Identification and analysis of recently duplicated genes in Channel Catfish (*Ictalurus punctatus*) genome

by

Yu Zhang

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 14, 2013

Keywords: gene duplication, catfish, genome, teleost

Approved by

Zhanjiang Liu, Chair, Alumni Professor of Fisheries and Allied Aquacultures
Eric Peatman, Assistant Professor of Fisheries and Allied Aquacultures
Rex Dunham, Professor of Fisheries and Allied Aquacultures
Nannan Liu, Professor of Entomology and Plant Pathology

Abstract

Gene duplication is an important source of genetic innovation and structure and usually associated to the lineage-specific adaptive trait during evolution, such as genes related to immunity, development and reproduction. Channel catfish (*Ictalurus punctatus*) is the primary aquaculture species in the United States, which represents 70% of the total aquaculture production. However, the set size of duplicated genes and duplication pattern in channel catfish are still less well understood.

In this study, a total of 2,130 duplicated gene sets were identified by using predicted mRNA in channel catfish. Similar overall distribution patterns of duplication size were observed among catfish and model fish species (zebrafish, medaka, stickle and *Tetraodon*) under comparison, significant differences were noted, with catfish harboring the highest percentage (69.4%) of duplicated gene sets with only two copies. A total of 1,128 duplication sets can be characterized with their duplication types with Intra-chromosomal duplications accounting for most proportion. The large number of duplicated genes with low Ks indicated that most of duplicated genes predicted in this study were recently duplicated genes. GO enrichment analysis indicated that GO terms associated to various metabolic and catabolic processes, response to stimuli and signal transduction were over-represented in duplicated genes. When compared with the four model fish species, a total of 239 catfish specific duplicated genes were identified, with 71.1% of duplication sets containing 2 copies and 59.4% of duplicated genes (pairwise) had small Ks values of ≤ 1. These results suggested that catfish species-specific duplicated genes were mostly recent. The catfish species-specific duplicated genes included a set of genes related to immune and stress responses suggesting a mechanism for catfish to adapt to its environment.

This study is the first study on surveying genome-wide duplicated genes in channel catfish. This research is not only useful for further catfish whole genome annotation, but also useful for understanding environmental adaption in channel catfish.

Acknowledgments

First and foremost, I would like to express my deep and sincere gratitude to my advisor Dr. Zhanjiang Liu for his constant guidance, support and patience throughout my degree program. His scientific attitude and philosophy have had a profound influence on me. Under his supervision, I learned how to do scientific research. Without him, it would be impossible for me to finish this dissertation. I would like to express my gratitude to my committee: Dr. Nannan Liu, Dr. Eric Peatman, Dr. Dunham and my dissertation outside reader, Dr. Bernhard Kaltenboeck for their advice and critical reading of my dissertation. Sincere thanks will also go to Huseyin Kucuktas and Ludmilla Kaltenboeck for technical help, and the colleagues in the laboratory for their help, collaboration, and friendship. I would like to express my gratitude to the Chinese Scholarship Council for the financial support through my PhD study. Finally, I would like to thank my parents and all my family members for their love and support.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

SD	Segmental duplication
SNPs	Single nucleotide polymorphisms
PSVs	Paralogous sequence variants
MSVs	Multisite variants
NHEJ	Non-homologous end joining
CNV	Copy number variations
NAHR	Non-allelic homologous recombination
aCGH	Array-based comparative genomic hybridization
WGD	Whole genome duplication
WSSD	Whole Genome Shotgun Sequence Detection
WGAC	Whole Genome Assembly Comparison
FISH	Fluorescent in situ hybridization
WGS	Whole genome shotgun

I. INTRODUCTION AND LITERATURE REVIEW

The earliest observation of gene duplication was reported by Bridges in 1936 [1]. Bridges observed that the mutant fruit fly Drosophila melanogaster with doubling of a chromosomal band had extreme reduction in eye size. In 1970s, Susumo Ohno further popularized that duplications is the main force on supplying raw genetic material to biological evolution [2]. Ohno suggested that gene duplication might be a more important mechanism in evolution than natural selection. The basic idea of ‘‘Ohno-mechanism’’ is that, after duplication, one copy of a gene would retain the original function and that the additional copy would be more flexible for mutation driven by natural selection and evolve into a gene with novel functions [2]. This hypothesis has been validated by analysis of whole genome sequences. For instance, there are about 38% duplicated gene in *Homo. Sapiens* [3] and are about 80% duplicated genes in *Arabidopsis thaliana* [4].

Mechanism of gene duplication

Gene duplication can range from affecting single genes, large chromosomal fragment and whole genomes (polyploidy) [5]. Three main mechanisms have been proposed to be responsible for generation of gene duplications: unequal crossing-over, retroposition, and chromosomal (or genome) duplication [6-7] (Figure1). The outcomes of these mechanisms are quite different. Unequal crossing-over usually generates tandem gene duplication, that is, duplicated genes are physically linked in close locations on a chromosome (Figure 1a). Retroposition is a process that a message RNA (mRNA) is retrotranscribed to complementary DNA (cDNA) and then inserted into the genome [6-9] (Figure 1b). There are several molecular features of duplicated genes generated by retroposition. First, retroposed gene copy tends to be pseudogene because promoter and regulatory sequences are not transcribed. Second, functional retroposed gene copy has to recruit a new regulatory sequence and form a chimeric structure. Third, retroposed gene copy haslost its introns and regulatory sequences, but harbors its poly A tracts and flanking short direct repeats. Forth, retroposed gene copy is usually not linked to original gene because the location of insertion is random [6-9]. Unequal crossing-over and retroposition usually generate small-scale duplications.

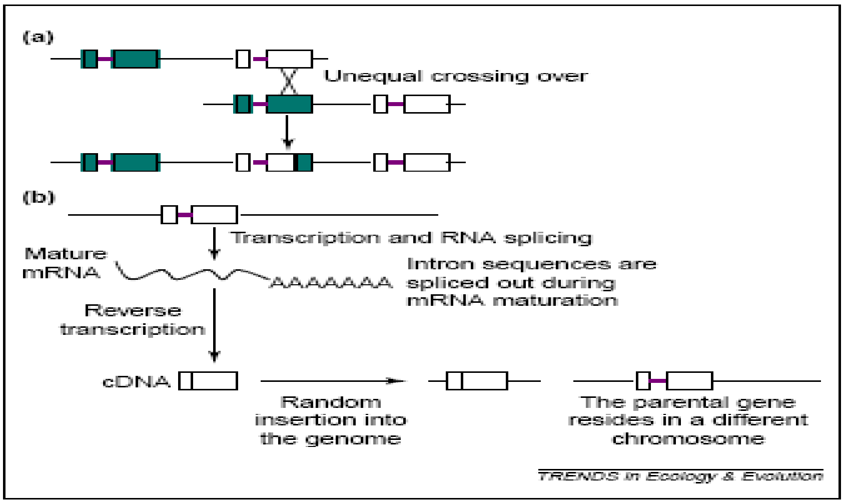


Figure 1. Mechanisms of gene duplication [6]. (a). Unequal crossing-over generates tandem gene duplication. (b). Mechanisms of retroposition duplications.

Large-scale duplications, such as chromosomal or genome duplications (polyploidy), occur because of a lack of disjunction among daughter chromosomes after DNA replication [7]. Chromosomal duplications have a major impact in the generation of genetic diversity because all the genes in the genome are duplicated simultaneously. Tandem duplication and genome duplication are considered as two dominant mechanisms contributing to vertebrate genome evolution [7, 10-11]. Human genome analysis reveals another type of large-scale duplication, segmental duplications, which are duplicated blocks of genomic DNA typically ranging in size from 1–200 kb. Further studies indicated that segmental duplications display an interspersed pattern of recent duplications in human and other primate genomes [12-14]. In contrast, in most mammalian lineages segmental duplications are organized in a tandem configuration, particularly in pericentromeric and subtelomeric regions. In spite of its significance, the exact mechanism of segmental duplication is unclear at present [15-17].

The fate of duplicated genes

There are various hypothesis on the fate of duplicated genes. The prevailing theory is that many duplicated genes are lost from the moment of their origin (Figure 2). The gene loss resulted from degenerative mutation (non-functionalization), which can become pseudogene or be totally deleted [18-19]. The "neofunctionalization" model indicated that the additional copy of gene evolves into a gene with a new function while the original one keeps its ancestral function. The neofunctionalization of duplicated genes requires varying numbers of amino acid substitutions and its diversification over the long period of time [20-21]. Another model called "subfunctionalization" indicated that two copies of a duplicated gene each lose a subset of functions and diversify by partitioning the ancestral function. After subfunctionalization, duplicated gene develops a distinct expression pattern. Subfunctionalization seems to be the most common phenomena and occurs in early stage after gene duplication [20-21] (Figure 2). On the other hand, duplicated genes can maintain the same function and expression patterns. However the genes with identical functions are unlikely to be stably maintained in the genome unless increased gene dosage is advantageous [22-23].

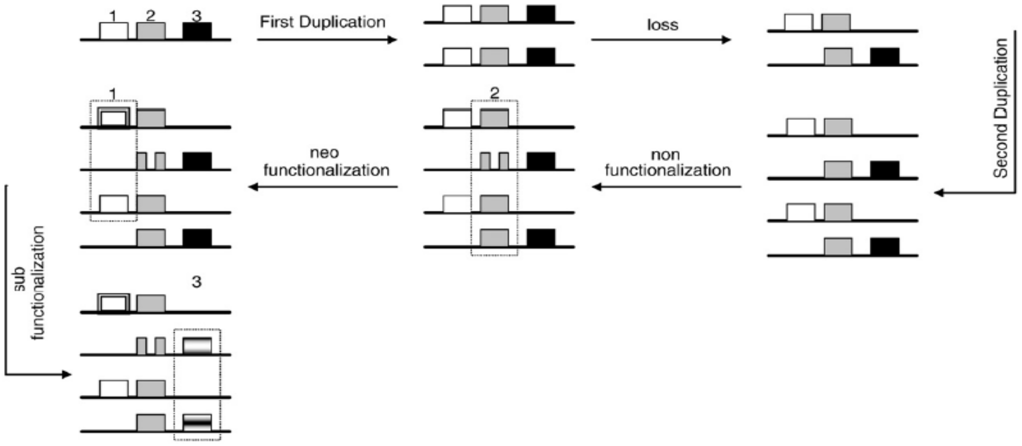


Figure 2. Gene duplication and modification of three genes (1, 2, 3) [25].

Gene families are a cluster of genes generated by gene duplication. Members of a gene family are paralogous one another. Paralogs reflect a horizontal event, that is, paralogs are created when duplication occur within a species. While orthologs are homologous genes derived from a same ancestral gene, which reflect a vertical event (speciation) [24].

Tandem duplication

Tandem duplication refers to a cluster of duplicated genes that are located next to each other in the same chromosome. It is considered as one of the the driving forces in the origin and maintenance of gene families in evolution [11]. Tandem duplications seemed to play an important role in fine-tuning of developmental stages and physiological functions. They also enhance disease resistance or satisfy the requirement for a large dosage of a gene product [26]. Lu et al 2012 found that tandem duplications account for a large number of duplicated genes in the model teleosts and are involved in immune and sensory response pathway [27]. Many immune-related genes in teleosts were verified as tandem duplications. For instance, the novel immune-type receptors (NITRs), a subset of innate immune receptors organized in a tandemly arrayed cluster, were identified in European sea bass [28]. Some other immune-related genes, such as STAT1 genes [29], tumor necrosis factor superfamily genes (TNFSF) [30], microfibrillar-associated protein 4 genes (MFAP4) [31], chemokine family genes [32-33], lectin genes [34], MHC genes [35], immunoglobulin D genes [36], have been found to be tandem duplications in teleosts. In addition, some genes associated with important physiological functions were also found to be present as tandem duplicated genes in teleosts, such as vasotocin genes [37], sodium channel alpha-subunit (SCNA) gene family [38] and neuroendocrine peptides genes [39].

Segmental duplications

Segmental duplication (SD), also known as low-copy repeat, is defined as duplicated blocks of genomic DNA in length greater than 1kb, which share a high level of sequence identity (>90%). SDs often contain gene sequences with intron-exon structure and high copy repeats [40-44]. Characterization of the human genome revealed that SDs were over-represented in unordered and unassigned contigs indicating that SDs are difficult to assign to their proper position. Thus, the exact location of SDs in the genome is difficult to be determined because of their high levels of sequence similarity [12, 40-41]. In addition, a single-base difference within SDs may be misjudged as single nucleotide polymorphisms (SNPs), which are present in different alleles. This genetic variation associated with SDs has been observed in human genome and called paralogous sequence variants (PSVs) [45-47], which don’t segregate the substitution within either paralogue.

In contrast to PSVs, another source of variation in polyploid genomes are multisite variants (MSV) which segregate for a base substitution in one or both of the paralogous loci [46]. However PSVs and true SNPs can be distinguished by quantitative genotyping in a set of ordinary individuals and double haploid individuals (complete hydatidiform moles) [45-46]. Through RNA-seq, a total of 4,878 PSVs or MSVs were detected in double haploid catfish [48]. Recent genotyping in Atlantic salmon also revealed 1,471 MSVs by iSelect SNP array [49].

Experimental and computational approaches indicated that the human and chimpanzee have greater level SDs (~5%) than other mammals (mouse, dog and cattle), whose SDs account for 2-4% in genomes [15-17, 50]. However, compared with primate and mammalian genomes, other eukaryote genomes (worm and fly) show less SD contents. The different content of SD probably reflects evolutionary constraints imposed by the smaller genome sizes of non-mammalian species, since SDs play an important role in genetic innovation and structural variation in genome [51]. SDs were detected as inter- and intrachromosomal duplications, but the distribution of SDs is nonrandom, with clustering particularly located in pericentromeric and subtelomeric regions and most are intrachromosal duplications [13-17]. In primate genomes (including human), the SD clusters represent complex mosaics architecture, that is, individual ancestral duplicated segments are located adjacent to other duplicons that derived from diverse ancestor [13, 50]. In contrast, in mammalian genomes such as cattle, mouse and dog, SDs were primarily organized into local tandem duplication clusters with preference for pericentromeric and subtelomeric regions [15-17]. Overall, the patterns of SD differ from lineage to lineage.

The origin and mechanism of SDs are still not well understood. Several mechanisms are proposed for different patterns of SD. Tandem duplications are thought to occur mainly through non-allelic homologous recombination (NAHR) and replication error [52]. However the origin and the mechanism of the dispersion of SDs are more complex. Within pericentromeric region *Alu*-mediated transposition is suggested as a probable mechanism of SDs because the enrichment of *Alu*, a type of SINE repeats, are observed in SD breakpoints in primate lineage. Non-homologous end joining (NHEJ)-mediated translocations followed by the serial transfer of sequences is proposed as a probably mechanism for SDs in subtelomeric region [53-55].

