321_g2_i1	XP_002587125.1	Hypothetical protein BRAFLDRAFT_241896, partial [<i>Branchiostoma floridae</i>]
c64496_g2_i1	XP_005157804.1	Lymphocyte antigen 6D [<i>Danio rerio</i>]
c64744_g1_i3	XP_009523604.1	Hypothetical protein PHYSODRAFT_387807, partial [<i>Phytophthora sojae</i>]
c64744_g1_i4	CDM33296.1	Unnamed protein product [<i>Penicillium roqueforti FMI64</i>]
c64959_g3_i2	XP_007249134.1	Leukemia NUP98 fusion partner 1 homolog isoform X3 [<i>Astyanax mexicanus</i>]
c64980_g1_i1	ELK29144.1	Keratin-associated protein 10-8 [<i>Myotis davidii</i>]
c64980_g4_i2	KDB17756.1	Hypothetical protein UV8b_1640 [<i>Ustilaginoidea virens</i>]

c65062_g1_i2	XP_668908.1	Hypothetical protein, partial [<i>Plasmodium berghei</i> strain ANKA]
c65062_g2_i3	EDL91225.1	Rcg56442 [<i>Rattus norvegicus</i>]
c65227_g1_i2	CDQ56715.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c65227_g1_i6	XP_007235286.1	Indoleamine 2,3-dioxygenase 2-like [<i>Astyanax mexicanus</i>]
c65340_g3_i2	NP_001070103.2	Putative transferase CAF17 homolog, mitochondrial [<i>Danio rerio</i>]
c65400_g2_i1	XP_002582316.1	Predicted protein [<i>Uncinocarpus reesii</i> 1704]
c65539_g1_i2	AGZ01206.1	CPXV166 protein [<i>Cowpox virus</i>]
c65549_g1_i4	WP_030857228.1	Hypothetical protein [<i>Streptomyces</i> sp. NRRL S-37]
c65551_g1_i2	XP_002099036.1	GE10695 [<i>Drosophila yakuba</i>]
c65774_g1_i1	XP_001351429.1	Conserved Plasmodium protein, unknown function [<i>Plasmodium falciparum</i> 3D7]
c65899_g1_i1	XP_002675717.1	Predicted protein [<i>Naegleria gruberi</i>]
c65927_g13_i1	BAE34717.1	Unnamed protein product [<i>Mus musculus</i>]
c65927_g14_i1	NP_001013763.1	Dendrin [<i>Mus musculus</i>]
c65960_g4_i1	XP_002951091.1	Hypothetical protein VOLCADRAFT_104948 [<i>Volvox carteri</i> f. <i>Nagariensis</i>]
c66104_g2_i2	XP_007235695.1	Predicted: uncharacterized protein LOC103030816 [<i>Astyanax mexicanus</i>]
c66104_g2_i3	XP_007251070.1	Predicted: uncharacterized protein LOC103030037 isoform X1 [<i>Astyanax mexicanus</i>]
c66121_g8_i1	KGQ96612.1	Hypothetical protein MG1_03593 [<i>Candida albicans</i> GC75]
c66153_g1_i1	WP_001035349.1	Hypothetical protein [<i>Streptococcus pneumoniae</i>]
c66153_g2_i1	XP_505603.1	YALI0F19030p [<i>Yarrowia lipolytica</i>]
c66191_g8_i3	KFM59794.1	Hypothetical protein X975_02288, partial [<i>Stegodyphus mimosarum</i>]
c66316_g5_i1	XP_006815999.1	Predicted: endochitinase 2-like [<i>Saccoglossus kowalevskii</i>]
c66335_g2_i6	XP_007244226.1	Predicted: cyclin-D1-binding protein 1 homolog [<i>Astyanax mexicanus</i>]
c66438_g1_i11	XP_001238633.1	Hypothetical protein, conserved [<i>Eimeria tenella</i> strain Houghton]
c66457_g1_i2	KGQ89436.1	Hypothetical protein MEO_02301 [<i>Candida albicans</i> P94015]
c66495_g1_i1	XP_002088215.1	GE13692 [<i>Drosophila yakuba</i>]
c66495_g1_i4	XP_006045092.1	Predicted: P-selectin glycoprotein ligand 1 isoform X2 [<i>Bubalus bubalis</i>]
c66513_g5_i1	WP_011108098.1	Hypothetical protein [<i>Bacteroides thetaiotaomicron</i>]
c66780_g6_i1	XP_007251487.1	Predicted: prepro-urotensin II-beta-like [<i>Astyanax mexicanus</i>]
c66958_g2_i1	XP_002648999.1	Hypothetical protein CBG21336, partial [<i>Caenorhabditis briggsae</i>]

c67007_g1_i2	XP_007231307.1	Predicted: uncharacterized protein LOC103022990 [<i>Astyanax mexicanus</i>]
c67309_g2_i10	ETW11322.1	Peptidase S8 and S53 subtilisin kexin sedolisin [<i>Roseivivax atlanticus</i>]
c67479_g5_i2	XP_502130.1	YALI0C22297p [<i>Yarrowia lipolytica</i>]
c67501_g8_i1	AAX30961.2	SJCHGC09650 protein, partial [<i>Schistosoma japonicum</i>]
c67578_g1_i1	XP_001339495.3	Predicted: protein FAM118B isoform X1 [<i>Danio rerio</i>]
c67578_g1_i2	CDQ62658.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c67644_g4_i1	KFQ16078.1	Hypothetical protein N330_03420, partial [<i>Leptosomus discolor</i>]
c67702_g3_i1	XP_008302761.1	Predicted: tigger transposable element-derived protein 6-like [<i>Stegastes partitus</i>]
c67727_g1_i1	KFM68833.1	Hypothetical protein X975_07143, partial [<i>Stegodyphus mimosarum</i>]
c67951_g7_i1	XP_005515022.1	Predicted: S-antigen protein-like [<i>Columba livia</i>]
c67953_g1_i2	XP_007232846.1	Predicted: uncharacterized protein LOC103037313 [<i>Astyanax mexicanus</i>]
c68238_g3_i2	NP_001070846.1	Nuclear-interacting partner of ALK [<i>Danio rerio</i>]
c68406_g5_i2	ETL79070.1	Hypothetical protein L917_20226, partial [<i>Phytophthora parasitica</i>]
c68458_g1_i2	ELU06704.1	Hypothetical protein CAPTEDRAFT_96104 [<i>Capitella teleta</i>]
c68545_g3_i1	KFD68221.1	Hypothetical protein M514_19703 [<i>Trichuris suis</i>]
c68585_g9_i1	WP_032512849.1	Fap1 adhesin, partial [<i>Streptococcus suis</i>]
c68955_g1_i3	WP_031661350.1	Alanine and proline rich protein [<i>Mycobacterium tuberculosis</i>]
c69022_g8_i1	XP_007236431.1	Predicted: uncharacterized protein LOC103026859 [<i>Astyanax mexicanus</i>]
c69038_g9_i1	BAC25475.1	Unnamed protein product, partial [<i>Mus musculus</i>]
c69193_g2_i1	XP_006794397.1	Predicted: uncharacterized protein LOC102794807 [<i>Neolamprologus brichardi</i>]
c69193_g2_i2	XP_005452781.1	Predicted: uncharacterized protein LOC102077212 isoform X1 [<i>Oreochromis niloticus</i>]
c69279_g5_i6	XP_005463300.1	Predicted: coiled-coil domain-containing protein 169-like isoform X1 [<i>Oreochromis niloticus</i>]
c69301_g4_i2	XP_005164735.1	Predicted: A-kinase anchor protein 14 [<i>Danio rerio</i>]
c69312_g2_i1	CDQ92388.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c69573_g3_i2	XP_007248062.1	Predicted: TBC1 domain family member 19 [<i>Astyanax mexicanus</i>]
c69617_g5_i2	EMC85899.1	Hypothetical protein A306_05657 [<i>Columba livia</i>]
c69758_g1_i4	KFG64603.1	Hypothetical protein TGRUB_246190 [<i>Toxoplasma gondii</i> RUB]
c69852_g1_i2	KFO99229.1	Hypothetical protein N300_03429, partial [<i>Calypte anna</i>]
c69898_g6_i3	EQK98425.1	Hypothetical protein OCS_05863 [<i>Ophiocordyceps sinensis</i> CO18]