Copy number variation

Copy number variation is a form of genomic variations involving differences in copy numbers of segments of DNA derived from deletions (fewer than the normal number) or duplications (more than the normal number) on certain chromosomes. Many studies have indicated that SDs are hotspots for formation of copy number variations (CNV), disease-causing rearrangement and evolutionary breakpoints during primate genome evolution. The significant relationship between SDs and CNVs have been found in studies with primate and mammalian genomes because the regions flanked by SDs are susceptible to rearrangement by non-allelic homologous recombination (NAHR) that is the main mechanism of CNV formation [13, 17, 56-57]. CNV is defined as gains and losses of genomic sequence (>50bp) between two individuals of species, which has been shown to play an important role in phenotypic variability and disease susceptibility [58-59]. Hypothesis was proposed that before fixation of SDs by positive selection, the SDs undergo lineage-specific expansions in copy numbers, which is essential for the evolution of species-specific adaptive traits [15, 57]. CNVs associate frequently with SDs and often occur in gene-rich regions. In cattle, CNV regions contain a range of 300 to 1,200 genes and there are 61.02 % of CNV regions directly overlapped with SDs [60-62]. In human, many diseases have been found to be associated with CNV through the change of gene dosage, such as autism [63] and schizophrenia [64]. In cattle, multiple studies found an enrichment of genes related to immunity and defense, sensory perception (e.g., olfactory receptors), response to external stimuli in CNV regions. [60-62]. Recent studies of zebrafish CNV revealed that high level of CNVs in zebrafish (69%) exhibited strain specificity and indicated that high level of CNVs may have functional effects that impact phenotype [65].

Traditional methods for the detection of CNV is array-based comparative genomic hybridization (aCGH) [66]. Because of ongoing developments and cost decreases in next-generation sequencing (NGS) , sequence-based approaches to detect CNV are becoming more popular and allow CNV reconstruction at a higher resolution and sensitivity [58, 67]. Sequence-based approaches include: Paired end mapping analysis (or Read pair analysis) is used to detect structural variations by tracking read pairs' span (deletions and insertions) and orientation (inversions and/or translocations) after mapping [68-69] (Figure 3A). Read depth analysis is used to construct personalized SD and CNV maps based on calculating the representation of reads [70-71] (Figure 3B). Split read method is used to identify breakpoint [72] (Figure 3C) and Local de novo assembly is used to identify novel insertions [73] (Figure 3D). Experimental approaches such as quantitative PCR (qPCR), fluorescent in situ hybridization (FISH), and sequencing can be used to validate CNVs [74-75].

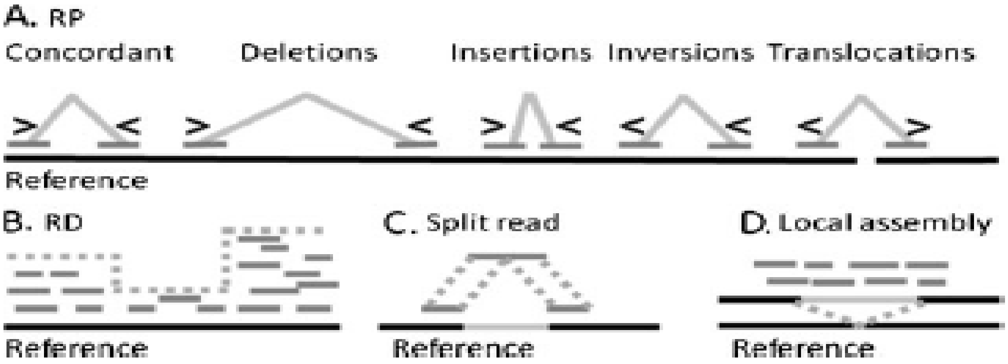


Figure 3. Approaches for detecting CNVs by DNA next-generation sequencing technologies [57].

Fish-specific genome duplication (FSGD) and Hox gene cluster

During evolution the entire genome of vertebrates underwent two rounds of duplication. Later, ~350 million years ago (mya), it is believed that only ray-finned fishes (actinopterygian) had a third round whole genome duplication, that is, the fish-specific genome duplication (FSGD or 3R) [76-81]. Study on Hox gene clusters from a spectrum of vertebrate species provided critical evidence in support of this hypothesis [80-83]. Hox genes, a subset of the homeobox-containing superfamily, encode transcription factors involved in specification of axial patterning and the development of appendages and organ systems [84]. Due to their conserved nucleotide sequence, organized order within clusters and key role in developmental stage, Hox genes are considered as an attractive gene family in evolutionary and developmental studies. [85].

In chordates, Hox genes are organized as tandem cluster with conservation of gene order. It has been suggested that the single-cluster Hox genes with 13 paralogy groups is the prototype in chordates and have arisen at the stage of protochordate during evolution [86]. The study of Hox gene cluster in amphioxus (*Branchiostoma floridae*), ancestor of vertebrate lineage, indicated that amphioxus possesses a single cluster containing 14 Hox genes. However the 14th Hox gene was considered as the duplication of one of the 13 Hox genes [94]. Therefore this observation was not in conflict with above hypothesis [87]. In mammals and tetrapods, Hox genes are organized into four clusters resulting from 2R whole genome duplication [88]. Compared with relatively stable of Hox loci in mammals and tetrapods, the patterns of Hox genes are more variable among different teleost species. Because of FSGD, Hox genes in teleosts are organized into about 7 clusters, however, dramatic Hox gene loss are displayed in teleost genomes (Figure 4). For instance, there are total of 49 Hox genes within 7 clusters in zebrafish, 46 Hox genes within 7 clusters in Medaka and 48 Hox genes within 7 clusters in Fugu. However in zebrafish HoxDb cluster was lost, while in Medaka and Fugu, HoxCb cluster was lost [82, 89-90]. In addition, thirteen putative Hox clusters were identified in Salmonidae indicated that subsequent genome duplication (4R) occurred after FSGD in the salmon lineage [91]. Phylogenetic analysis of Hox genes in American Paddlefish indicated an independent duplication of the genome happened [92].

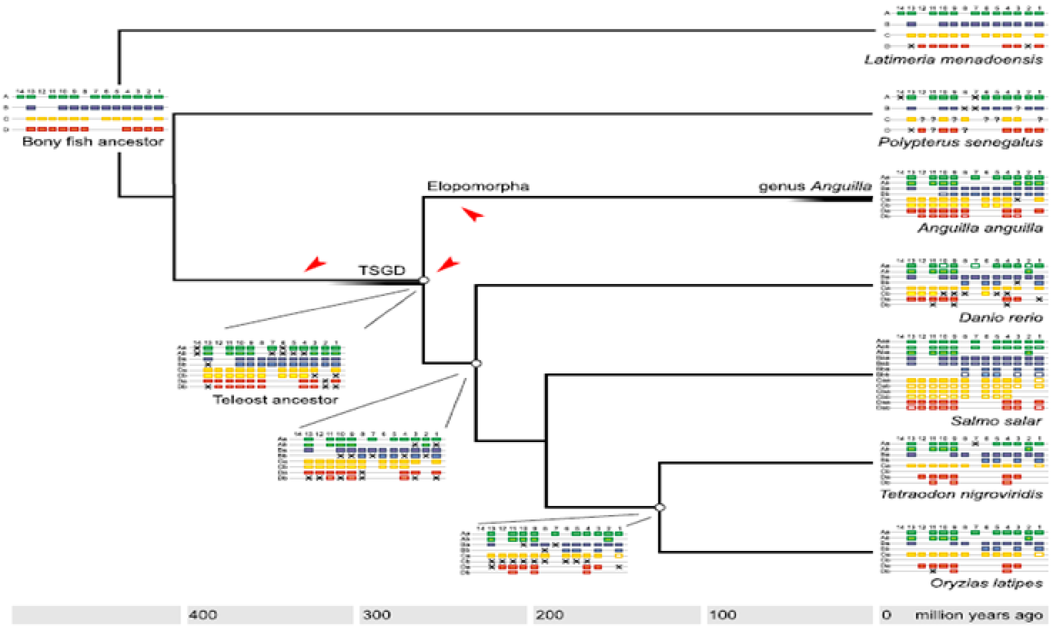


Figure 4. Model for the evolution of teleost Hox gene organization [93].

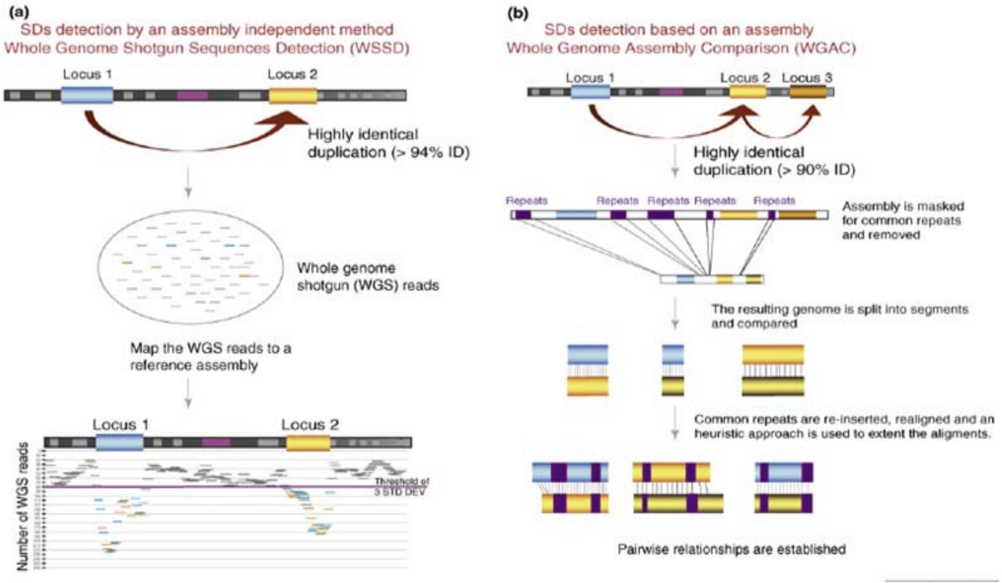
Identification of gene duplication

Accurate identification of gene duplication is difficult because the small duplication (tandem duplication and retrotransposition) and large duplication (segmental duplication and whole genome duplication) underwent genome rearrangement, gene loss, further duplication and mutation after creation [10, 76]. However, different computational approaches were established to diagnose gene duplication process from spatial and temporal aspects. Based on the spatial signature associated with a specific duplication mechanism, map-based and sequence-based approaches were developed [95].

Map-based method is used to detect homology between different chromosomal segments. Chromosomes are considered as lists of genes (the map), which are sorted according to their position on that chromosome, ignoring intergenic sequence. From these gene lists, genes that have homologues in other chromosomal segments are identified by all against all similarity searches. After that, local similarities in gene order and content are detected by comparing the map with itself [95-96]. This method has been applied to detect recent tandem duplications and other large-scale duplications, which are collinear between genomic segments with conservation of gene content and order [95-96]. Sequence-based approaches are most applied for identification of retroposed gene duplication and SDs, since both of them have featured structures [95]. Retroposed genes are often characterized as lack of introns, sometimes the presence of poly-A tails and flanking repeat sequences associated with the integration sites of transposable elements. While SDs are featured with high sequence identity and their flanking regions enriched with repeat sequences [8, 97]. For instance, Joseph et al. (2003) utilized genome sequence self-blast searches for identifying inter-chromosomal segmental duplications [47]. Similar studies using protein sequences through all-blast-all similarity searches followed single-linkage clustering or Markov clustering algorithm were also conducted for the identification of paralogs [19, 98]. With this method, Lu et al. (2012) compared duplication patterns among various sequenced model fish species [27].

Temporal and spatial analysis are also used to detect large-scale duplications, such as whole genome duplication (WGD), since large-scale duplications result in a set of paralogs that form simultaneously and duplication regions have similar gene content and similar order [95]. For temporal analysis, the fraction of synonymous substitutions per synonymous site (Ks) is a broadly used approach to estimate time of duplication for individual gene pairs. WGD event can be identified by a large peak in the histogram that plotted by the number of duplicated genes against estimated duplication times [96]. For spatial analysis to detect WGD, it starts with identification of putative paralogon by map-based method. Paralogon is defined as homologous regions in different chromosomes containing a similar set of paralogs in a similar order. Further temporal or phylogenetic evidence can be used to restrict the set of paralogs under consideration to those duplicated in the relevant time frame [95, 25]. WGD is evidenced by the uniform distributions of paralogon along the chromosomes of the entire genome [25, 99]. WGD is also evidenced by a two-to-one relationship between paralogons of the duplicated genome as compared to a non-duplicated genome [25, 79]. For instance, Hox gene clusters provide strong evidence for 2R and FSGD hypotheses, since invertebrates possess a single Hox cluster, while vertebrate possess four Hox clusters (2R) and teleost fish possess seven Hox clusters (3R) [80-83].

There are two well-established computational approaches to detect putative SDs, Whole Genome Shotgun Sequence Detection (WSSD) [40] and Whole Genome Assembly Comparison (WGAC) [12] (Figure 5b). These approaches have been applied in many organisms such as human [12, 40], chimpanzee [50], mouse [15], cattle [16] and dog [17]. WGAC was an assembly-dependent method using a BLAST-based strategy, which was first applied to describe global duplication content in the preliminary version of the human genome assembly [12] . Briefly, genome assembly is broken into 400 kb fragments and subsequently high-copy repeats are removed by RepeatMasker. Unique gnomic fragments were then compared by BLAST. After initial seed duplication alignments, common repeats reinsert to the gnomic fragments and local pairwise alignments are computed. The alignment ends undergo heuristic trimming allowing for refinement of alignment end points which may lie within common repetitive sequence. WGAC is advantageous in that it provides estimate of copy numbers and structural details of duplicated regions and the location of all duplications. However it depends on the quality of genome assembly and cannot be applied in organisms without a finished genome assembly [12, 44].



TRENDS in Genetics

WSSD is an assembly-independent approach. With this approach,each public sequence is analyzed at the clone level for overrepresentation within a whole-genome shotgun (WGS) reads against a reference assembly (with a defined identity threshold, usually 94%) [40] (Figure 5a). Duplications will be detected as an excess of read-depth and number of reads and average sequence identity are calculated across window intervals. This method is simple and sensitive to high identity duplications larger than 20 Kb. However, WSSD require WGS reads be randomly distributed and that all sequences are represented at least once within the reference assembly [40, 44]. WSSD can be applied in species without finished genome assembly, but with reference assembly that share genetic similarity. WSSD has been shown to work well in all the great ape WGS sequences mapped to the human assembly [50, 100]. In cattle, both WSSD and WGAC were applied to detect SDs and fluorescent in situ hybridization (FISH) was used for validation [16].

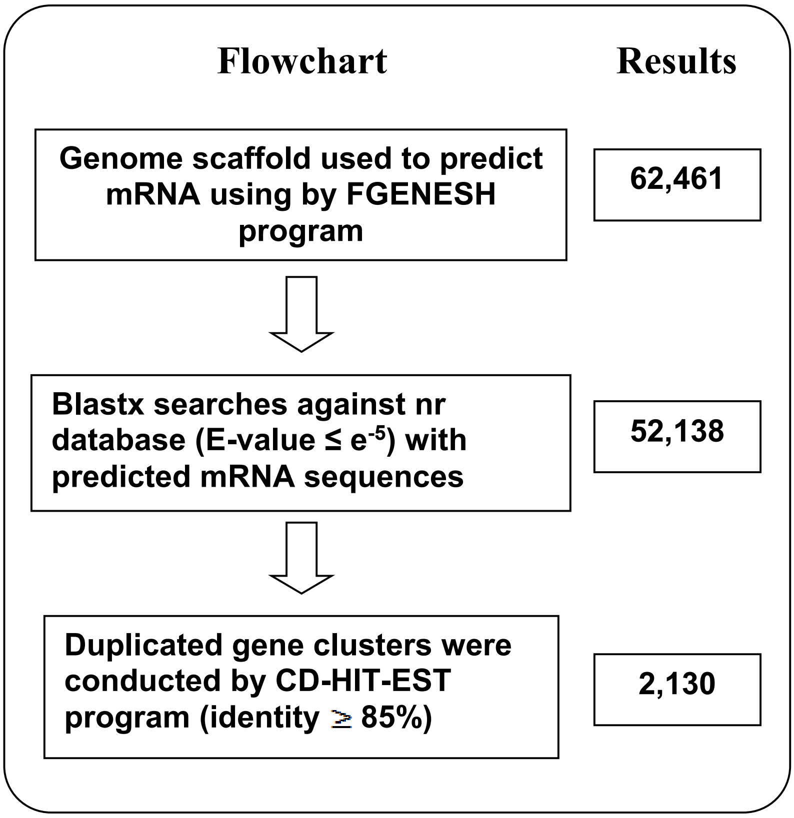
Channel catfish (*Ictalurus punctatus*) is the primary aquaculture species in the United States. However, gene duplication situation in catfish is not well understood. Understanding the pattern and nature of gene duplications is important for following reasons. First, Gene duplication is an important source of genetic innovation and structure variation that are main forces of evolution [2]. Second, duplicated genes are often associated to the lineage-specific adaptive trait during evolution, such as genes related to immunity, development and reproduction [101]. Taking the advantage of doubled haploid templates[102], Liu et al. (2012) used reads from RNA-seq to detect PSVs [48]. The idea is that there should be no SNPs from a single individual of the doubled haploid fish. If any "SNPs" are detected through informatics analysis, they would have to be derived from PSVs or MSVs due to gene duplications. A total of 2,659 duplicated genes were predicted in this study [48]. In spite of being very efficient, this approach had several limitations: 1) Duplicated genes with identical sequences would be missed; 2) expression patterns and levels limit the analysis of all genes; 3) sequence errors would lead to inflation of duplicated genes; 4) sequence variations derived from RNA editing would lead to the false identification of duplicated genes. In addition, the RNA-Seq approach provided no information as to the location and nature of gene duplications. In this study, we attempted to determine the duplicated genes using the draft genome sequences in channel catfish.

II. MATERIALS AND METHODS

mRNA prediction, clustering and identification of duplicated gene sets

In order to capture candidate mRNAs for the identification of duplicated gene sets in catfish, mRNAs were predicted from the catfish whole genome scaffolds by MOLQUEST 2.4 with FGENESH program. The predicted mRNA sequences were then used for BLASTX searches against the nr database with e-value $1E^{-5}$ for the generation of the gene list for further analysis (Figure 6). mRNA sequences with gene matches were then used as input sequences for cluster analysis using CD-HIT-EST program in CD-HIT v 4.6 [103], with band width of 20, sequence identity 0.85 (85%) and word length of 5. CD-HIT use a greedy algorithm that identifies a representative sequence for each cluster and assigns a new sequence to that cluster if it is sufficiently similar to the representative. Each cluster was considered as one duplicated gene set for our analysis.

Figure 6. Schematic presentation of analysis for duplicated genes in catfish.



Analysis for the nature of duplicated genes in catfish

In order to provide hints into the nature of duplicated genes, the duplicated genes were classified into three general categories: 1) Tandem duplications: if the duplicated gene copies are located on the same scaffold within 10 kb of one another, they were categorized as tandem duplications; 2) Intrachromosomal duplications: if the duplicated gene copies are located on different scaffold but on the same chromosome, or if they are located in the same scaffold but are separated by a distance greater than 10 kb between members, they were classified as intra-chromosomal duplication (Non-tandem); 3) Interchromosomal duplications: if the duplicated gene copies were located on different chromosomes, they were classified as inter-chromosomal duplications. In addition to these three categories, there are mixed types of duplications that included combinations of the above defined categories of duplications. Due to the low map resolution of the catfish linkage map, many duplicated genes cannot yet be classified, and they were placed into “non-tandem unassigned duplication” category.

Identification of the catfish species-specific duplicated genes

The representative mRNA sequences in each duplication cluster of catfish were used as query to perform BLASTX searches against protein sequences of duplicated genes from other fish species respectively, including zebrafish, medaka, stickleback and Tetraodon [27]. If a duplicated gene in catfish that exist as a single copy gene in all these fish species, the gene is then regarded as catfish species-specific duplicated gene.

Analysis of synonymous substitution (Ks) for duplicated gene pairs

In order to analyze the synonymous substitution mutation rate of duplicated genes in catfish, all the mRNA sequences in each duplication set were retrieved. For each pair of the homologs, their sequences were aligned with CLUSTALW v2.0 [104] and every pairwise alignments were then converted to axt format by ParaAT v1.0 [105]. The Ks values were calculated using the KaKs Calculator v2.0 [106] with maximum-likelihood model averaging algorithm [107].

Gene enrichment analysis of duplicated genes

In order to identify over-represented GO terms in the duplicated genes, GO enrichment analysis was performed by using Ontologizer 2.0 [108], using the Parent-Child-Intersection method with a Benjammini-Hochberg multiple testing correction [109]. GO terms for each duplicated gene were obtained by utilizing Uniprot annotations for the representative mRNA sequence in each cluster. The difference of the frequency of assignment of gene ontology terms in the different duplication set size were compared to the overall catfish annotated genes. The difference of the frequency of assignment of gene ontology terms in the species-specific duplicated genes were compared to the overall catfish duplicated genes. The threshold was set as FDR p value $\leqslant 0.05$.

III. RESULT

Generation of the duplicated gene pool

The very first question we were interested in is how many duplicated genes existed in the catfish genome. In order to answer this question, we first generated the duplicated gene pool. The draft whole genome assembly scaffolds (unpublished) were subjected to the analysis of gene sequences using MOLQUEST 2.4. A total of 82,837 mRNA sequences were initially identified. Due to the non-continuous nature of the assembly, exons of the same gene may be captured in one predicted mRNA sequence, or multiple predicted mRNA sequences. In order to generate the duplicated gene pool, we then conducted BLASTX searches. Of the 82,837 predicted mRNA sequences, 52,138 had significant hits that represented 26,896 unique genes (Table 1) The 52,138 mRNA sequences were then clustered by CD-HIT v 4.6 for further analysis that allowed identification of genes with two or more members. A total of 2,130 clusters containing 7,886 mRNA sequences were identified to contain two or more mRNA sequences that make up the duplicated gene pool for further analysis.

Table 1. Identify duplicated gene sets by cluster analysis of predicted mRNAs using CD-HIT v4.6

Category	**Number**
Total genome scaffold sequences	62,461
The number of predicted mRNA sequences	82,837
The number of predicted mRNA sequences with gene hits	52,138
The number of unique gene	26,896
Number of clusters containing two or more mRNAs	2,130

Analysis of duplicated copy numbers

The second question we were interested was the copy numbers of the duplicated genes. As shown in Table 2, the duplicated genes had highly variable copy numbers ranging from 2 to 303 in the catfish genome. Among these 2,130 putative duplicated gene sets, the vast majority (1,478 sets, 69.4%) had 2 members, followed by 255 sets with 3 copies, 149 sets with 4 copies, 58 sets with 5 copies, 40 sets with 6 copies, 23 with 7 copies, 24 with 8 copies, 13 with 9 copies, and 11 with 10 copies. The number of genes with very high gene copy numbers are relatively small, with 42 genes having 11-20 copies, 18 genes with 21-30 copies, 7 genes with 31-40 copies, 3 genes with 41-50 copies, 5 genes with 51-100 copies, and 4 genes with over 100 copies (Table 2). The gene with the highest copy number was reverse transcriptase gene with 303 copies, followed by a gene similar to the ORF2-encoded protein with 131 copies, a gene similar to the hypothetical protein LOC100535052 with 124 copies, and CD2-associated protein like gene with 121 copies (Supplement Table 1).

Table 2. Summary of gene copies in the duplicated gene sets

The size of duplication	**Number of sets**	**Percentage of genes**
2	1,478	69.4
3	255	12.0
4	149	7.0
5	58	2.7
6	40	1.9
7	23	1.1
8	24	1.1
9	13	0.6
10	11	0.5
11-20	42	2.0
21-30	18	0.8
31-40	7	0.3
41-50	3	0.1
51-100	5	0.2
>100	4	0.2
Total sets	2,130	100%

Comparison of duplication set size among catfish and model fish

Figure 7 showed the comparison of the relative distribution of duplication size among the duplicated gene sets of several fish species. Although similar overall distribution patterns of duplication size were observed among the five fish species under comparison, significant differences were noted, with catfish harboring the highest percentage (69.4%) of duplicated gene sets with only two copies, followed by Tetraodon (40.1%), stickleback (37.7%), medaka (37.5%), and zebrafish (32.8). Thus, over two thirds of the catfish duplicated gene sets had just two duplicates. As a result, the percentage of duplicated genes with three or more copies in catfish were much lower as compared with the other four fish species (Figure 7). In spite of minor spikes, overall, the percentage of gene sets at higher copy numbers was smaller and smaller at higher and higher copy numbers (Figure 7).

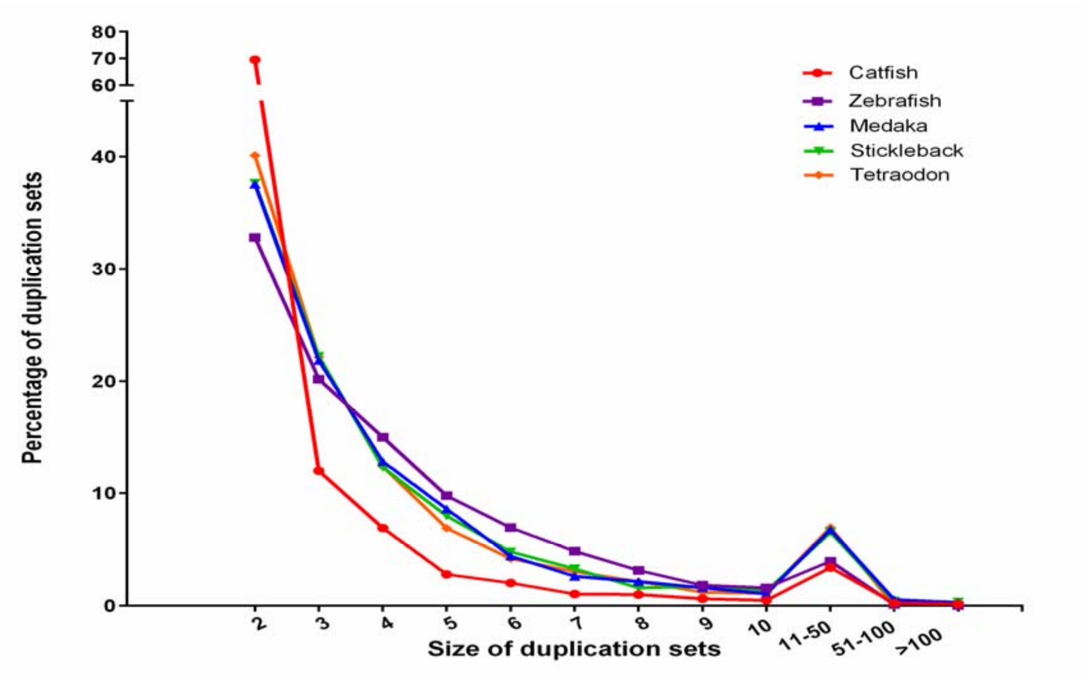


Figure 7. The distribution of different size of duplication sets among different fish species. Except the data of catfish, the data of other fish species were adopted from Lu et al., 2012 [27].

Characterization of duplicated genes

Duplicated genes were classified into three non-exclusive types based on the nature of duplications: 1) Tandem duplication was defined as duplicated gene copies located within 10 kb of each other along the same chromosome [27]. Intra-chromosomal duplication (Non-tandem) was defined as duplicated gene copies located on the same chromosome, but with a distance of greater than 10 kb between members; 3) Inter-chromosomal duplication was defined as duplicated gene copies located on different chromosomes. In some cases, duplicated genes sets may contain tandemly duplicated copies as well as additional copies elsewhere, or copies of a duplicated gene are present both intra- and inter-chromosomal, making them "mixed type" of duplications. In addition, due to the limitations of the catfish genetic linkage map, the chromosomal locations of many non-tandem duplicated genes are unknown at present, and they were categorized as "Non-tandem unassigned duplications".

Table 3. Summary of nature of duplicated genes

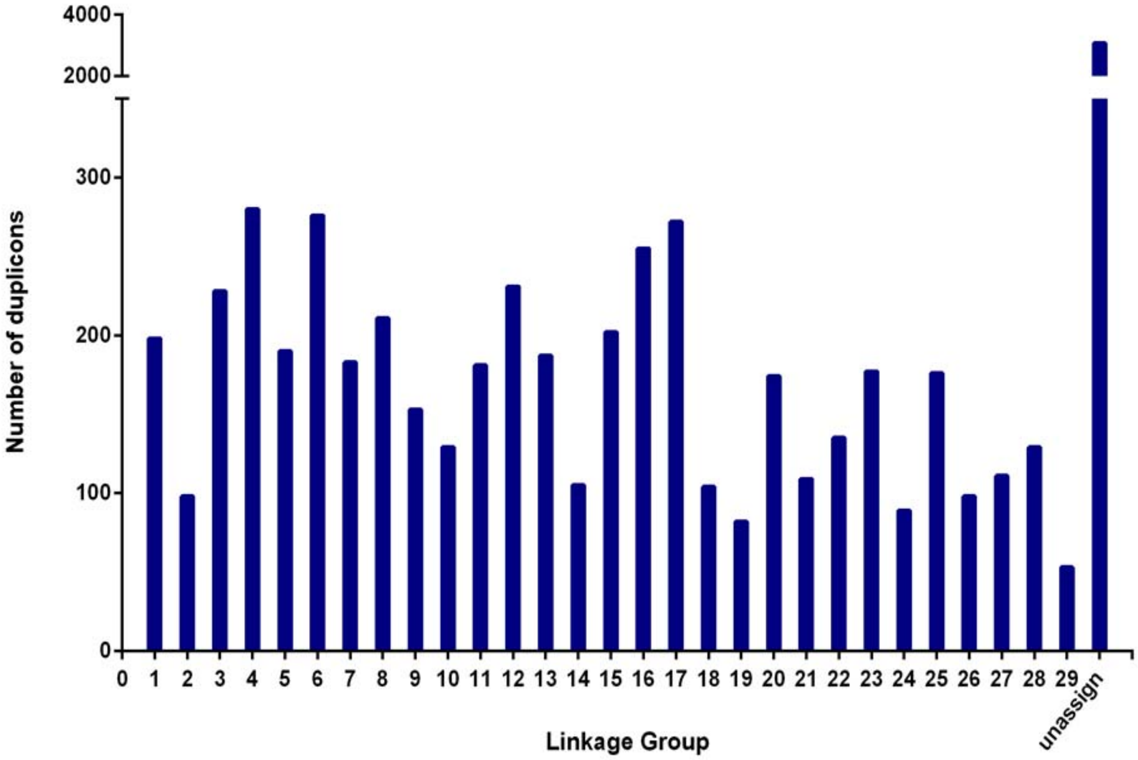
Catogery	Number	Percentage(%)
Tandom duplication sets	101	4.7
Inter-chromosomal duplication sets	586	27.5
Intra-chromosomal duplication sets	680	31.9
Mixed duplication sets	190	8.9
Non-tandem unassigned duplication sets	1,002	47.0

A total of 1,128 duplication sets can be characterized with their duplication types. Among them, Intra-chromosomal duplications account for most proportion, around 32.6% of all duplication sets with total 680 sets (Table 3). Inter-chromosomal duplications are the second prevalent types, accounting for around 27.5% of duplication sets with total 586 sets. Tandem duplications have the lowest proportion of the duplicated sets, accounting for around 4.7% with total 101 sets. The duplication sets with combination of several types account for around 10.3% with total 190 sets. In addition, 47% of duplication sets are classified into “Non-tandem unassigned duplications” types, because their chromosomal locations are yet unknown.