c69994_g8_i3	CDQ94673.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c70019_g5_i10	AAI71522.1	Si:ch211-51e12.4 protein [<i>Danio rerio</i>]
c70019_g5_i3	CDQ74559.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c70019_g5_i7	XP_007237838.1	Predicted: transmembrane protein 181-like [<i>Astyanax mexicanus</i>]
c70034_g10_i1	XP_001351573.1	Interspersed repeat antigen, putative [<i>Plasmodium falciparum 3D7</i>]
c70091_g7_i1	CDQ73868.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c70180_g3_i9	XP_002586172.1	Hypothetical protein BRAFLDRAFT_109674 [<i>Branchiostoma floridae</i>]
c70327_g4_i2	XP_001638575.1	Predicted protein, partial [<i>Nematostella vectensis</i>]
c70366_g7_i4	XP_706387.3	Predicted: uncharacterized protein si:dkeyp-69c1.9 [<i>Danio rerio</i>]
c70439_g2_i4	KFM68802.1	Hypothetical protein X975_05742, partial [<i>Stegodyphus mimosarum</i>]
c70497_g3_i3	WP_022080930.1	Wgs project CAEQ00000000 data annotated contig 1564 [<i>Ruminococcus sp. CAG:488</i>]
c70635_g1_i2	AHE54744.1	Hypothetical protein NX02_15305 [<i>Sphingomonas sanxanigenens</i> DSM 19645 = NX02]
c70635_g3_i5	XP_002262466.1	Hypothetical protein [<i>Plasmodium knowlesi</i> strain H]
c70635_g3_i6	WP_031876890.1	Clumping factor A, partial [<i>Staphylococcus aureus</i>]
c70635_g3_i9	XP_003286425.1	Hypothetical protein DICPUDRAFT_150384 [<i>Dictyostelium purpureum</i>]
c70736_g1_i1	EHY66668.1	Hypothetical protein NERG_00308 [<i>Nematocida sp. 1 ertm2</i>]
c70736_g2_i3	XP_007604311.1	Hypothetical protein VICG_00862 [<i>Vittaforma corneae ATCC 50505</i>]
c70938_g10_i3	XP_006498571.1	Predicted: keratin-associated protein 10-7-like isoform X1 [<i>Mus musculus</i>]
c70942_g6_i1	KFW78511.1	Hypothetical protein N305_08832, partial [<i>Manacus vitellinus</i>]
c71070_g1_i1	KGS84071.1	Hypothetical protein X947_3108 [<i>Burkholderia pseudomallei MSHR7334</i>]
c71072_g13_i3	ABG78036.1	Pbi-SS(Tom)(AP)51-EGFP [<i>synthetic construct</i>]
c71072_g6_i1	AIR10077.1	Putative cell-wall-anchored protein (LPXTG motif) [<i>Lactobacillus salivarius</i>]
c71072_g7_i3	XP_009229797.1	Hypothetical protein GGTG_13616 [<i>Gaeumannomyces graminis var. Tritici R3-111a-1</i>]
c71077_g2_i2	XP_007230370.1	Predicted: uncharacterized protein LOC103045361 [<i>Astyanax mexicanus</i>]
c71121_g6_i1	P01172.1	Somatostatin-2
c71217_g10_i1	EDL85728.1	Myeloid/lymphoid or mixed-lineage leukemia (trithorax homolog, <i>Drosophila</i>); translocated to, 11, isoform CRA_a [<i>Rattus norvegicus</i>]
c71256_g3_i1	ERL85978.1	Hypothetical protein D910_03392 [<i>Dendroctonus ponderosae</i>]
c71408_g1_i2	ETP72619.1	Hypothetical protein UYO_1395 [<i>Lachnospiraceae bacterium JC7</i>]

c71511_g2_i5	EWM24835.1	Phosphoesterase pa-phosphatase related protein [<i>Nannochloropsis gaditana</i>]
c7183_g1_i1	XP_007254859.1	Predicted: uncharacterized protein LOC103029740 [<i>Astyanax mexicanus</i>]
c71915_g4_i7	XP_001118777.1	Predicted: putative uncharacterized protein FLJ45035-like [<i>Macaca mulatta</i>]
c71915_g7_i1	CDU17396.1	Chitinase [<i>Plasmodium yoelii</i>]
c72042_g2_i3	ETJ17848.1	Hypothetical protein Q620_VSAC00650G0001, partial [<i>Veillonella sp.</i>]
c72106_g4_i2	XP_007239467.1	Predicted: IGF-like family receptor 1-like [<i>Astyanax mexicanus</i>]
c72131_g1_i1	EQB78249.1	Hypothetical protein CB1_000709011 [<i>Camelus ferus</i>]
c72141_g4_i1	XP_311477.4	AGAP010468-PA, partial [<i>Anopheles gambiae str. PEST</i>]
c72141_g4_i5	WP_013080559.1	Hemolysin-type calcium-binding region [<i>Caulobacter segnis</i>]
c72141_g5_i1	XP_002067064.1	GK24221 [<i>Drosophila willistoni</i>]
c72141_g5_i4	WP_008888391.1	Calcium binding hemolysin protein [<i>Thalassospira profundimaris</i>]
c72141_g5_i7	WP_022730875.1	Hypothetical protein [<i>Thalassospira lucentensis</i>]
c72291_g1_i1	XP_009050328.1	Hypothetical protein LOTGIDRAFT_78510, partial [<i>Lottia gigantea</i>]
c72392_g2_i1	KGO69473.1	Protein of unknown function DUF3533 [<i>Penicillium italicum</i>]
c72702_g4_i1	EHH51534.1	Hypothetical protein EGM_10925, partial [<i>Macaca fascicularis</i>]
c72744_g4_i1	XP_003686865.1	Hypothetical protein TPHA_0H02270 [<i>Tetrapisispora phaffii CBS 4417</i>]
c72754_g4_i1	KFM59159.1	Hypothetical protein X975_17642, partial [<i>Stegodyphus mimosarum</i>]
c72848_g5_i1	WP_005609620.1	Hemolysin-type calcium-binding repeat family protein [<i>Ruegeria mobilis</i>]
c72848_g5_i2	XP_002614831.1	Hypothetical protein CLUG_04845 [<i>Clavisporea lusitaniae ATCC 42720</i>]
c72956_g1_i1	XP_008333367.1	Predicted: uncharacterized protein LOC103396902, partial [<i>Cynoglossus semilaevis</i>]
c72994_g5_i2	EFZ17295.1	Hypothetical protein SINV_14571, partial [<i>Solenopsis invicta</i>]
c72994_g5_i5	XP_001326580.1	Viral A-type inclusion protein [<i>Trichomonas vaginalis G3</i>]
c72994_g5_i6	XP_001314181.1	Latent nuclear antigen [<i>Trichomonas vaginalis G3</i>]
c73042_g2_i3	XP_009049557.1	Hypothetical protein LOTGIDRAFT_238721 [<i>Lottia gigantea</i>]
c73053_g1_i1	WP_013539738.1	Yada domain-containing protein [<i>Variovorax paradoxus</i>]
c73170_g2_i1	KFV62840.1	Hypothetical protein N307_09981, partial [<i>Picooides pubescens</i>]
c73188_g5_i2	ELU00532.1	Hypothetical protein CAPTEDRAFT_95794 [<i>Capitella teleta</i>]
c73200_g1_i3	XP_706537.4	Predicted: uncharacterized protein LOC570021 [<i>Danio rerio</i>]
c73345_g2_i2	XP_007253957.1	Predicted: A-kinase anchor protein 12-like [<i>Astyanax mexicanus</i>]