Distribution of duplicated genes on the 29 chromosomes

The distribution of duplicated genes on the 29 chromosomes are shown in Figure 8. It appeared that the distribution of duplicated genes are not evenly distributed among the 29 chromosomes of catfish, nor correlated with the chromosome size. Overall, the number of duplicated genes on the 29 linkage groups varied between 53 and 280, with LG4 containing the largest numbers of duplicated genes with a total of 280 duplicated genes, followed by LG6, LG17, LG16, LG3, and LG12, all of which had more than 200 duplicated genes. In contrast, LG29 contained the smallest numbers of duplicated genes with only 53 duplicated genes, followed by LG19, LG24, LG2, and LG26, all of which has fewer than 100 duplicated genes. In terms of chromosomal sizes, the physical size of the 29 catfish chromosomes are yet unknown at present. The number of duplicated genes within the linkage groups is not correlated with the genetic size as determined by recombination map size [110], although the physical size of the chromosomes may not be perfectly correlated with the genetic map size. For instance, LG17 was the smallest linkage group, but was one of the linkage groups containing the large numbers of duplicated genes.

Figure 8. Distribution of duplicated genes within 29 linkage groups (LG).

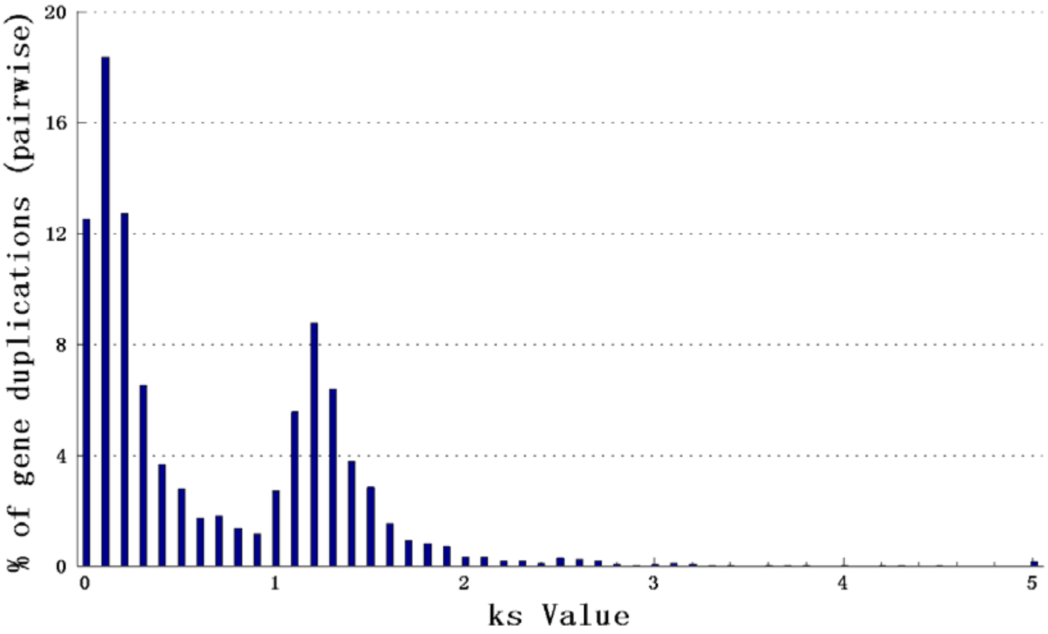


Analysis of Ks distance in duplicated genes

In order to determine the evolutionary rate of the duplicated genes, the Ks distance (synonymous substitutions/synonymous site) between pairs of homologs sequences were calculated. Approximately 62.8% duplicated genes (pairwise) had lower Ks value of ≤1.0 and 34.2% duplicated genes (pairwise) had Ks value between 1 and 2 (Figure 9) while only 3.0%

duplicated genes (pairwise) had Ks value greater than 2.0. The large number of duplicated genes with low Ks indicated that most of duplicated genes predicted in this study were recently duplicated genes. The distribution of all duplicated genes across varying Ks values showed two peaks with Ks value of 0.1 and 1.2 respectively, indicating two broad age categories of duplicated genes in catfish.

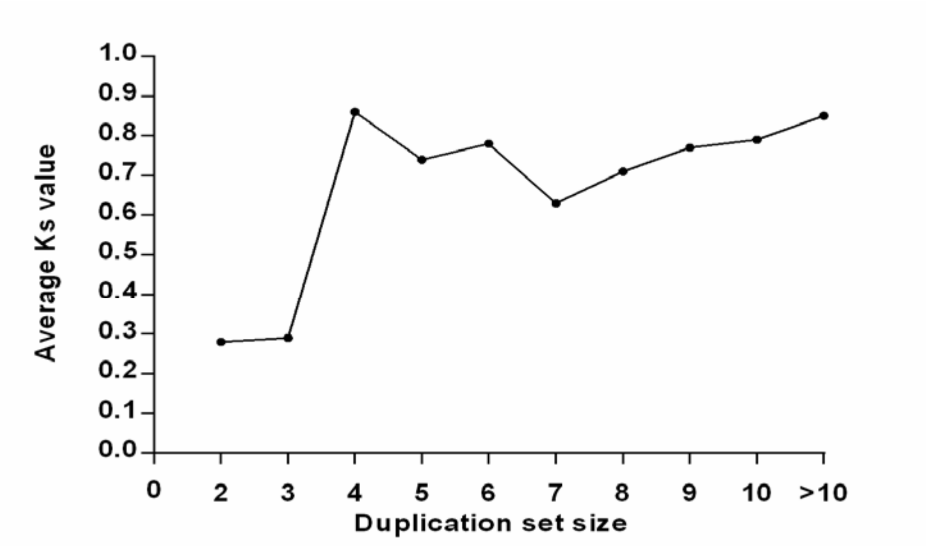
Figure 9. Distribution of duplicated genes along various Ks values.



In order to examine if Ks values are correlated with the duplication set size (copy numbers), the average Ks values within each duplication set size were plotted against the duplication set size. As shown in Figure 10, there was apparently a strong correlation of the average Ks values with the duplication set size. The duplication sets with two copies or three copies had relatively low Ks values while those with four or more copies had much greater Ks values. For instance, duplication sets with 2 copies and 3 copies had low Ks value of 0.28 and

0.29 respectively, while duplication sets with four or more copies had Ks values of 0.63 to 0.85, suggesting that genes with two or three copies were mostly derived from very recent gene duplications whereas genes with four or more copies were mostly derived from older gene duplications.

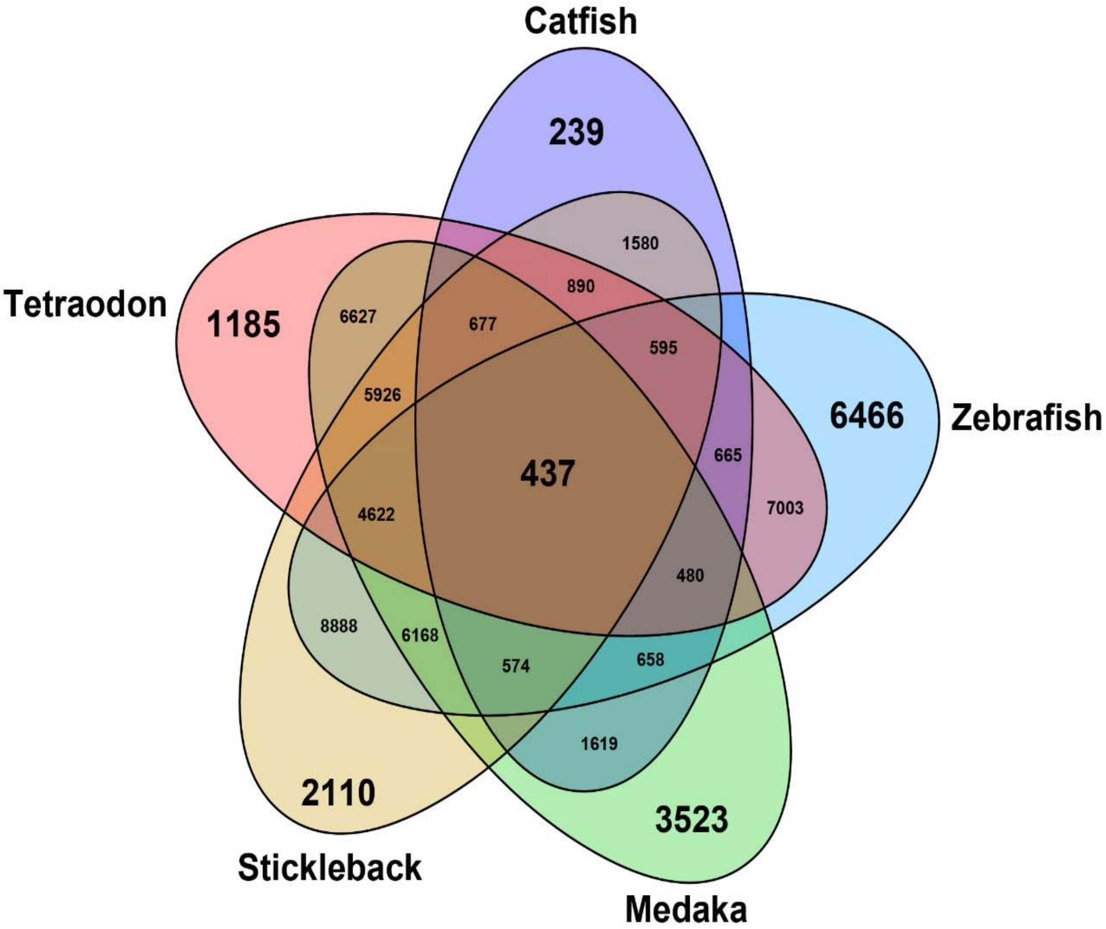
Figure 10. Relationship of average Ks values with duplication set size.



Evolutionary conservation of duplicated genes among teleost species

The 3R hypothesis suggested that teleost genomes went through one additional round of whole genome duplication after the split of teleost from the tetrapod. The duplicated gene sets at present is the resultant genes following random gene loss after the whole genome duplication event [79-81]. In addition, additional lineage-specific gene duplications may have occurred after the whole genome duplications. One interesting question is if the duplicated gene sets are conserved among various teleost species. As shown in Figure 11, catfish shared 1,341 duplicated genes sets with zebrafish, 1,619 sets with medaka, 1,580 sets with stickleback, and 1,264 sets with Tetraodon. There are 437 shared duplication sets among the five teleost species of catfish, zebrafish, medaka, stickleback and Tetraodon (Figure 11).

Figure 11. Venn diagram of duplicated genes among five species of fish, catfish, zebrafish, medaka, stickleback, and Tetraodon.



Catfish species-specific duplicated genes

While it was interesting to know the evolutionarily conserved duplicated gene sets among various sequenced fish genomes, it is even more interesting to know the catfish species-specific duplicated genes. To determine catfish species-specific duplicated genes, BLASTX searches were performed by using all catfish duplicated genes as query against duplicated gene database [27] from model fish respectively. When compared with the four model fish species, 789, 511, 550 and 866 genes were duplicated in catfish but not in zebrafish, medaka, stickleback and Tetraodon, respectively. A total of 239 genes were catfish-specific duplicated genes that are not duplicated in any of the four model fish species (Table 4).

Table 4. Number of unique duplicated gene in catfish when compared with other fish

	Zebrafish	**Medaka**	**Stickleback**	**Tetraodon**	**All the four fish**
Number	789	511	550	866	239

Of the 239 catfish specific duplications, the vast majority (170 duplication sets, 71.1%) had only two copies, followed by 24 duplication sets with three copies, 15 duplication sets with four copies, and 5 duplication sets with 5 copies (Table 5). There were 25 duplication sets with copies numbers greater than 5. However, it is noteworthy that some of the most highly duplicated genes were among the catfish-specific duplications. For instance, the gene "hypothetical protein LOC100535052" had a high copy number of 124, the gene "hypothetical protein

LOC100705656” had a high copy numbers of 89, novel protein similar to DNA polymerase had 58 copies, and dynein heavy chain 9 gene had 47 copies (Supplement Table 2).

Table 5. Summary of gene copies in the catfish specific duplications

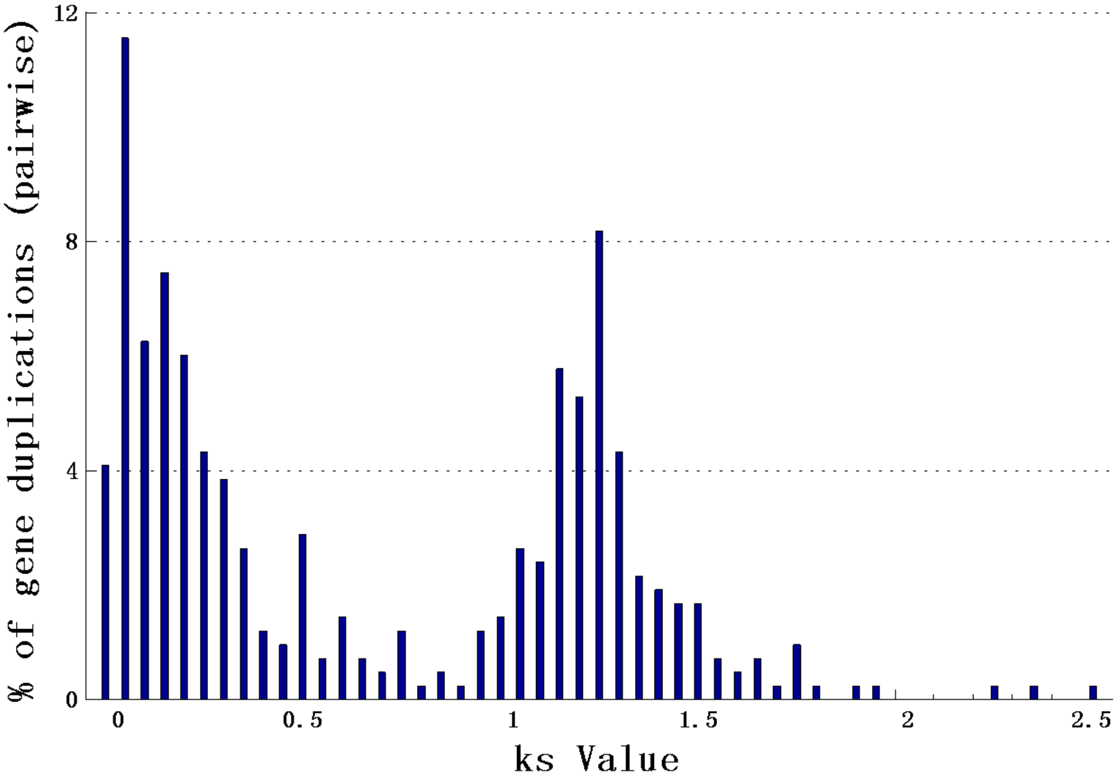
The size of duplication	**Number of sets**	**Percentage of genes**
2	170	71.1
3	24	10.0
4	15	6.3
5	5	2.1
≥6	25	10.5
Total sets	239	100%

Of the 239 catfish-specific duplicated genes, 13 were tandem duplications, 72 were intrachromosomal duplications, 72 were interchromosomal duplications, 21 were mixed type of duplications, and 109 were non-tandem unassigned duplications (Table 6). Analysis of Ks values of the catfish-specific duplicated genes indicated that the Ks values for these genes were lower than 2.5, and almost 60% of the catfish-specific duplicated genes (pairwise) had small Ks values of ≤ 1 (Figure 12), suggesting they were recent duplications. It is interesting that two peaks of Ks values were also observed with catfish-specific duplicated genes (Figure 12), one with very small Ks values peaking at 0.05 and 1.25, suggesting two periods with very high activities of gene duplications.

Table 6. Summary of nature of catfish specific duplicated genes

Catogery	Number	Percentage(%)
Tandem duplication sets	13	5.4
Inter-chromosomal duplication sets	72	30.1
Intra-chromosomal duplication sets	72	30.1
Mixed duplication sets	21	8.8
Non-tandem unassigned duplication sets	109	45.6

Figure 12. Distribution of catfish species-specific duplicated genes along various Ks values.



Functional enrichment analysis of the catfish duplication genes

In order to provide insight into the potential linkage between duplicated genes and functions, parent-child GO term enrichment analysis was performed to identify over-represented GO terms by using Ontologizer. A total of 36 terms with p-value (FDR-corrected) $\leqslant$ 0.05 were considered significantly over-represented, when using all the 2,130 duplicated genes as study group. The 36 GO terms were distributed across the three general categories provided by the GO annotation scheme: biological processes (BF), molecular functions (MF), and cellular components (CC) (Table 7). In the category of biological processes, the top three GO terms are purine-containing compound metabolic process, nucleobase-containing small molecule metabolic process and phosphorus metabolic process. In addition, the terms related to responses to stimulus and related to signal transduction were also over-represented in the catfish duplicated genes. Under the category of molecular function, the three most significant terms were nucleoside binding, nucleoside phosphate binding and catalytic activity. For cellular components, the three most significant terms were non-membrane-bounded organelle, intracellular non-membrane-bounded organelle and plasma membrane (Table 7).

Enrichment analyses were also performed in relation to the duplication set size to determine if the enriched GO terms among the duplicated genes were also correlated with copy numbers. A total of 20 GO terms are significant enriched in duplicated genes with 2 copies, which are distributed in two GO subdivision, biological processes and molecular functions (Table 8). Out of 20 terms, total of 15 significant GO terms belong to the biological processes which are all related to metabolic and catabolic processes and the term with most significant enrichment is nucleobase-containing small molecule metabolic process. In addition, total of 5 terms belong to molecular functions category and most significantly enriched term is catalytic activity. For duplicated genes with 3 members, only one term is significant enriched, which is protein-DNA complex in cellular components category. For duplicated genes with 4 members, two terms are significant enriched, which are the term antigen processing and presentation, and the term MHC protein complex. The two terms belong to biological processes and cellular components separately. A total of 27 terms are significant enriched in duplicated genes with 5-10 members, including terms related to various catabolic and metabolic processes as well as immune response. Out of the 27 terms, total of 19 terms belong to biological processes category, with most significant term being aromatic compound catabolic process in BP category. In addition, 3 out of 27 terms belong to molecular functions category with most significant term nucleoside binding. Total of 6 terms belong to cellular components with most significant term Non-membrane-bounded organelle. However, no GO is significantly enriched in duplicated genes with members above 10.

One of the major findings of the gene ontology analysis was that many genes related to immune response and stress response are highly enriched among the duplicated gene sets in catfish (Table 9). A total of 33 duplicated gene sets were related to immune responses, and 12 duplicated gene sets were related to stress responses. Although some of these duplicated gene sets were not statistically significantly enriched ($P>0.05$), their abundance in the duplicated gene sets may reflect importance of gene duplications related to function. Some examples of the duplicated immune response genes included CC chemokines, interleukins, metrix metalloproteinases, MHC genes, and TLR receptors. Some examples of the stress response duplicated genes included hemoglobins and members of the cytochrome P450 genes.

In order to examine the functional bias in catfish species-specific duplicated genes, the 239 catfish species-specific duplicated genes were used for GO enrichment analysis. Four terms were found to be significantly enriched, which are intracellular transport, cellular localization, intracellular part and intracellular (Table 10). Although gene enrichment analysis is useful in providing the initial insight into functional relationships with duplicated gene sets, detailed analysis of duplicated genes and their functions are necessary to understand the functional importance of the duplicated genes. Several categories of genes are particularly visible when catfish species-specific duplicated genes were analyzed. These included genes involved in immune and stress responses, intracellular transport, and cell signaling and interactions (Table 11).

Table 7. Gene ontology (GO) enriment analysis all duplicated genes. The significant GO ID and term ($p \leq 0.05$) are listed in each category

GO ID	**Go Term**	**GO Subdivision**	**P-value**
GO:0072521	Purine-containing compound metabolic process	BP	3.92E-07
GO:0055086	Nucleobase-containing small molecule metabolic process	BP	1.00E-06
GO:0006793	Phosphorus metabolic process	BP	2.22E-06
GO:0044270	Cellular nitrogen compound catabolic process	BP	4.78E-06
GO:0034655	Nucleobase-containing compound catabolic process	BP	4.78E-06
GO:0046700	Heterocycle catabolic process	BP	5.55E-06
GO:0019439	Aromatic compound catabolic process	BP	5.55E-06
GO:0019637	Organophosphate metabolic process	BP	9.95E-06
GO:1901361	Organic cyclic compound catabolic process	BP	9.95E-06
GO:1901564	Organonitrogen compound metabolic process	BP	1.99E-05
GO:1901135	Carbohydrate derivative metabolic process	BP	2.55E-04
GO:1901657	Glycosyl compound metabolic process	BP	2.93E-04
GO:0044248	Cellular catabolic process	BP	3.33E-04
GO:0044699	Single-organism process	BP	4.14E-04
GO:1901575	Organic substance catabolic process	BP	5.02E-04

GO:0044710	Single-organism metabolic process	BP	0.001324
GO:0009056	Catabolic process	BP	0.003211
GO:0050790	Regulation of catalytic activity	BP	0.006588
GO:0050896	Response to stimulus	BP	0.007952
GO:0019693	Ribose phosphate metabolic process	BP	0.008410
GO:0007165	Signal transduction	BP	0.008613
GO:1901658	Glycosyl compound catabolic process	BP	0.043623
GO:0006753	Nucleoside phosphate metabolic process	BP	0.048402
GO:0001882	Nucleoside binding	MF	1.00E-06
GO:1901265	Nucleoside phosphate binding	MF	4.26E-06
GO:0003824	Catalytic activity	MF	2.67E-05
GO:0043167	Ion binding	MF	3.17E-05
GO:0005488	Binding	MF	4.20E-04
GO:0016817	Hydrolase activity, acting on acid anhydrides	MF	0.001913
GO:0036094	Small molecule binding	MF	0.004372
GO:0043228	Non-membrane-bounded organelle	CC	0.002544
GO:0043232	Intracellular non-membrane-bounded organelle	CC	0.002544
GO:0005886	Plasma membrane	CC	0.003845
GO:0071944	Cell periphery	CC	0.018176
GO:0005623	Cell	CC	0.020583
GO:0044464	Cell part	CC	0.020583

Table 8. Gene ontology (GO) enriment analysis within different duplication size. The significant GO ID and term ($p<0.05$) are listed in each category

Duplication copies	**GO ID**	**Go Term**	**GO Subdivision**	**P-value**
2	GO:0055086	Nucleobase-containing small molecule metabolic process	BP	0.000269
	GO:0072521	Purine-containing compound metabolic process	BP	0.000464
	GO:0034655	Nucleobase-containing compound catabolic process	BP	0.001215
	GO:0044270	Cellular nitrogen compound catabolic process	BP	0.001215
	GO:0046700	Heterocycle catabolic process	BP	0.001319
	GO:0019439	Aromatic compound catabolic process	BP	0.001319
	GO:1901361	Organic cyclic compound catabolic process	BP	0.001319
	GO:0006793	Phosphorus metabolic process	BP	0.00175
	GO:1901564	Organonitrogen compound metabolic process	BP	0.002876
	GO:0019637	Organophosphate metabolic process	BP	0.00732
	GO:0050790	Regulation of catalytic activity	BP	0.019411
	GO:0044248	Cellular catabolic process	BP	0.036755
	GO:0044699	Single-organism process	BP	0.036755
	GO:1901657	Glycosyl compound metabolic process	BP	0.036755
	GO:1901135	Carbohydrate derivative metabolic process	BP	0.036755
	GO:0003824	Catalytic activity	MF	0.000269
	GO:0043167	Ion binding	MF	0.000269
	GO:1901265	Nucleoside phosphate binding	MF	0.000853
	GO:0001882	Nucleoside binding	MF	0.001319
	GO:0005488	Binding	MF	0.00732
3	GO:0032993	Protein-DNA complex	CC	0.004672
4	GO:0019882	Antigen processing and presentation	BP	0.031208
	GO:0042611	MHC protein complex	CC	0.00118
5-10	GO:0019439	Aromatic compound catabolic process	BP	0.001139
	GO:0046700	Heterocycle catabolic process	BP	0.001139
	GO:0044270	Cellular nitrogen compound catabolic process	BP	0.001829
	GO:1901361	Organic cyclic compound catabolic process	BP	0.002277
	GO:0006793	Phosphorus metabolic process	BP	0.002515
	GO:0034655	Nucleobase-containing compound catabolic process	BP	0.00252
	GO:1901575	Organic substance catabolic process	BP	0.00252
	GO:0072521	Purine-containing compound metabolic process	BP	0.006086
	GO:1901657	Glycosyl compound metabolic process	BP	0.01307
	GO:0019637	Organophosphate metabolic process	BP	0.013426
	GO:0009056	Catabolic process	BP	0.014734
	GO:0055086	Nucleobase-containing small molecule metabolic process	BP	0.019324
	GO:0044248	Cellular catabolic process	BP	0.020335
	GO:0044267	Cellular protein metabolic process	BP	0.027496
	GO:0060341	Regulation of cellular localization	BP	0.027496
	GO:0006955	Immune response	BP	0.034191
	GO:0043412	Macromolecule modification	BP	0.036432
	GO:0007017	Microtubule-based process	BP	0.043105

	GO:0032940	Secretion by cell	BP	0.045689
	GO:0001882	Nucleoside binding	MF	0.012533
	GO:0003774	Motor activity	MF	0.019324
	GO:0005200	Structural constituent of cytoskeleton	MF	0.020335
	GO:0043228	Non-membrane-bounded organelle	CC	0.001013
	GO:0043232	Intracellular non-membrane-bounded organelle	CC	0.001013
	GO:0016459	Myosin complex	CC	0.001829
	GO:0044430	Cytoskeletal part	CC	0.00252
	GO:0044446	Intracellular organelle part	CC	0.00252
>10		No significant GO enrichment		

Table 9. Duplicated genes with impartant function related to immune and stress response

Function	Gene Name	Copy Number
Immune response	Beta-arrestin-1	2
	CC chemokine SCYA110	2
	CC chemokine SCYA117	2
	CC chemokine SCYA120	5
	Dorsal-ventral patterning tolloid-like protein 1	2
	Interferon 4	3
	Interleukin 1 beta type b	2
	Interleukin 8	3
	Kit receptor tyrosine kinase a	2
	Macrophage erythroblast attacher	2
	Matrix metalloproteinase 13	2
	Matrix metalloproteinase 9	2
	MHC class I alpha chain	4
	MHC class II beta chain	4
	Neurobeachin-like protein 1	2
	Niemann-Pick C1 protein	2
	Nuclear receptor coactivator 2	7
	Phosphatidylinositol-3,4,5-trisphosphate 5-phosphatase 1-like	2
	Pim1 protein	4
	Protein kinase C beta type-like	6
	Protein numb homolog	2
	SAM domain and HD domain-containing protein 1	2
	Spectrin beta chain, erythrocyte	2
	Splicing factor 3A subunit 3-like	2
	Squamous cell carcinoma antigen recognised by T cells 3	2
	Stress-70 protein, mitochondrial-like	2
	TATA box-binding protein-like protein 2	3
	Tlr5	5
	TNF-related apoptosis inducing ligand TRAIL	2
	Toll-interacting protein	2
	Toll-like receptor 21	2
	Transducin beta-like protein 3-like	2
	Tumor necrosis factor receptor superfamily member 1A	2
Stress response	Heme oxygenase	6
	Hemoglobin subunit alpha	4
	Hemoglobin-beta	3
	Niemann-Pick C1 protein	2
	Syntenin-1	2
	T-complex protein 1 subunit delta	2
	Catalase	2
	Cryptochrome 1a	2
	Lon protease homolog, mitochondrial	2
	Per2 protein	2
	Peroxidasin-like	3
	Prostaglandin endoperoxide synthase-2	2

Table 10. Gene ontology (GO) enriment analysis of species-specific duplicated genes

GO ID	Go term	GO Subdivision	P-value
GO:0046907	intracellular transport	BP	0.009494
GO:0051641	cellular localization	BP	0.022052
GO:0044424	intracellular part	CC	2.93E-04
GO:0005622	intracellular	CC	0.020478

Table 11. Species-specific duplicated genes with important function in immune and stress response, intracellular transport and cell signaling

Function	**Gene Name**	**Copy Number**
Immune and stress responses	Catalase	2
	Chromatin assembly factor 1 subunit A	2
	DNA repair protein RAD50	3
	Interleukin 1 beta type b	2
	Interleukin-8	3
	Mitochondria-eating protein	2
	TFIIH basal transcription factor complex helicase XPB subunit	2
	Toll-interacting protein	2
Intracellular transport	Basic leucine zipper nuclear factor 1	2
	Exportin 6	2
	Intraflagellar transport protein 46	3
	Mannose-6-phosphate receptor	3
	Nucleoporin 107	2
	Nucleoporin like 1	2
	Syntaxin 17	6
	Vacuolar protein sorting-associated protein 16	2
Cell signaling and interaction	Breast carcinoma amplified sequence 3	3
	Cytochrome c-1	2
	Elongator complex protein 1	4
	General transcription factor IIIC, polypeptide 3, 102kda	2
	Glycine N-methyltransferase	2
	Listerin E3 ubiquitin protein ligase 1	2
	Low density lipoprotein receptor-related protein 6	2
	MAD1 mitotic arrest deficient-like	2
	Mediator complex subunit 16	2
	Microsomal triglyceride transfer protein	2
	Mitochondrial ribosomal protein L42	2
	NADH dehydrogenase (ubiquinone) flavoprotein 1, 51kda	2
	Negative elongation factor complex member B	3
	Polymerase (RNA) III (DNA directed) polypeptide C (62kd)	4
	RAB, member RAS oncogene family-like 6	2
	RNA (guanine-7-) methyltransferase	2
	Sequestosome 1	2
	Small nuclear ribonucleoprotein polypeptides B and B1	2
	Suppressor of Ty 16 homolog	2
	TRK-fused gene	2
	Ubiquitin fusion degradation 1 like	2

IV. DISCUSSIONS

Gene duplication is considered one of the most powerful forces in evolution [2, 6, 8], and it is widely believed that teleost fish have gone through an additional round of whole genome duplication. As a consequence, rapid genome rearrangement and gene loss occurred in the course of evolution after the whole genome duplication event. Teleost fish, therefore, may harbor genes at variable copies. In this study, we determined the duplicated genes in the catfish genome, and analyzed their copy numbers, their distribution in the genome among and on each chromosome. In addition, enrichment analysis was conducted to determine any connection between gene duplication levels and functions.