c73417_g2_i1	XP_005474738.1	Predicted: uncharacterized protein LOC102082235 [<i>Oreochromis niloticus</i>]
c73474_g1_i5	XP_007542231.1	Predicted: uncharacterized protein LOC103130702 [<i>Poecilia formosa</i>]
c73502_g1_i1	CCE29783.1	Uncharacterized protein CPUR_03630 [<i>Claviceps purpurea 20.1</i>]
c73635_g2_i2	XP_005460634.1	Predicted: uncharacterized protein LOC102080148 isoform X1 [<i>Oreochromis niloticus</i>]
c73676_g3_i1	XP_007809759.1	Hypothetical protein MAC_03419 [<i>Metarhizium acridum cqma 102</i>]
c73676_g3_i2	WP_014934850.1	Collagen triple helix repeat-containing protein [<i>Cardinium endosymbiont of Encarsia pergandiella</i>]
c73676_g3_i3	XP_846278.1	Hypothetical protein [<i>Trypanosoma brucei brucei TREU927</i>]
c73751_g3_i1	XP_008768000.1	Predicted: uncharacterized protein LOC103693684 [<i>Rattus norvegicus</i>]
c73887_g3_i2	XP_001629254.1	Predicted protein, partial [<i>Nematostella vectensis</i>]
c7404_g1_i1	CAA35792.1	Unnamed protein product [<i>Mus musculus</i>]
c74243_g1_i1	EGV99860.1	Hypothetical protein I79_000131 [<i>Cricetulus griseus</i>]
c7436_g1_i1	WP_011464043.1	PE-PGRS family protein [<i>Rhodiferax ferrireducens</i>]
c7436_g1_i2	EEH04749.1	Proteoglycan [<i>Histoplasma capsulatum G186AR</i>]
c75309_g1_i1	WP_015828478.1	Hypothetical protein [<i>Hirschia baltica</i>]
c76384_g1_i1	EDL24583.1	Mcg130546, isoform CRA_b, partial [<i>Mus musculus</i>]
c76827_g1_i1	BAC30155.1	Unnamed protein product [<i>Mus musculus</i>]
c77205_g1_i1	WP_027663054.1	Hypothetical protein [<i>Escherichia coli</i>]
c773_g1_i1	NP_079911.1	ZW10 interactor [<i>Mus musculus</i>]
c77310_g1_i1	NP_001277950.1	Tyrosine-protein phosphatase non-receptor type substrate 1 isoform 4 precursor [<i>Mus musculus</i>]
c7733_g2_i1	AAF69645.1	AF119917_53 PRO2706 [<i>Homo sapiens</i>]
c78027_g1_i1	EGV92483.1	Hypothetical protein I79_011368 [<i>Cricetulus griseus</i>]
c78220_g1_i1	XP_005081165.1	Predicted: uncharacterized protein LOC101826732 [<i>Mesocricetus auratus</i>]
c78229_g1_i1	EDL06238.1	Neuronatin, isoform CRA_b, partial [<i>Mus musculus</i>]
c79694_g1_i1	EYB95389.1	Hypothetical protein Y032_0160g3330 [<i>Ancylostoma ceylanicum</i>]
c79714_g1_i1	XP_006817491.1	Predicted: uncharacterized protein LOC102809945 [<i>Saccoglossus kowalevskii</i>]
c80130_g1_i1	EDM16824.1	ATPase, Ca ⁺⁺ transporting, plasma membrane 1, isoform CRA_d [<i>Rattus norvegicus</i>]
c80812_g1_i1	XP_001631534.1	Predicted protein, partial [<i>Nematostella vectensis</i>]
c81352_g1_i1	EDL80381.1	Similar to microfilament and actin filament cross-linker protein isoform b [<i>Rattus norvegicus</i>]
c81462_g1_i1	WP_020607007.1	Hypothetical protein [<i>Spirosoma spitsbergense</i>]

c81712_g1_i1	XP_007261131.1	Predicted: uncharacterized protein LOC103023039 [<i>Astyanax mexicanus</i>]
c81731_g1_i1	EDL95259.1	Rcg58321 [<i>Rattus norvegicus</i>]
c81800_g1_i1	EGV92323.1	Hypothetical protein I79_017216 [<i>Cricetulus griseus</i>]
c82010_g1_i1	XP_005065564.1	Predicted: skin secretory protein xp2-like [<i>Mesocricetus auratus</i>]
c82266_g1_i1	BAC25645.1	Unnamed protein product [<i>Mus musculus</i>]
c82692_g1_i1	EDL25637.1	Sodium channel, type IV, beta, partial [<i>Mus musculus</i>]
c8289_g1_i1	KGB32951.1	Universal stress protein [<i>Schistosoma haematobium</i>]
c8289_g2_i1	EUB58410.1	Hypothetical protein EGR_06697 [<i>Echinococcus granulosus</i>]
c83063_g1_i1	EDL99399.1	Serine/arginine-rich protein specific kinase 2 (predicted), isoform CRA_d [<i>Rattus norvegicus</i>]
c83155_g1_i1	XP_007228416.1	Predicted: RIB43A-like with coiled-coils protein 2-like [<i>Astyanax mexicanus</i>]
c8351_g1_i1	NP_001289983.1	Small integral membrane protein 18 [<i>Danio rerio</i>]
c83576_g1_i1	XP_005168232.2	Predicted: uncharacterized protein LOC100329771 [<i>Danio rerio</i>]
c83685_g1_i1	XP_004088922.1	Predicted: uncharacterized protein LOC101178713 [<i>Nomascus leucogenys</i>]
c83838_g1_i1	XP_009164084.1	Hypothetical protein T265_01730 [<i>Opisthorchis viverrini</i>]
c84084_g1_i1	KFW72756.1	Hypothetical protein AS28_15112, partial [<i>Pygoscelis adeliae</i>]
c843_g1_i1	WP_011456798.1	Hemolysin-type calcium-binding protein [<i>Jannaschia sp. CCS1</i>]
c84397_g1_i1	AAH27781.1	Srrm2 protein [<i>Mus musculus</i>]
c85197_g1_i1	CDQ71087.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c85337_g1_i1	XP_007249628.1	Predicted: protein phosphatase 1 regulatory subunit 17 [<i>Astyanax mexicanus</i>]
c85658_g1_i1	EGW02335.1	Hypothetical protein I79_006438 [<i>Cricetulus griseus</i>]
c85730_g1_i1	AAH24457.1	D2Ert391e protein [<i>Mus musculus</i>]
c86382_g1_i1	WP_017138397.1	Hypothetical protein [<i>Pseudomonas fluorescens</i>]
c86398_g1_i1	EGW07852.1	Hypothetical protein I79_010928 [<i>Cricetulus griseus</i>]
c86418_g1_i1	CDQ62141.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c86555_g1_i1	XP_006544771.1	Predicted: uncharacterized protein LOC102642291 isoform X1 [<i>Mus musculus</i>]
c87153_g1_i1	XP_002614519.1	Hypothetical protein CLUG_05296 [<i>Clavisporea lusitaniae ATCC 42720</i>]
c87171_g1_i1	NP_001153828.1	Ubiquinol-cytochrome-c reductase complex assembly factor 3 precursor [<i>Mus musculus</i>]
c87261_g1_i1	CDY48599.1	Bnaa01g07660d [<i>Brassica napus</i>]
c87560_g1_i1	CAX73426.1	Hypothetical protein [<i>Schistosoma japonicum</i>]