The vast majority, almost 70% of the catfish duplicated gene sets have two copies in the genome. This may suggest that the vast majority of duplicated genes may have been derived from the whole genome duplication event. However, detailed analysis suggested that the current existing duplicated genes may not be mainly derived from the whole genome duplication. For instance, while there were modest levels of recent tandem duplications, interchromosomal and non-tandem intrachromosomal duplications had almost equal shares of the duplicated gene sets (Table 3).

In this study, a total of 2,130 duplicated gene sets were identified in channel catfish. Compared with previous observation, the total duplication set number of catfish is smaller than that of zebrafish (3,991), Medaka (2,584), stickleback (2,669), but larger than that of *Tetradon* (2,020) [27]. Part of this difference may be due to the use of different methodologies. In this study, we used mRNA rather than protein sequences for the identification of the duplication pool. With amino acid sequences, many related genes may have been placed within a single duplication set, thus inflating the levels of duplication. For instance, TLR 2 and TLR 3 are apparently different genes with different functions, but they share domains with a relatively high identity. When amino acid sequences were used as used by Lu et al. (2012), these non-duplicated genes may be counted as duplications because they were brought into the duplicated gene pool. In this study, we initially predicted mRNA sequences from the catfish genome sequence scaffolds, followed by cluster analysis of related genes with a cut off level of 85% sequence similarity. There are following advantages of using predicted mRNA to identify duplicated genes, as compared with using protein sequences: 1) each predicted mRNA has unique position along the scaffold, which is useful for the analysis of the nature of duplications; 2) Relatively high stringent cut off values would avoid inflation of duplicated gene sets due to pooling of related genes into a single duplication set. However, our use of mRNA may underestimate the duplicated genes. The similarity cut off value was set at 85%, i.e., duplicated genes were identified only when their similarities were between 85%-100%. Any duplicated genes less than 85% similarities were excluded from the duplication pool.

Similar overall distribution patterns of duplication size were observed among the five fish species under comparison. However catfish contained the highest percentage (69.4%) of duplicated gene sets with only two copies and much lower the percentage of duplicated genes with three or more copies as compared with the other four fish species (Figure 7). This difference could be partially caused by the parameters used for the identification of the duplication sets (see above). Lu et al. (2012) used protein self-blast (cutoff $1e^{-5}$) followed Markov clustering algorithm to identify the duplication sets. Here in this study we used CD-HIT-EST (cutoff identity ≥ 85) to cluster predicted mRNA to identify the duplication sets. As discussed above, the duplication sets may have been inflated in Lu et al.'s study due to clustering of related different genes into a single duplication set. However, the differences in methodologies may have accounted only in part of the differences.

Analysis of Ks distance provided clues for the evolution rate of duplicated genes and the age of duplication in catfish. In catfish, a total of 62.8% duplicated genes (pairwise) had low Ks values of ≤1.0, indicating that most duplicated genes were recently duplicated genes. Previous study indicated that zebrafish had over 24.4% of duplicated genes with low Ks values of ≤1.0 and even smaller percentage of duplicated genes with low Ks value in medaka (1.3%), stickleback (0.97%) and Tetraodon (0.05%) [27]. Taken together, these results suggested that catfish had much more recently duplicated genes than other model fish species.

The distribution of all duplicated genes across varying Ks values indicated that there were two peaks around Ks values of 0.1 and 1.2 (Figure 9), which is different from the patterns in the four model fish species [27]. Although two peaks of Ks values were also observed in the model fish species, but they were around Ks values of 2 and 4 in all analyzed model fish species except zebrafish where a high peak of duplications was also observed with very low Ks values. It is possible that some duplicated genes with large Ks value may not be included in the present study due to the use of high stringency of parameters for identifying duplications in catfish, but it is also clear that catfish had two most recent peaks of gene duplications that are absent in the other four model fish species.

Many studies indicated that duplicated genes, especially recently duplicated genes, contribute to adaptation of the organism to various forms of environmental stress [111], such as extreme temperatures [112-113], nutrient limitations [114-115], response to various toxins [116-117], antibiotics [118], and pesticide treatments [119]. In order to provide clues for the connection of gene duplication with functionality, GO enrichment analysis was conducted with duplicated genes. A large number of GO terms were found to be over represented in the duplicated gene sets. For instance, genes involved in purine-containing compound metabolic process, nucleobase-containing compound catabolic process, phosphorus metabolic process, cellular nitrogen compound catabolic process, heterocycle catabolic process and aromatic compound catabolic process were significantly over represented (Table 7). In addition, the GO terms "response to stimulus" and "signal transduction" were also over-represented in catfish duplicated genes.

Many genes involved in immune and stress responses were among the duplicated gene sets in catfish (Table 9). These genes included chemokines, cytokines, metalloproteinases, MHC class I and class II genes, and Toll-like receptors. Although the direct relationship of gene duplication with immune and stress responses are yet to be established, many of these genes were previously found to play important roles in responding to bacterial infections with *E. ictaluri* and *F. columnare* in catfish, such as CC chemokine SCYA117, CC chemokine SCYA120, Interleukin 1 beta type b, Interleukin 8, MHC class I alpha chain, MHC class II beta chain, Toll-like receptor 21 and Catalase [120-122].

A total of 239 catfish specific duplicated genes were identified, with 71.1% of duplication sets containing 2 copies and 59.4% of duplicated genes (pairwise) had small Ks values of ≤ 1. These results suggested that catfish species-specific duplicated genes were mostly recent duplications. GO enrichment analysis of catfish-specific duplicated genes indicated that GO term "intracellular transport", "cellular localization", "intracellular part" and "intracellular" were over-represnted. Some of the genes included in the "intracellular transport" included mannose-6-phosphate receptor, syntaxin 17, and basic leucine zipper nuclear factor 1. Intracellular transport is fundamental for cellular function, survival and morphogenesis [123]. For instance, mannose-6-

phosphate receptor is responsible for transporting of phosphorylated lysosomal enzymes from the Golgi complex and the cell surface to lysosomes. It also acts as a positive regulator of T-cell coactivation by binding DPP4 [124]. Syntaxin 17 is an essential proteins for fusion of cellular membranes and also play a role in the early secretory pathway [125]. Basic leucine zipper nuclear factor 1 is responsible for protein transport from the endoplasmic reticulum (ER) to the cell surface [126].

The catfish species-specific duplicated genes also included a set of genes related to immune and stress responses (Table 11) such as interleukins, and Toll-interacting proteins. These species-specific duplications could be a mechanism for catfish to adapt to its environment. For instance, Interleukin 1 beta type b is an important cytokines which respond to infection and induce a cascade of reactions leading to inflammation [127]. In this study, we identified Interleukin 1 beta type b as a tandem duplication with 2 copies, as previously reported by Wang et al. (2005) [120]. Phylogenetic analysis indicated Interleukin 1 beta type b gene is a species-specific duplication, which also agreed with our observation. Both copies of Interleukin 1 beta type b responded to infection with *Edwardsiella ictaluri* in catfish and showed distinct expression profiles in various catfish tissues [120]. Interleukin-8 is another important pro-inflammatory cytokines which also respond to inflammation [128]. In the catfish with resistant to *F. columnare*, the expression of Interleukin-8 showed rapidly upregulated after infection [121]. The expression of Interleukin-8 also showed upregulated after infection of *Edwardsiella ictaluri* in channel and blue catfish [128]. In this study, Interleukin-8 are identified as intra-chromosomal duplication with 3 copies and are catfish-specific duplication. In addition, gene Catalase was also identified as catfish specific duplicated gene in this study, which was tandem duplication with 2 copies. Catalase is one of the key antioxidant enzymes, which is responsible for protection against immune infection and oxidative stress by decomposing hydrogen peroxide [129-130]. Upregulated expression of catalase was observed in resistance catfish after infection with *F. columnare* [121]. Other catfish specific duplicated genes were also important in cell signaling and interaction, such as Glycine N-methyltransferase [131], TRK-fused gene [132], and Mediator complex subunit 16 [133].

V. CONCLUSION

This study is the first study on surveying genome-wide duplicated genes in channel catfish. By using predicted mRNA, a total of 2,130 duplicated gene sets were identified in channel catfish, with high percentage of duplicated gene sets containing only two copies. The large number of duplicated genes with low Ks suggested that most of duplicated genes predicted in this study were recently duplicated genes. GO enrichment analysis indicated that GO terms associated to various metabolic and catabolic processes, response to stimuli and signal transduction were over-represented in duplicated genes. A total of 239 catfish specific duplicated genes were identified, which contained a set of genes related to immune and stress responses, suggesting a mechanism for catfish to adapt to its environment. This research is not only useful for further catfish whole genome annotation, but also provide clues of understanding environmental adaption from immunology and physiological aspect in channel catfish.

Supplement table 1. The list of duplicated genes in channel catfish.

Group	Copy number	NR id	Gene name
1	2	EFN62203.1	120.7 kda protein in NOF-FB transposable element
2	2	NP_001070178.2	12-lipoxygenase
3	2	ADF97628.1	14 kda apolipoprotein
4	2	XP_004649731.1	26S protease regulatory subunit 4
5	2	ADO28347.1	26S protease regulatory subunit s10b
6	2	NP_001156671.1	2'-deoxynucleoside 5'-phosphate N-hydrolase 1
7	2	XP_003757788.1	2-oxoglutarate dehydrogenase, mitochondrial isoform 4
8	2	NP_001187067.1	40S ribosomal protein S2
9	2	XP_003445580.1	4-hydroxyphenylpyruvate dioxygenase-like
10	2	NP_957327.2	5'-3' exoribonuclease 1
11	2	XP_004559292.1	80 kda MCM3-associated protein-like
12	2	1VJY_A	A chain a, crystal structure of a naphthyridine inhibitor of human tgf-beta type i receptor
13	2	NP_956714.1	A disintegrin and metalloprotease domain 10b precursor
14	2	XP_001919066.1	A disintegrin and metalloproteinase with thrombospondin motifs 18
15	2	XP_004550220.1	Abhydrolase domain-containing protein 15-like
16	2	NP_001116845.1	Abnormal spindle-like microcephaly-associated protein
17	2	XP_003451707.1	Acetyl-coa carboxylase 2
18	2	NP_001264046.1	Acetyl-coenzyme A synthetase, cytoplasmic
19	2	XP_003446076.1	Acetyl-coenzyme A synthetase, cytoplasmic-like
20	2	NP_001002452.1	Acid phosphatase 5b, tartrate resistant precursor
21	2	NP_999956.1	Acid-sensing ion channel 1
22	2	ACO09534.1	ACN9 protein homolog, mitochondrial precursor
23	2	XP_003961633.1	Actin filament-associated protein 1-like 2-like
24	2	NP_001090226.1	Actin, alpha skeletal muscle 2
25	2	NP_853632.3	Actin, cytoplasmic 2
26	2	NP_001007773.1	Actin-related protein 2
27	2	XP_001919038.3	Activated CDC42 kinase 1-like
28	2	XP_003454919.1	Activating signal cointegrator 1 complex subunit 3
29	2	XP_002667626.1	Activin receptor type-2B-like
30	2	AAI52247.1	Adal protein
31	2	XP_002733696.1	Adaptor-related protein complex 2, beta 1 subunit-like
32	2	AAI55778.1	Adaptor-related protein complex 3, mu 1 subunit
33	2	NP_001161821.1	Adenylate cyclase type 1
34	2	NP_001137224.1	Adenylate cyclase type 8
35	2	NP_001166036.1	Adenylate kinase 7
36	2	XP_001919948.2	A-kinase anchor protein 6
37	2	NP_001077015.1	Aldehyde dehydrogenase 18A1
38	2	NP_001073423.2	Aldehyde dehydrogenase family 16 member A1
39	2	AAH66668.1	Aldh9a1a protein

40	2	CAL18610.1	Alpha 2,3-sialyltransferase ST3Gal I-r4
41	2	CAL18607.1	Alpha 2,6-sialyltransferase ST6GalNAc I
42	2	NP_001020705.2	Alpha globin regulatory element containing-like
43	2	NP_956919.1	Alpha-1,2-mannosyltransferase ALG9
44	2	XP_001922207.2	Alpha-1-antiproteinase 2-like
45	2	XP_004571257.1	Alpha-2 adrenergic receptor-like
46	2	NP_955880.1	Alpha-actinin-4
47	2	ACQ58328.1	Alpha-enolase
48	2	NP_001005390.1	Alpha-ketoglutarate-dependent dioxygenase alkb homolog 6
49	2	XP_002661434.1	Alpha-N-acetylneuraminide alpha-2,8-sialyltransferase-like
50	2	XP_002662342.2	Alsin isoform X1
51	2	XP_003446648.1	Alsin-like
52	2	NP_001166124.1	Amylo-1, 6-glucosidase, 4-alpha-glucanotransferase
53	2	XP_693879.5	Amyloid beta A4 precursor protein-binding family A member 1
54	2	XP_002927784.1	Amyloid beta A4 precursor protein-binding family A member 2-like
55	2	XP_005025270.1	AN1-type zinc finger protein 4 isoform X1
56	2	XP_004575247.1	Angio-associated migratory cell protein-like
57	2	XP_691871.5	Angiomotin
58	2	XP_002666042.2	Angiopoietin-related protein 5
59	2	Q1LVW0.2	Ankyrin repeat and BTB/POZ domain-containing protein BTBD11-A; altname: Full=BTB/POZ domain-containing protein 11-A
60	2	XP_003198948.1	Ankyrin repeat and death domain-containing protein 1A-like
61	2	NP_001119867.1	Ankyrin repeat and LEM domain-containing protein 2
62	2	XP_004770732.1	Ankyrin repeat and SAM domain-containing protein 1A-like isoform X9
63	2	XP_004542869.1	Ankyrin repeat and SOCS box protein 10-like isoform X2
64	2	XP_696789.5	Ankyrin repeat and SOCS box protein 14-like isoform X2
65	2	XP_002662990.2	Ankyrin repeat domain-containing protein 11
66	2	CBN81678.1	Ankyrin-3
67	2	XP_003448779.1	Ankyrin-3-like
68	2	NP_001244037.1	Annexin a1
69	2	XP_003201441.1	AP-3 complex subunit beta-2-like
70	2	EMP31323.1	Apolipoprotein A-I-binding protein
71	2	NP_955834.2	Apoptosis inhibitor 5
72	2	XP_003452039.1	Apoptosis-stimulating of p53 protein 2-like
73	2	XP_003439111.1	Arf-GAP with coiled-coil, ANK repeat and PH domain-containing protein 3-like
74	2	XP_001921526.2	Arf-GAP with gtpase, ANK repeat and PH domain-containing protein 1
75	2	XP_688145.5	Arf-GAP with Rho-GAP domain, ANK repeat and PH domain-containing protein 2-like
76	2	XP_687312.3	Arf-GAP with Rho-GAP domain, ANK repeat and PH domain-containing protein 3 isoform X2
77	2	XP_002187508.1	Arginine-glutamic acid dipeptide repeats protein
78	2	XP_688618.3	Armadillo repeat-containing protein 3

79	2	NP_001165565.1	Arrestin red cell
80	2	XP_004069735.1	Arylsulfatase A-like
81	2	XP_695808.3	Asc-type amino acid transporter 1
82	2	NP_998222.1	Aspartate aminotransferase, cytoplasmic
83	2	XP_003197930.1	Astrotactin-1
84	2	AAI63272.1	Atf7b protein
85	2	NP_001006016.1	Atlastin-3
86	2	Q9PTY0.1	ATP synthase subunit beta, mitochondrial; Flags: Precursor
87	2	NP_956512.1	ATP synthase subunit s, mitochondrial
88	2	XP_003441308.1	Atpase family AAA domain-containing protein 3-A-like
89	2	XP_001920862.3	ATP-binding cassette sub-family A member 1
90	2	XP_686632.5	ATP-binding cassette sub-family A member 12
91	2	NP_001017544.1	ATP-binding cassette sub-family B member 8, mitochondrial
92	2	NP_001167126.1	ATP-binding cassette sub-family G member 2
93	2	XP_687685.2	ATP-binding cassette sub-family G member 4 isoform 1
94	2	NP_001002649.1	ATP-citrate synthase
95	2	NP_001188373.1	ATP-dependent RNA helicase A
96	2	NP_001184253.1	ATP-dependent RNA helicase DDX1
97	2	NP_775336.1	ATP-dependent RNA helicase DDX55
98	2	XP_697933.5	ATP-dependent RNA helicase Dhx29
99	2	XP_003198936.1	ATP-sensitive inward rectifier potassium channel 10-like
100	2	XP_696565.5	AT-rich interactive domain-containing protein 1A
101	2	XP_003200444.1	B(0,+)-type amino acid transporter 1-like isoform X1
102	2	XP_001334285.4	BAH and coiled-coil domain-containing protein 1-like
103	2	AAI62155.1	Bardet-Biedl syndrome 1
104	2	NP_001091715.1	B-cell receptor CD22 precursor
105	2	XP_003199193.1	BDNF/NT-3 growth factors receptor
106	2	XP_692598.4	Beta-1,3-glucosyltransferase-like
107	2	NP_001002175.1	Beta-1-syntrophin
108	2	NP_001128197.1	Beta-adrenergic receptor kinase 2
109	2	NP_001153294.1	Beta-arrestin-1
110	2	NP_999888.1	Beta-enolase
111	2	NP_001003853.1	Beta-galactoside alpha-2,6-sialyltransferase 1
112	2	NP_998049.1	Beta-lactamase-like protein 2
113	2	XP_004563856.1	Beta-parvin-like isoform X1
114	2	XP_003216036.1	Bifunctional aminoacyl-trna synthetase-like isoform 2
115	2	XP_003455263.1	Bifunctional heparan sulfate N-deacetylase/N-sulfotransferase 1-like
116	2	NP_001017887.1	Bis(5'-adenosyl)-triphosphatase enpp4 precursor
117	2	AAH55578.1	Blzf1 protein
118	2	XP_004066703.1	Bone morphogenetic protein receptor type-2-like
119	2	XP_001922208.1	BR serine/threonine-protein kinase 2-like
120	2	XP_003199686.1	Brain-specific angiogenesis inhibitor 1-associated protein 2-like
121	2	XP_003200273.1	Brain-specific angiogenesis inhibitor 1-like, partial

122	2	NP_001107092.1	Breast cancer anti-estrogen resistance protein 3
123	2	XP_001922198.2	Bromodomain and PHD finger-containing protein 3 isoform X1
124	2	NP_001089560.1	BTB (POZ) domain containing 10
125	2	EMC86350.1	BTB/POZ domain-containing protein 6
126	2	ADO27936.1	Btg1
127	2	ELK00187.1	Bullous pemphigoid antigen 1, isoforms 6/9/10
128	2	XP_003973271.1	Bystin-like
129	2	XP_004561690.1	C2 domain-containing protein 2-like isoform X1
130	2	F1Q8W0.1	Cactin; Short=zcactin
131	2	NP_001009884.1	CAD protein
132	2	XP_003439228.1	Cadherin-12
133	2	NP_999974.1	Cadherin-23 precursor
134	2	XP_003442792.1	Cadherin-4-like
135	2	EMP28743.1	Cadherin-related family member 1
136	2	NP_001004010.1	Calcitonin gene-related peptide type 1 receptor precursor
137	2	NP_957000.1	Calcium and integrin-binding family member 2
138	2	NP_001139622.1	Calcium channel, voltage-dependent, L type, alpha 1S subunit, a
139	2	NP_001003458.1	Calcium homeostasis modulator 2
140	2	XP_004572881.1	Calcium-binding mitochondrial carrier protein scamc-2-B-like isoform X3
141	2	XP_004540070.1	Calpain-3-like isoform X2
142	2	NP_001076532.2	Calpain-like
143	2	ADX33320.1	Calreticulin
144	2	XP_004320607.1	Calsequestrin-2
145	2	XP_004081844.1	Calsequestrin-2-like isoform 1
146	2	XP_683242.5	Camp-specific 3',5'-cyclic phosphodiesterase 4B, partial
147	2	XP_004564869.1	Camp-specific 3',5'-cyclic phosphodiesterase 4D-like isoform X7
148	2	WP_006729443.1	Carbohydrate-binding protein
149	2	NP_001187680.1	Carbonic anhydrase 7
150	2	XP_694982.2	Carbonic anhydrase 9
151	2	NP_001070801.1	Carbonic anhydrase XV c precursor
152	2	XP_002665524.1	Carcinoembryonic antigen-related cell adhesion molecule 5
153	2	XP_693888.4	Cartilage acidic protein 1-like
154	2	XP_002664490.2	Casein kinase I isoform alpha isoform X1
155	2	XP_003199651.1	Caskin-2
156	2	XP_004075421.1	Caspase-1-like
157	2	XP_004612805.1	Caspase-6-like
158	2	AGM34043.1	Caspase-9
159	2	NP_570987.1	Catalase
160	2	NP_001180580.1	Catenin delta-2
161	2	NP_001187379.1	Cathepsin K precursor
162	2	NP_001187996.1	Cathepsin L precursor
163	2	ABA54957.1	CC chemokine SCYA110

164	2	ABA54964.1	CC chemokine SCYA117
165	2	AAI55265.1	Ccne2 protein
166	2	NP_001082810.1	CCR4-NOT transcription complex subunit 11
167	2	XP_697509.2	CCR4-NOT transcription complex subunit 4-like
168	2	NP_001008583.2	CD2-associated protein
169	2	NP_001187191.1	CD3 zeta precursor
170	2	XP_001332626.4	CD97 antigen-like
171	2	NP_001007404.1	Cdc42-interacting protein 4 homolog
172	2	XP_003442763.1	CDK5 regulatory subunit-associated protein 1-like
173	2	XP_004556684.1	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase, mitochondrial-like isoform X1
174	2	NP_001133237.1	Cell division protein kinase 9
175	2	NP_001107081.1	Centromere protein H
176	2	XP_697409.4	Centromere protein J
177	2	XP_002663091.2	Centrosomal protein of 120 kda
178	2	NP_001188107.1	Charged multivesicular body protein 2a
179	2	XP_001332348.1	Chemokine XC receptor 1
180	2	XP_003451769.1	Chemokine-like receptor 1-like
181	2	AAI63083.1	Chka protein
182	2	BAA78329.1	Chmadrin (short type)
183	2	XP_004575094.1	Chondroitin sulfate synthase 3-like
184	2	NP_001038478.2	Chromatin assembly factor 1 subunit A
185	2	XP_003452731.1	Chymotrypsin-like elastase family member 2A-like
186	2	XP_003199092.1	Citron Rho-interacting kinase-like
187	2	AFL70283.1	Class II major histocompatibility complex transactivator
188	2	XP_693753.5	Clathrin coat assembly protein AP180-like
189	2	XP_002667655.2	Claudin-20-like
190	2	XP_003971031.1	Claudin-4-like
191	2	NP_001108611.1	CLIP-associating protein 1
192	2	AAH76079.1	CNDP dipeptidase 2 (metallopeptidase M20 family)
193	2	NP_001092248.1	Coagulation factor II (thrombin) receptor-like 1 precursor
194	2	NP_001188180.1	Coiled-coil domain-containing protein 124
195	2	NP_001082851.1	Coiled-coil domain-containing protein 79
196	2	NP_001035414.1	Coiled-coil domain-containing protein 93
197	2	XP_003200359.1	Coiled-coil domain-containing protein c6orf97
198	2	XP_694040.5	Collagen alpha-1(IV) chain
199	2	XP_003445052.1	Collagen alpha-1(VII) chain-like
200	2	XP_002663415.2	Collagen alpha-3(VI) chain
201	2	XP_004572221.1	Collagen alpha-3(VI) chain-like
202	2	CAD54661.1	Collagen type XI alpha 2
203	2	BAF63404.1	Cone transducin alpha subunit
204	2	XP_001919009.1	Connector enhancer of kinase suppressor of ras 2 isoform X1
205	2	XP_003446534.1	Constitutive coactivator of PPAR-gamma-like protein 1-like

206	2	XP_001338473.4	Contactin associated protein-like 5
207	2	NP_956098.1	Contactin-5
208	2	XP_003783998.1	Coproporphyrinogen-III oxidase, mitochondrial
209	2	XP_003198104.1	CREB-binding protein-like
210	2	XP_001489820.2	Crooked neck-like protein 1-like isoform 1
211	2	NP_001070765.2	Cryptochrome 1a
212	2	XP_700009.2	CTD small phosphatase-like protein 2-like
213	2	XP_004573747.1	Cubilin-like
214	2	XP_003201787.1	Cullin-9-like
215	2	AAO16601.1	Cyclic nucleotide gated channel
216	2	NP_001139240.1	Cyclin-dependent kinase-like 5 isoform 1
217	2	ADO27795.1	Cystatin-b
218	2	XP_004066644.1	Cysteine and tyrosine-rich protein 1-like
219	2	CAN88772.1	Cysteine conjugate-beta lyase; cytoplasmic (glutamine transaminase K, kyneurenine aminotransferase)
220	2	NP_001018407.1	Cystinosin precursor
221	2	CBN81058.1	Cytochrome b-c1 complex subunit 1, mitochondrial
222	2	Q92088.1	Cytochrome P450 2M1; altname: Full=CYPIIM1; altname: Full=LMC1; altname: Full=Lauric acid omega-6-hydroxylase
223	2	ADO19749.1	Cytochrome P450 3A
224	2	ACI33584.1	Cytochrome P450 4F3
225	2	BAN16583.1	Cytochrome P450 family 8 subfamily b polypeptide 1-1
226	2	NP_001018658.1	Cytochrome P450 precursor
227	2	XP_004860889.1	Cytohesin-1 isoform X3
228	2	XP_697830.5	Cytohesin-2
229	2	XP_003199223.1	Cytohesin-interacting protein-like
230	2	NP_001036210.1	Cytoplasmic dynein 1 heavy chain 1
231	2	XP_003459594.1	Cytoplasmic dynein 1 heavy chain 1-like isoform 1
232	2	XP_001339287.3	Cytoplasmic dynein 1 intermediate chain 1-like, partial
233	2	ADO27757.1	Cytoplasmic dynein 2 light intermediate chain 1
234	2	XP_003970541.1	Cytoplasmic FMR1-interacting protein 2-like isoform 1
235	2	XP_002666445.2	Cytoplasmic polyadenylation element-binding protein 2 isoform X1
236	2	XP_692712.4	Cytoskeletal protein Sojo
237	2	NP_001074047.2	Cytosolic endo-beta-N-acetylglucosaminidase
238	2	XP_693755.4	Cytosolic phospholipase A2 zeta-like
239	2	NP_898890.1	D(3) dopamine receptor
240	2	XP_004571177.1	DC-STAMP domain-containing protein 2-like
241	2	ELK32047.1	DDB1- and CUL4-associated factor 15
242	2	NP_001189401.1	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 32
243	2	ABG91568.1	Death receptor 6
244	2	CBN81567.1	Dedicator of cytokinesis protein 1
245	2	XP_004545493.1	Dedicator of cytokinesis protein 2-like
246	2	XP_004082983.1	Dedicator of cytokinesis protein 4-like
247	2	XP_002661310.1	Dedicator of cytokinesis protein 7

248	2	XP_001920312.3	Dedicator of cytokinesis protein 9
249	2	XP_003442551.1	Dehydrogenase/reductase SDR family member 7C-B-like
250	2	XP_683226.3	Deleted in lung and esophageal cancer protein 1-like
251	2	XP_004076636.1	Delta(14)-sterol reductase-like
252	2	XP_004574152.1	Delta(3,5)-Delta(2,4)-dienoyl-coa isomerase, mitochondrial-like, partial
253	2	ACN11415.1	DENN domain-containing protein 2D
254	2	XP_004065699.1	Deoxycytidylate deaminase-like
255	2	NP_001070920.1	Desmin
256	2	XP_001922445.2	Diacylglycerol kinase zeta
257	2	NP_001187387.1	Diamine N-acetyltransferase 1
258	2	AGN95665.1	Diet1
259	2	NP_001018353.1	Dihydropyrimidinase-related protein 2
260	2	XP_003199094.1	Dihydropyrimidinase-related protein 2-like
261	2	XP_004540414.1	Diphthine--ammonia ligase-like
262	2	XP_002663338.1	Disco-interacting protein 2 homolog A-like
263	2	NP_998128.1	Disco-interacting protein 2 homolog B-A
264	2	NP_001025307.1	Disheveled-associated activator of morphogenesis 1
265	2	XP_004078972.1	Disks large homolog 1-like
266	2	NP_001139075.1	Disks large homolog 5
267	2	XP_003974815.1	Dmx-like protein 1-like
268	2	XP_002920469.1	Dmx-like protein 2-like
269	2	XP_002665161.2	DNA helicase MCM8-like isoform X1
270	2	NP_001119860.1	DNA ligase 1
271	2	NP_001121995.1	DNA polymerase epsilon catalytic subunit A
272	2	XP_686065.4	DNA polymerase sigma
273	2	NP_001133624.1	DNA polymerase subunit alpha B
274	2	XP_002665692.2	DNA polymerase zeta catalytic subunit
275	2	XP_004566328.1	DNA polymerase zeta catalytic subunit-like
276	2	NP_957438.1	DNA repair and recombination protein RAD54-like
277	2	XP_004067195.1	DNA repair protein complementing XP-G cells homolog
278	2	NP_001006081.1	DNA replication complex GINS protein PSF1
279	2	NP_001001838.1	DNA topoisomerase I, mitochondrial precursor
280	2	XP_001919588.2	DNA-dependent protein kinase catalytic subunit
281	2	XP_001922839.2	DNA-directed RNA polymerase I subunit RPA1
282	2	XP_004079921.1	DNA-directed RNA polymerase II subunit RPB1-like
283	2	NP_001188119.1	DNA-directed RNA polymerase III subunit rpc10
284	2	AAH49406.1	Dnaj (Hsp40) homolog, subfamily C, member 3
285	2	XP_002667092.2	Dnaj homolog subfamily C member 11-like
286	2	XP_004551997.1	Dnaj homolog subfamily C member 13-like isoform X1
287	2	XP_003978958.1	Dnaj homolog subfamily C member 25-like
288	2	XP_004070906.1	Dnaj homolog subfamily C member 5-like
289	2	ADP89572.1	Dopamine D2 receptor