c88638_g1_i1	XP_009166169.1	Hypothetical protein T265_03421 [<i>Opisthorchis viverrini</i>]
c88674_g1_i1	XP_006537881.1	Predicted: protein KIAA1045 isoform X10 [<i>Mus musculus</i>]
c889_g1_i1	XP_003587737.3	Predicted: uncharacterized protein c6orf132 homolog isoform X2 [<i>Bos taurus</i>]
c89191_g1_i1	XP_007245805.1	Predicted: stathmin domain-containing protein 1 [<i>Astyanax mexicanus</i>]
c89723_g1_i1	BAE32632.1	Unnamed protein product [<i>Mus musculus</i>]
c90334_g1_i1	EDL14314.1	Mcg145223, partial [<i>Mus musculus</i>]
c90724_g1_i1	AGO28203.1	Agtw-amide [<i>Aurelia aurita</i>]
c90833_g1_i1	EDL29460.1	Lymphocyte antigen 6 complex, locus E, isoform CRA_b [<i>Mus musculus</i>]
c9087_g1_i1	AAH24711.1	D3Bwg0562e protein, partial [<i>Mus musculus</i>]
c90902_g1_i1	ERE80872.1	Hypothetical protein H671_3g8557 [<i>Cricetulus griseus</i>]
c91162_g1_i1	KFZ09592.1	Hypothetical protein V501_05529 [<i>Pseudogymnoascus pannorum VKM F-4519 (FW-2642)</i>]
c91384_g1_i1	XP_009305084.1	Predicted: E3 ubiquitin-protein ligase RBBP6-like [<i>Danio rerio</i>]
c91720_g1_i1	CAJ57446.1	TPA: interleukin-11, partial [<i>Ictalurus punctatus</i>]
c91784_g1_i1	WP_029470441.1	Transporter [<i>Clostridiales bacterium</i> VE202-06]
c91817_g1_i1	AAH09094.1	Cd47 protein, partial [<i>Mus musculus</i>]
c92071_g1_i1	CAA43690.1	E161 [<i>Mus musculus</i>]
c92323_g1_i1	XP_007231496.1	Predicted: PIH1 domain-containing protein 2-like [<i>Astyanax mexicanus</i>]
c92330_g1_i1	XP_007253378.1	Predicted: coiled-coil domain-containing protein 37 [<i>Astyanax mexicanus</i>]
c92822_g1_i1	EDL08413.1	Mcg145104, partial [<i>Mus musculus</i>]
c92932_g1_i1	XP_007245798.1	Predicted: alanine and arginine-rich domain-containing protein-like [<i>Astyanax mexicanus</i>]
c93096_g1_i1	EDL96641.1	RGD1563912 (predicted), isoform CRA_a [<i>Rattus norvegicus</i>]
c93457_g1_i1	EGV98724.1	Hypothetical protein I79_013418 [<i>Cricetulus griseus</i>]
c94060_g1_i1	ELT90326.1	Hypothetical protein CAPTEDRAFT_133476 [<i>Capitella teleta</i>]
c94135_g1_i1	EHH50940.1	Hypothetical protein EGM_10246, partial [<i>Macaca fascicularis</i>]
c94711_g1_i1	NP_001270376.1	Uncharacterized protein LOC101866292 [<i>Macaca fascicularis</i>]
c94907_g1_i1	EGW11848.1	Hypothetical protein I79_017120 [<i>Cricetulus griseus</i>]
c95166_g1_i1	XP_007571278.1	Predicted: adherens junction-associated protein 1 isoform X1 [<i>Poecilia formosa</i>]
c95285_g1_i1	EFN73558.1	Hypothetical protein EAG_11370, partial [<i>Camponotus floridanus</i>]
c96326_g1_i1	EDK99933.1	Gamma-aminobutyric acid (GABA(A)) receptor-associated protein-like 1 [<i>Mus musculus</i>]

c96457_g1_i1	XP_003795420.1	Predicted: Duffy antigen/chemokine receptor [<i>Otolemur garnettii</i>]
c96744_g1_i1	BAE22816.1	Unnamed protein product, partial [<i>Mus musculus</i>]
c96882_g1_i1	CCQ43166.1	Alternative protein GNB1 [<i>Homo sapiens</i>]
c97595_g1_i1	XP_006505091.1	Predicted: protachykinin-1 isoform X2 [<i>Mus musculus</i>]
c97899_g1_i1	CDQ62288.1	Unnamed protein product [<i>Oncorhynchus mykiss</i>]
c98285_g1_i1	XP_006525312.1	Predicted: uncharacterized protein LOC102640032 [<i>Mus musculus</i>]
c98537_g1_i1	EDL86229.1	Protein phosphatase 3, catalytic subunit, beta isoform, isoform CRA_a [<i>Rattus norvegicus</i>]
c98866_g1_i1	EDL81932.1	Rcg28534 [<i>Rattus norvegicus</i>]
c9888_g1_i1	BAG57046.1	Unnamed protein product [<i>Homo sapiens</i>]
c99070_g1_i1	XP_004073053.1	Predicted: putative proline-rich protein 21-like [<i>Oryzias latipes</i>]
c99467_g1_i1	EDL12498.1	Mcg1036263 [<i>Mus musculus</i>]
c99606_g1_i1	XP_007247951.1	Predicted: protein phosphatase 1 regulatory subunit 36-like [<i>Astyanax mexicanus</i>]
c99647_g1_i1	KES06986.1	Nicotinate-nucleotide--dimethylbenzimidazole phosphoribosyltransferase [<i>Streptomyces toyocaensis</i>]

CHAPTER V. DISCUSSIONS

Biological characteristics or phenotypes are controlled by three factors: genes, environment, and gene-environment interactions. The control of traits by genes, are governed by the composition of the genome, and complex sets of regulatory mechanisms leading to the eventual “expression” of the genetic material. The traditional central dogma of genetics from DNA to RNA through transcription, and from RNA to proteins through translation, is apparently becoming obsolete, with the rapid discoveries of many levels of regulation of gene expression. At the genomic level, alleles may make a difference at each locus and the combination of all the haplotypes makes the genome composition. The genome must be transcribed into RNA to have its biological functions realized. The entire RNA of all cells and tissues of an organism makes up its transcriptome that are dynamic at any given moment with variable development stages, physiological conditions, and environment. Once again, the combination of the quality and quantities of expressed transcriptome as a whole determines the phenotypes. The transcription and the transcribed RNAs are under various layers of regulation, including being regulated by the transcripts themselves such as microRNA or long non-coding RNAs. Therefore, characterization of transcriptome is crucially important for the understanding of function and phenotypes.

The primary goal of my dissertation was to generate and assemble the transcriptome of channel catfish, common carp, and pleco for comparative transcriptome analysis. This goal was reached efficiently through the application of high throughput RNA-Seq. In my study, scaleless channel catfish skin transcriptome, scaled common carp skin transcriptome and scaled pleco skin transcriptome were sequenced, assembled and annotated. A total of 20,474 unique genes, 25,136 unique genes, and 21,105 unique genes were identified, respectively. It is unknown at present as to if the entire transcriptome is revealed by the RNA-Seq analysis, but it is apparent that RNA-Seq has the deep sequencing coverage, allowing the most of the transcripts to be sequenced.

The second level of goals of my dissertation is to conduct comparative subtraction using transcriptome from species with contrasted phenotypes of interest. One of the most significant characteristics of catfishes is scaleless, and this feature is shared by most catfish species. However, some catfish species are covered by bony dermal plates, such as *Loricaridae* (pleco). It is of great interest, not only for evolutionary biologist, but also for aquaculturist, to understand what the genomic causes are for the evolutionary lack of scales, and how the scaleless skin is adapted to the environmental changes to be able to resist against various disease infections without physical barrier of scales as a mechanism of protection.

Several possibilities may cause the loss of scales, including: 1) loss or mutations in genes known to control scale formation; 2) loss or mutations in unknown genes or gene pathways controlling scale formation; 3) genes are present but not expressed in the skin that would cause

the lack of scales. A number of genes have been reported to be involved in scale formation or regeneration in previous studies. For example, mutations in ectodysplasia A (EDA) and ectodysplasia A receptor (EDAR) lead to the loss or abnormal scales in medaka and zebrafish (Kondo et al., 2001; Harris et al., 2008; Lida et al., 2014), as well as failures for the development of teeth and hair in human and mice. Mutations in lymphocyte enhancer factor-1 (LEF1) cause the loss of caudal fin and defect in fin and tooth formation (McGraw et al., 2011). Mutations in fibroblast growth factors receptors (FGFR) caused scaleless in zebrafish (Kolanczyk, 2009; Rohner et al., 2009). In order to check the first possibility, EDA, EDAR, LEF1 and FGFR genes were first examined. However, analysis of open reading frames of these genes and expression profiles in skin transcriptome indicated that these genes are present in channel catfish, and are likely functional because they encode very similar amino acid sequences as compared to those of zebrafish, suggesting their irrelevance to lack of scales in channel catfish.

Direct functional analysis is essentially impossible with channel catfish at the genome scale. However, expression analysis often is very insightful in relation to function. In order to determine what genes are involved in scale formation, differentially expressed genes were first identified during carp scale regeneration. The notion is that if a gene is important for scale formation, it should be differentially expressed for scale regeneration. A total of 1,173 genes were found significantly differentially expressed during scale regeneration. The number of differentially expressed genes was generally decreasing with time after scale removal (from 638 genes at 12 h time point to 213 genes at 21 days time point), indicating more biological events

were involved in early stages after scale removal. Specifically, immediately after scale removal, it is expected that wound healing and related immune responses would be involved, and an examination of the induced gene list certainly support that. For instance, a large number of chemokines were significantly induced at early stages of scale regeneration. Although the analysis of differentially expressed genes during scale regeneration was useful, itself did not allow it to pin down the list of genes.

We further reasoned that the genes important for scale formation may not be expressed in the scaleless channel catfish. Interspecific comparative transcriptome analysis between common carp skin and channel catfish skin were conducted and a total of 836 genes were identified, which were expressed in the common carp skin, but not in the channel catfish skin. Functional annotation of these genes indicated that genes were involved in biological regulation, signaling, metabolic, developmental, and immune process and other biological process, providing an important candidate gene pool for scale formation. Once again, the ration was straightforward, but the interspecific transcriptome analysis did not allow me to narrow the list of candidate genes either.

Along with the candidate gene pool that are only expressed in scaled carp but not in scaleless channel catfish, the differentially expressed genes during scale regeneration in common carp allowed me to narrow the list of candidate genes for scale formation. The cross intersection of these two gene lists significantly reduced because the genes must satisfy two conditions: 1) they are only expressed in the scaled carp skin but not in the scaleless catfish, and 2) they must

be expressed differentially during scale regeneration. These conditions should be met if the genes are important for scale formation.

Using the strategies developed in this study, a total of 18 genes were identified that both are only expressed in scaled carp skin but not in scaleless channel catfish skin and differentially expressed during carp scale regeneration. Of the 13 known genes, 10 were up-regulated. They were 14 kDa apolipoprotein, actinoporin-like protein, apolipoprotein A-I, apolipoprotein AIb1, lymphocyte antigen 6D, protein FAM133-like, protein GAPT-like, secretory calcium-binding phosphoprotein 7, si:dkey-22i16.3 (fa93e10), and uroplakin 2-like. The functional analysis of some of these proteins were already known from fish or mammal, but many were novel. Of these, protein GAPT and uroplakin2-like gene were functionally related to immune response. Protein GAPT was highly differentially expressed after carp scale removal at early stage, with 32.17 fold at 12 hours and 27.46 fold at 24 hours up-regulation, which was reported that GAPT was expressed in B cells and meloid cells, regulating B cell proliferation involved in immune response (Liu and Zhang, 2008). Uroplakin 2 is up-regulated in response to inflammation (Saban et al., 2007). In addition, uroplakins play a fundamental role in wound healing (Sun, 2006), bacterial infection (Xie et al., 2006). Here, uroplakin 2-like gene was significantly up-regulated after scale removal at 12 hours, with 7.69 fold, indicated it might play an crucial role in response to wound healing or inflammation at the early stage after scale removal.

Actinoporin-like protein was up-regulated at 3 days after scale removal with 2.38 fold, which is functioned as predation, denfence, and digestion in sea anemone (Garc á-Ortega et al.,

2011) and play important role in pore complex assembly. However, there are no publications have indicated that this gene is related with bone, scale or fin formation. Lymphocyte antigen 6D (Ly-6D) has been demonstrated that is expressed exclusively at the outer cell surface of transitional epithelium and keratinocyte of stratified squamous epithelia in normal tissues (Quak et al., 1990). The wounds were generated after fish scale losing, and the wound would heal in the epidermis first before new scale rebuilt. During this process, keratinocytes migrated faster toward the center of the wound (Graham et al., 2013). The expression level of Ly-6D was up-regulated after carp scale removal at 24 hours, and differentially expressed at 3 days, with induced 7.24 fold, suggesting lymphocyte antigen 6D may play an important role during the process of scale regeneration, especially for the early stages. More strikingly, si:dkey-22i16.3 (fa93e10) was highly up-regulated over a thousand-fold from 5 days to 14 days after scale removal, and secretory calcium-binding phosphoprotein 7 gene was dramatically induced beginning 5 days through 21 days after scale removal (75-fold at 5 days, 102-fold at 7 days, and 186-fold at 14 days, and 12.96-fold at 21 days increasing, respectively). Secretory calcium-binding phosphoprotein gene, which initially arose from SPARC, are well known to be involved in tissue mineralization of bone, skeletons, enameloid and dentin in osteichthyans (Kawasaki and Amemiya, 2014; Kawasaki et al., 2007; Kawasaki and Weiss, 2006). Interestingly, three apolipoprotein genes including 14 kDa apolipoprotein, apolipoprotein A-I, apolipoprotein AIIb1 were differentially expressed at 14 days after scale removal. It is reported that partial mineralization of the basal plate occurred at days 14-28 (Ohira et al., 2007).

Apolipoproteins promote fat efflux (Steinmetz et al., 1990); they were reported to be involved in the process of fin regeneration and scale development in zebrafish, suggesting their roles in development of mineralized tissues (Monnot et al., 1999; Sire and Akimenko, 2004).

Although direct analysis of scale regeneration was not feasible with pleco, analysis of its skin transcriptome provided insights for narrowing down the candidate gene pool of scale formation. The cross intersection between the 18 genes and genes that are only expressed in scaled pleco but not in scaleless channel catfish allowed identification of three genes. These three genes, lymphocyte antigen 6D, secretory calcium-binding phosphoprotein, and uncharacterized protein si:dkey-30j10.5, were only expressed in both scaled fish but not in scaleless fish, providing strong evidence that they might account for the evolutionary loss of scale in catfish.

In the present study, a narrow list of genes were identified that may be required for scale formation, and their loss or mutation may be responsible for the evolutionary loss of scales. However, as powerful as the RNA-Seq and related comparative transcriptome analysis, genomic and transcriptomic analysis alone cannot provide direct evidence for the functions of these candidate genes. Future functional studies such as knock out experiments using CRISPR or its related technologies are warranted.

CHAPTER VI. CONCLUSIONS

In my dissertation project, transcriptomes from channel catfish, common carp, and pleco were generated and assembled. Comparative transcriptome subtraction analysis, coupled to scale regeneration experiments, provided insights into the evolutionary scaleless phenotype. As a result, a total of 863 specific expressed genes were identified in common carp skin comparing with channel catfish skin, and a set of 1,173 differentially expressed genes during scale regeneration were identified. The cross intersection of these two lists allowed identification of 18 genes, that were only expressed in scaled common carp but not in channel catfish and were differentially expressed during scale regeneration, providing a pool of genes important for scale formation. A total of 704 genes were only expressed in pleco skin but not in channel catfish skin. Of the shared 18 genes, that were only expressed in scaled carp but not in scaleless catfish and were differentially expressed during carp scale regeneration, three genes, lymphocyte antigen 6D, secretory calcium-binding phosphoprotein, and uncharacterized protein si:dkey-30j10.5, were only expressed in scaled pleco but not in scaleless channel catfish. These genes may be among the pool of genes responsible for the evolutionary lack of scales in catfish. Future functional studies such as knock out experiments are required to demonstrate that these genes indeed are the genomic basis of the scalelessness in catfish. My research also demonstrated that comparative

transcriptome subtraction is an efficient and effective alternative for searching candidate genes of structural significance, and this methodology is broadly applicable to other biological characteristics.

REFERENCE

- Arce, M., Reis, R.E., Geneva, A.J., Pérez, M.H.S., 2013. Molecular phylogeny of thorny catfishes (*Siluriformes: Doradidae*). *Molecular phylogenetics and evolution* 67, 560-577.
- Armbruster, J.W., 2004. Phylogenetic relationships of the suckermouth armoured catfishes (*Loricariidae*) with emphasis on the Hypostominae and the Ancistrinae. *Zoological Journal of the Linnean Society* 141, 1-80.
- Armbruster, J.W., 2011. Global catfish biodiversity, *American Fisheries Society Symposium*, pp. 15-37.
- Balon, E.K., 1995. Origin and domestication of the wild carp, *Cyprinus carpio*: from Roman gourmets to the swimming flowers. *Aquaculture* 129, 3-48.
- Bauer, S., Grossmann, S., Vingron, M., Robinson, P.N., 2008. Ontologizer 2.0--a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics* 24, 1650-1651.
- Bereiter-Hahn, J., Zylberberg, L., 1993. Regeneration of teleost fish scale. *Comparative Biochemistry and Physiology Part A: Physiology* 105, 625-641.
- Birol, I., Jackman, S.D., Nielsen, C.B., Qian, J.Q., Varhol, R., Stazyk, G., Morin, R.D., Zhao, Y., Hirst, M., Schein, J.E., Horsman, D.E., Connors, J.M., Gascoyne, R.D., Marra, M.A., Jones, S.J., 2009. *De novo* transcriptome assembly with ABySS. *Bioinformatics* 25, 2872-2877.
- Bolger, A.M., Lohse, M., Usadel, B., 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30, 2114-2120.

Bostock, J., McAndrew, B., Richards, R., Jauncey, K., Telfer, T., Lorenzen, K., Little, D., Ross,

L., Handisyde, N., Gatward, I., 2010. Aquaculture: global status and trends. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365, 2897-2912.

Bruet, B.J., Song, J., Boyce, M.C., Ortiz, C., 2008. Materials design principles of ancient fish armour. *Nature materials* 7, 748-756.

Chauhan, P., Hansson, B., Kraaijeveld, K., de Knijff, P., Svensson, E.I., Wellenreuther, M., 2014.

De novo transcriptome of *Ischnura elegans* provides insights into sensory biology, colour and vision genes. *BMC genomics* 15, 808.

Colbert, E.H., 1955. Evolution of the vertebrates. A history of the backboned animals through time. *Evolution of the vertebrates. A history of the backboned animals through time.*

Conesa, A., Götz, S., García-Gómez, J.M., Terol, J., Talón, M., Robles, M., 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research.

Bioinformatics 21, 3674-3676.

de Vrieze, E., Sharif, F., Metz, J.R., Flik, G., Richardson, M.K., 2011. Matrix metalloproteinases in osteoclasts of ontogenetic and regenerating zebrafish scales. *Bone* 48, 704-712.

Eschmeyer, W., Fong, J., 2015. Species by family/subfamily in the Catalog of Fishes.

Ferraris Jr, C.J., Vari, R.P., 2012. Review of the African catfish genus *Andersonia* (*Teleostei: Siluriformes*). *Zootaxa* 3210, 39-49.

- García-Ortega, L., Alegre-Cebollada, J., García-Linares, S., Bruix, M., Martínez-del-Pozo, Á., Gavilanes, J.G., 2011. The behavior of sea anemone actinoporins at the water–membrane interface. *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1808, 2275-2288.
- Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature biotechnology* 29, 644-652.
- Graham, D.M., Huang, L., Robinson, K.R., Messerli, M.A., 2013. Epidermal keratinocyte polarity and motility require Ca²⁺(+) influx through TRPV1. *Journal of cell science* 126, 4602-4613.
- Guttman, M., Garber, M., Levin, J.Z., Donaghey, J., Robinson, J., Adiconis, X., Fan, L., Koziol, M.J., Gnirke, A., Nusbaum, C., Rinn, J.L., Lander, E.S., Regev, A., 2010. *Ab initio* reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nature biotechnology* 28, 503-510.
- Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., Li, B., Lieber, M., 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature protocols* 8, 1494-1512.
- Harris, M.P., Rohner, N., Schwarz, H., Perathoner, S., Konstantinidis, P., Nusslein-Volhard, C., 2008. Zebrafish *eda* and *edar* mutants reveal conserved and ancestral roles of ectodysplasin signaling in vertebrates. *PLoS Genet* 4, e1000206.

- Holtan, P., 1998. Catfish. On-line. Accessed April 20, 2014.
- Huyseune, A., Sire, J.Y., 1998. Evolution of patterns and processes in teeth and tooth- related tissues in non- mammalian vertebrates. *European journal of oral sciences* 106, 437-481.
- Iger, Y., Abraham, M., 1990. The process of skin healing in experimentally wounded carp. *Journal of Fish Biology* 36, 421-437.
- Iida, Y., Hibiya, K., Inohaya, K., Kudo, A., 2014. Eda/Edar signaling guides fin ray formation with preceding osteoblast differentiation, as revealed by analyses of the medaka all- fin less mutant afl. *Developmental Dynamics* 243, 765-777.
- Iimura, K., Tohse, H., Ura, K., Takagi, Y., 2012. Expression patterns of runx2, sparx, and bglp during scale regeneration in the goldfish *Carassius auratus*. *Journal of experimental zoology. Part B, Molecular and developmental evolution* 318, 190-198.
- Kawasaki, K., Amemiya, C.T., 2014. SCPP genes in the coelacanth: tissue mineralization genes shared by sarcopterygians. *Journal of experimental zoology. Part B, Molecular and developmental evolution* 322, 390-402.
- Kawasaki, K., Buchanan, A.V., Weiss, K.M., 2007. Gene duplication and the evolution of vertebrate skeletal mineralization. *Cells, tissues, organs* 186, 7-24.
- Kawasaki, K., Weiss, K.M., 2006. Evolutionary genetics of vertebrate tissue mineralization: the origin and evolution of the secretory calcium-binding phosphoprotein family. *Journal of experimental zoology. Part B, Molecular and developmental evolution* 306, 295-316.

- Kolanczyk, M.E., 2009. Signaling mechanisms and developmental function of fibroblast growth factor receptors in zebrafish. Gdansk, Polen.
- Kondo, S., Kuwahara, Y., Kondo, M., Naruse, K., Mitani, H., Wakamatsu, Y., Ozato, K., Asakawa, S., Shimizu, N., Shima, A., 2001. The medaka *rs-3* locus required for scale development encodes ectodysplasin-A receptor. *Current biology* 11, 1202-1206.
- Kottelat, M., Freyhof, J., 2007. Handbook of European freshwater fishes. Publications Kottelat Cornol.
- Lagler, K.F., 1962. Ichthyology: The Study Of Fishes. Ontario, Canada.
- Lanzing, W., Wright, R., 1976. The ultrastructure and calcification of the scales of *Tilapia mossambica* (*Peters*). *Cell and tissue research* 167, 37-47.
- Li, C., Beck, B.H., Peatman, E., 2014. Nutritional impacts on gene expression in the surface mucosa of blue catfish (*Ictalurus furcatus*). *Developmental and comparative immunology* 44, 226-234.
- Li, C., Wang, R., Su, B., Luo, Y., Terhune, J., Beck, B., Peatman, E., 2013. Evasion of mucosal defenses during *Aeromonas hydrophila* infection of channel catfish (*Ictalurus punctatus*) skin. *Developmental and Comparative Immunology* 39, 447-455.
- Li, C., Zhang, Y., Wang, R., Lu, J., Nandi, S., Mohanty, S., Terhune, J., Liu, Z., Peatman, E., 2012. RNA-seq analysis of mucosal immune responses reveals signatures of intestinal barrier disruption and pathogen entry following *Edwardsiella ictaluri* infection in channel catfish, *Ictalurus punctatus*. *Fish and shellfish immunology* 32, 816-827.

- Li, R., Yu, C., Li, Y., Lam, T.W., Yiu, S.M., Kristiansen, K., Wang, J., 2009. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* 25, 1966-1967.
- Liu, S., Wang, X., Sun, F., Zhang, J., Feng, J., Liu, H., Rajendran, K.V., Sun, L., Zhang, Y., Jiang, Y., Peatman, E., Kaltenboeck, L., Kucuktas, H., Liu, Z., 2013. RNA-Seq reveals expression signatures of genes involved in oxygen transport, protein synthesis, folding, and degradation in response to heat stress in catfish. *Physiological genomics* 45, 462-476.
- Liu, S., Zhang, Y., Zhou, Z., Waldbieser, G., Sun, F., Lu, J., Zhang, J., Jiang, Y., Zhang, H., Wang, X., Rajendran, K.V., Khoo, L., Kucuktas, H., Peatman, E., Liu, Z., 2012. Efficient assembly and annotation of the transcriptome of catfish by RNA-Seq analysis of a doubled haploid homozygote. *BMC genomics* 13, 595.
- Liu, S., Zhou, Z., Lu, J., Sun, F., Wang, S., Liu, H., Jiang, Y., Kucuktas, H., Kaltenboeck, L., Peatman, E., 2011. Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. *BMC genomics* 12, 53.
- Liu, Y., Zhang, W., 2008. Identification of a new transmembrane adaptor protein that constitutively binds Grb2 in B cells. *Journal of leukocyte biology* 84, 842-851.
- Luo, C., Tsementzi, D., Kyrpides, N., Read, T., Konstantinidis, K.T., 2012. Direct comparisons of Illumina vs. Roche 454 sequencing technologies on the same microbial community DNA sample. *PLoS one* 7, e30087.
- Martin, J.A., Wang, Z., 2011. Next-generation transcriptome assembly. *Nature Reviews Genetics* 12, 671-682.

- McGraw, H.F., Drerup, C.M., Culbertson, M.D., Linbo, T., Raible, D.W., Nechiporuk, A.V.,
2011. Lef1 is required for progenitor cell identity in the zebrafish lateral line primordium.
Development 138, 3921-3930.
- Metzker, M.L., 2010. Sequencing technologies—the next generation. *Nature Reviews Genetics*
11, 31-46.
- Monnot, M.J., Babin, P.J., Poleo, G., Andre, M., Laforest, L., Ballagny, C., Akimenko, M.A.,
1999. Epidermal expression of apolipoprotein E gene during fin and scale development and
fin regeneration in zebrafish. *Developmental Dynamics* 214, 207-215.
- Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., Snyder, M., 2008. The
transcriptional landscape of the yeast genome defined by RNA sequencing. *Science* 320,
1344-1349.
- Ohira, Y., Shimizu, M., Ura, K., Takagi, Y., 2007. Scale regeneration and calcification in
goldfish *Carassius auratus*: quantitative and morphological processes. *Fisheries Science* 73,
46-54.
- Page, L., Burr, B.M., 1991. A field guide to freshwater fishes of North America north of Mexico.
Peterson Field Guide Series, Houghton Mifflin Co., Boston.
- Peatman, E., Li, C., Peterson, B.C., Straus, D.L., Farmer, B.D., Beck, B.H., 2013. Basal
polarization of the mucosal compartment in *Flavobacterium columnare* susceptible and
resistant channel catfish (*Ictalurus punctatus*). *Molecular immunology* 56, 317-327.

- Quak, J., Balm, A., van Dongen, G., Brakkee, J., Scheper, R., Snow, G., Meijer, C., 1990. A 22-kd surface antigen detected by monoclonal antibody E 48 is exclusively expressed in stratified squamous and transitional epithelia. *The American journal of pathology* 136, 191.
- Quilhac, A., Sire, J.Y., 1999. Spreading, proliferation, and differentiation of the epidermis after wounding a cichlid fish, *Hemichromis bimaculatus*. *The Anatomical record* 254, 435-451.
- Raschi, W., Tabit, C., 1992. Functional aspects of placoid scales: a review and update. *Marine and Freshwater Research* 43, 123-147.
- Reif, W.E., 1982. *Evolution of dermal skeleton and dentition in vertebrates*. Springer.
- Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., Mungall, K., Lee, S., Okada, H.M., Qian, J.Q., Griffith, M., Raymond, A., Thiessen, N., Cezard, T., Butterfield, Y.S., Newsome, R., Chan, S.K., She, R., Varhol, R., Kamoh, B., Prabhu, A.L., Tam, A., Zhao, Y., Moore, R.A., Hirst, M., Marra, M.A., Jones, S.J., Hoodless, P.A., Birol, I., 2010. *De novo* assembly and analysis of RNA-seq data. *Nature methods* 7, 909-912.
- Rohner, N., Bercs ényi, M., Orb án, L., Kolanczyk, M.E., Linke, D., Brand, M., Nüsslein-Volhard, C., Harris, M.P., 2009. Duplication of *fgfr1* permits Fgf signaling to serve as a target for selection during domestication. *Current Biology* 19, 1642-1647.
- Russo, G., Zegar, C., Giordano, A., 2003. Advantages and limitations of microarray technology in human cancer. *Oncogene* 22, 6497-6507.

- Saban, R., D'Andrea, M.R., Andrade-Gordon, P., Derian, C.K., Dozmorov, I., Ihnat, M.A., Hurst, R.E., Simpson, C., Saban, M.R., 2007. Regulatory network of inflammation downstream of proteinase-activated receptors. *BMC physiology* 7, 3.
- Schulz, M.H., Zerbino, D.R., Vingron, M., Birney, E., 2012. Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. *Bioinformatics* 28, 1086-1092.
- Sharpe, P.T., 2001. Fish scale development: Hair today, teeth and scales yesterday? *Current biology* 11, R751-752.
- Sire, J.Y., G éraudie, J., 1984. Fine structure of regenerating scales and their associated cells in the cichlid *Hemichromis bimaculatus* (Gill). *Cell and tissue research* 237, 537-547.
- Sire, J.Y., Akimenko, M.A., 2004. Scale development in fish: a review, with description of sonic hedgehog (*shh*) expression in the zebrafish (*Danio rerio*). *The International journal of developmental biology* 48, 233-247.
- Sire, J.Y., Allizard, F., Babiari, O., Bourguignon, J., Quilhac, A., 1997. Scale development in zebrafish (*Danio rerio*). *Journal of Anatomy* 190, 545-561.
- Sire, J.Y., Huysseune, A., 2003. Formation of dermal skeletal and dental tissues in fish: a comparative and evolutionary approach. *Biological reviews of the Cambridge Philosophical Society* 78, 219-249.
- Steinmetz, A., Barbaras, R., Ghalim, N., Clavey, V., Fruchart, J.C., Ailhaud, G., 1990. Human apolipoprotein A-IV binds to apolipoprotein A-I/A-II receptor sites and promotes cholesterol efflux from adipose cells. *J Biol Chem* 265, 7859-7863.

- Sun, L., Liu, S., Wang, R., Jiang, Y., Zhang, Y., Zhang, J., Bao, L., Kaltenboeck, L., Dunham, R., Waldbieser, G., 2014. Identification and analysis of genome-wide SNPs provide insight into signatures of selection and domestication in channel catfish (*Ictalurus punctatus*). PloS one, 9, e109666.
- Sun, T.T., 2006. Altered phenotype of cultured urothelial and other stratified epithelial cells: implications for wound healing. American journal of physiology. Renal physiology 291, F9-21.
- Thamamongood, T.A., Furuya, R., Fukuba, S., Nakamura, M., Suzuki, N., Hattori, A., 2012. Expression of osteoblastic and osteoclastic genes during spontaneous regeneration and autotransplantation of goldfish scale: a new tool to study intramembranous bone regeneration. Bone 50, 1240-1249.
- Thesleff, I., Sharpe, P., 1997. Signalling networks regulating dental development. Mechanisms of development 67, 111-123.
- Trapnell, C., Williams, B.A., Pertea, G., Mortazavi, A., Kwan, G., van Baren, M.J., Salzberg, S.L., Wold, B.J., Pachter, L., 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nature biotechnology 28, 511-515.
- Vieira, F.A., Gregorio, S.F., Ferrareso, S., Thorne, M.A., Costa, R., Milan, M., Bargelloni, L., Clark, M.S., Canario, A.V., Power, D.M., 2011. Skin healing and scale regeneration in fed and unfed sea bream, *Sparus auratus*. BMC genomics 12, 490.

- Wang, R., Sun, L., Bao, L., Zhang, J., Jiang, Y., Yao, J., Song, L., Feng, J., Liu, S., Liu, Z., 2013. Bulk segregant RNA-seq reveals expression and positional candidate genes and allele-specific expression for disease resistance against enteric septicemia of catfish. *BMC genomics* 14, 929.
- Wang, Z., Gerstein, M., Snyder, M., 2009. RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews. Genetics* 10, 57-63.
- Wellborn, T.L., 1988. Channel catfish: life history and biology. SRAC publication (USA).
- Wilson, D.E., Reeder, D.M., 2005. Mammal species of the world: a taxonomic and geographic reference. JHU Press.
- Wu, J., Xiong, S., Jing, J., Chen, X., Wang, W., Gui, J.F., Mei, J., 2015. Comparative Transcriptome Analysis of Differentially Expressed Genes and Signaling Pathways between XY and YY Testis in Yellow Catfish. *PloS one* 10, e0134626.
- Xie, B., Zhou, G., Chan, S.Y., Shapiro, E., Kong, X.P., Wu, X.R., Sun, T.T., Costello, C.E., 2006. Distinct glycan structures of uroplakins Ia and Ib: structural basis for the selective binding of FimH adhesin to uroplakin Ia. *J Biol Chem* 281, 14644-14653.
- Xu, C., Evensen, O., Munang'andu, H.M., 2015. *De novo* assembly and transcriptome analysis of Atlantic salmon macrophage/dendritic-like TO cells following type I IFN treatment and Salmonid alphavirus subtype-3 infection. *BMC genomics* 16, 96.
- Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome research* 18, 821-829.

Zhu, D., Ortega, C.F., Motamedi, R., Szewciw, L., Vernerey, F., Barthelat, F., 2012. Structure and mechanical performance of a “modern” fish scale. *Advanced Engineering Materials* 14, B185-B194.