290	2	EAW48672.1	Dopey family member 1, isoform CRA_a
291	2	NP_571085.1	Dorsal-ventral patterning tolloid-like protein 1 precursor
292	2	XP_003198125.1	Double C2-like domain-containing protein alpha-like
293	2	AAU89440.1	Doublesex and mab-3 related transcription factor 3
294	2	XP_003447952.1	Dual specificity mitogen-activated protein kinase kinase 7-like
295	2	XP_701125.5	Dymeclin-like, partial
296	2	XP_003227355.1	Dynein heavy chain 1, axonemal-like
297	2	XP_003200184.1	Dynein heavy chain 5, axonemal-like, partial
298	2	XP_004574595.1	Dynein heavy chain domain-containing protein 1-like
299	2	ACN10102.1	Dysbindin
300	2	BAM10995.1	Dysferlin
301	2	NP_001184229.1	E2F transcription factor 5 isoform 1
302	2	XP_003445419.1	E3 SUMO-protein ligase ranbp2
303	2	XP_700123.5	E3 ubiquitin/ISG15 ligase TRIM25-like
304	2	XP_004010935.1	E3 ubiquitin-protein ligase ARIH1
305	2	XP_003444981.1	E3 ubiquitin-protein ligase ARIH2-like isoform 2
306	2	XP_001923094.2	E3 ubiquitin-protein ligase Hakai-like isoform X1
307	2	XP_003448547.1	E3 ubiquitin-protein ligase HECTD3-like
308	2	XP_001923900.1	E3 ubiquitin-protein ligase HUWE1 isoform 2
309	2	XP_001340400.3	E3 ubiquitin-protein ligase listerin
310	2	NP_001083011.1	E3 ubiquitin-protein ligase RFWD2
311	2	XP_003457758.1	E3 ubiquitin-protein ligase RNF123
312	2	XP_004559461.1	E3 ubiquitin-protein ligase RNF213-like isoform X2
313	2	XP_004079095.1	E3 ubiquitin-protein ligase RNF220-like
314	2	NP_001001943.1	E3 ubiquitin-protein ligase SMURF1
315	2	XP_683951.5	E3 ubiquitin-protein ligase TRIM33-like
316	2	XP_003977898.1	E3 ubiquitin-protein ligase TRIM39-like
317	2	XP_003447432.1	E3 ubiquitin-protein ligase UBR3-like
318	2	XP_693147.3	E3 ubiquitin-protein ligase UBR4 isoform X11
319	2	XP_004077674.1	E3 ubiquitin-protein ligase UBR5-like isoform 2
320	2	XP_004538114.1	E3 ubiquitin-protein ligase UHRF2-like isoform X1
321	2	XP_003200485.1	Early endosome antigen 1
322	2	XP_004935407.1	Echinoderm microtubule associated protein like 4 isoform X6
323	2	NP_001002419.1	Ectonucleoside triphosphate diphosphohydrolase 4
324	2	NP_001135181.1	Ectonucleoside triphosphate diphosphohydrolase 6
325	2	XP_002664034.2	Ectonucleotide pyrophosphatase/phosphodiesterase family member 7-like
326	2	NP_957261.1	EF-hand domain-containing protein 1
327	2	XP_003452041.1	EGF-containing fibulin-like extracellular matrix protein 1-like
328	2	XP_003197798.1	EGF-like module-containing mucin-like hormone receptor-like 1-like
329	2	AER39526.1	Egl nine 1-like protein
330	2	XP_003444774.1	EH domain-containing protein 1
331	2	AAI42794.1	Eif4ba protein

332	2	XP_687688.5	Electrogenic sodium bicarbonate cotransporter 4
333	2	XP_002798491.1	ELKS/Rab6-interacting/CAST family member 1-like
334	2	NP_001002371.1	Elongation factor 1-alpha 2
335	2	XP_699102.2	Emilin-3
336	2	XP_692516.3	Endoplasmic reticulum aminopeptidase 1 isoform X3
337	2	NP_001116524.1	Endoplasmic reticulum aminopeptidase 2 precursor
338	2	NP_957186.1	Endoplasmic reticulum resident protein 44
339	2	NP_998256.1	Engulfment and cell motility protein 1
340	2	ABB52638.1	Envelope polyprotein
341	2	NP_001112369.1	Envoplakin
342	2	NP_001005919.1	Ephrin type-A receptor 4a precursor
343	2	XP_001344560.4	Ephrin type-A receptor 6
344	2	XP_004076390.1	Ephrin type-B receptor 4-like
345	2	AAH55594.1	Ephx1 protein, partial
346	2	NP_956536.1	Epidermal growth factor receptor kinase substrate 8
347	2	XP_002665258.2	Epithelial cell-transforming sequence 2 oncogene-like
348	2	EGW05689.1	Epsin-1
349	2	XP_693770.3	ER membrane protein complex subunit 1 isoform X2
350	2	AEO89688.1	Ercc2
351	2	NP_956039.2	ESF1 homolog
352	2	NP_001187548.1	Ester hydrolase c11orf54-like protein
353	2	ADO28051.1	Ets domain-containing protein elk-3
354	2	NP_958868.1	Eukaryotic peptide chain release factor subunit 1
355	2	NP_001265821.1	Eukaryotic translation initiation factor 2-alpha kinase 3-like precursor
356	2	XP_003441279.1	Eukaryotic translation initiation factor 3 subunit A-like
357	2	XP_003442557.1	Eukaryotic translation initiation factor 3 subunit D-like
358	2	NP_001073669.1	Eukaryotic translation initiation factor 4 gamma 1
359	2	NP_956273.1	Excitatory amino acid transporter 2
360	2	NP_001008400.1	Exostosin-2
361	2	AAH44132.1	Exportin 6
362	2	XP_003441490.1	Exportin-2
363	2	XP_699731.5	Extended synaptotagmin-1
364	2	XP_003198886.1	Extended synaptotagmin-1-like
365	2	NP_001116705.1	Extended synaptotagmin-3
366	2	XP_004563576.1	FACT complex subunit SPT16-like
367	2	NP_001188001.1	Fam18b
368	2	NP_001134170.1	Fam82b
369	2	NP_001070780.2	Fanconi anemia, complementation group I
370	2	NP_998561.1	Fasciculation and elongation protein zeta-1
371	2	NP_001116988.1	Fascin homolog 2, actin-bundling protein, retinal (Strongylocentrotus purpuratus) a
372	2	NP_001158785.1	FAST kinase domain-containing protein 1
373	2	XP_004566143.1	Fatty acyl-coa reductase 1-like isoform X2

374	2	NP_001070215.1	F-box only protein 36
375	2	XP_003445831.1	F-box/LRR-repeat protein 3-like
376	2	NP_998484.1	F-box/WD repeat-containing protein 5
377	2	XP_003444342.1	F-box-like/WD repeat-containing protein TBL1XR1-like
378	2	XP_692355.4	Fc receptor-like protein 5-like
379	2	XP_001922278.1	Fer-1-like protein 4-like
380	2	XP_004546346.1	FERM domain-containing protein 4B-like isoform X4
381	2	XP_003928260.1	FERM, rhogef and pleckstrin domain-containing protein 1
382	2	ACI33779.1	Fermitin family homolog 3
383	2	ADO28147.1	Ferritin middle subunit
384	2	AAI62517.1	Fgfr3 protein
385	2	XP_001920354.2	Fibroblast growth factor 11-like
386	2	XP_004542677.1	Fibroblast growth factor 12-like isoform X1
387	2	NP_878276.1	Fibroblast growth factor 8b precursor
388	2	XP_004067255.1	Fibulin-7-like
389	2	XP_004546009.1	Filamin-A-like isoform X2
390	2	XP_698846.5	Filamin-C, partial
391	2	CAQ14318.1	Fintrim family protein
392	2	NP_001187811.1	Flap endonuclease 1-b
393	2	AAI14255.1	Flj13639 protein
394	2	XP_004547040.1	Follistatin-A-like isoform X2
395	2	NP_001134013.1	Follistatin-related protein 1 precursor
396	2	XP_003975342.1	Forkhead box C1-A-like
397	2	XP_001922273.1	Forkhead box protein J3
398	2	NP_001070725.2	Forkhead box protein O1-A
399	2	XP_690538.5	Formin-binding protein 1-like
400	2	NP_956505.2	Fragile X mental retardation syndrome-related protein 2
401	2	XP_003450311.1	Fucolectin-1-like
402	2	XP_004547114.1	Fumarylacetoacetate hydrolase domain-containing protein 2-like isoform X1
403	2	XP_003201276.1	FYVE and coiled-coil domain-containing protein 1-like
404	2	XP_688794.3	FYVE, rhogef and PH domain-containing protein 1, partial
405	2	XP_004568307.1	FYVE, rhogef and PH domain-containing protein 4-like isoform X2
406	2	XP_001920703.1	G protein-activated inward rectifier potassium channel 1-like isoform X1
407	2	NP_001003992.1	G1 to S phase transition 1
408	2	XP_004562124.1	G2/M phase-specific E3 ubiquitin-protein ligase-like isoform X4
409	2	AAH45459.1	GA repeat binding protein, beta 2
410	2	BAC82608.1	Gag-like protein
411	2	NP_956709.1	Gamma-glutamyltranspeptidase 1
412	2	NP_001018511.1	Ganglioside-induced differentiation-associated protein 1
413	2	XP_004071603.1	Gap junction gamma-1 protein-like
414	2	XP_002662207.2	GAS2-like protein 2-like
415	2	XP_003769093.1	Gastricsin-like

416	2	XP_697820.3	GDH/6PGL endoplasmic bifunctional protein-like
417	2	XP_693326.4	GDNF family receptor alpha-2-like
418	2	XP_003437831.1	Gelsolin-like
419	2	NP_001243171.1	General transcription factor 3C polypeptide 3
420	2	EFN62669.1	General transcription factor II-I repeat domain-containing protein 2B
421	2	XP_001978729.1	Gg19746
422	2	NP_001187736.1	Glia maturation factor beta
423	2	ADO28093.1	Globoside alpha-13-n-acetylgalactosaminyltransferase 1
424	2	AAI63091.1	Glucuronyl C5-epimerase b
425	2	NP_001121809.1	Glutamate [NMDA] receptor subunit epsilon-2
426	2	XP_003453758.1	Glutamate receptor 4-like isoform 1
427	2	XP_003458295.1	Glutamate receptor, ionotropic kainate 2-like
428	2	XP_005022094.1	Glutamine and serine-rich protein 1
429	2	NP_001187282.1	Glutathione S-transferase kappa 1
430	2	ADO27976.1	Glycine n-acyltransferase
431	2	NP_997981.1	Glycine N-methyltransferase
432	2	XP_004546800.1	Glycine--trna ligase-like
433	2	XP_004543375.1	Glypican-6-like isoform X2
434	2	NP_001187640.1	GMP reductase 2
435	2	ADO27762.1	Gpn-loop gtpase 1
436	2	NP_957116.2	G-protein coupled receptor 143
437	2	NP_001018314.1	G-protein coupled receptor 182
438	2	XP_004564084.1	G-protein coupled receptor 64-like
439	2	NP_001187138.1	Granzyme-like I precursor
440	2	NP_001003430.1	Grifin
441	2	XP_683507.4	GRINL1A complex locus protein 1
442	2	NP_956272.1	Growth arrest-specific protein 6 precursor
443	2	XP_003800891.1	Growth arrest-specific protein 8
444	2	XP_004080481.1	Growth hormone-inducible transmembrane protein-like
445	2	NP_001129727.2	GTP cyclohydrolase 1
446	2	XP_002661284.1	Gtpase IMAP family member 7-like isoform X1
447	2	XP_002664713.2	Gtpase IMAP family member 8
448	2	NP_001002420.2	Gtpase-activating protein and VPS9 domain-containing protein 1
449	2	XP_004067107.1	GTP-binding protein RAD-like
450	2	XP_005009131.1	Guanine nucleotide-binding protein G(o) subunit alpha-like
451	2	XP_004538254.1	Guanine nucleotide-binding protein G(z) subunit alpha-like
452	2	BAH28868.1	Guanylate cyclase activating protein 4
453	2	BAH28864.1	Guanylate cyclase retinal cone
454	2	BAH28862.1	Guanylate cyclase retinal rod1
455	2	AAK95948.1	Guanylate cyclase-activating protein 2
456	2	XP_004542884.1	Guanylate kinase-like
457	2	NP_001187735.1	Guanylin precursor
458	2	NP_571856.1	Gylcine receptor, beta a precursor

459	2	XP_003964536.1	H(+)/Cl(-) exchange transporter 7-like
460	2	NP_001099062.1	Hairy and enhancer of split-related 15.2
461	2	NP_001018465.1	#Name?
462	2	BAI68365.1	Hatching enzyme
463	2	ADX32514.1	Heat shock 70 kda protein
464	2	AGO01991.1	Heat shock 70kda protein 12a
465	2	NP_999881.1	Heat shock protein 4a
466	2	XP_004548558.1	Helicase ARIP4-like isoform X1
467	2	XP_004028111.1	Hemicentin-1
468	2	NP_001187948.1	Hemoglobin subunit beta
469	2	NP_001074908.1	Heparan sulfate (glucosamine) 3-O-sulfotransferase 1-like1 precursor
470	2	XP_004558410.1	Heparan sulfate glucosamine 3-O-sulfotransferase 3B1-like
471	2	NP_957316.1	Hepatic triacylglycerol lipase precursor
472	2	XP_004836836.1	Hepatitis A virus cellular receptor 1 homolog
473	2	XP_004558585.1	Hepatocyte nuclear factor 1-alpha-like
474	2	NP_571955.3	Hepatocyte nuclear factor 1-beta-A
475	2	XP_003968508.1	High affinity cgmp-specific 3',5'-cyclic phosphodiesterase 9A-like
476	2	ABX79926.1	Histidine-isoleucine/vasoactive intestinal polypeptide
477	2	NP_001004586.1	Histidyl-trna synthetase, cytoplasmic
478	2	XP_003746420.1	Histone H3-like
479	2	NP_998589.2	Histone H4 transcription factor
480	2	XP_004542769.1	Histone-lysine N-methyltransferase MLL3-like isoform X7
481	2	XP_001920852.3	Histone-lysine N-methyltransferase SETD1A isoform X1
482	2	XP_003418047.1	Histone-lysine N-methyltransferase SUV39H1-like
483	2	XP_004570733.1	Homeobox protein SIX3-like
484	2	AAI71514.1	Hypothetical LOC559844
485	2	PC1123	Hypothetical protein - bloodfluke planorb (fragment)
486	2	ABD77536.1	Hypothetical protein
487	2	WP_009782202.1	Hypothetical protein
488	2	XP_001432014.1	Hypothetical protein
489	2	WP_019505193.1	Hypothetical protein
490	2	XP_808638.1	Hypothetical protein
491	2	XP_002590453.1	Hypothetical protein BRAFLDRAFT_124569
492	2	XP_002605224.1	Hypothetical protein BRAFLDRAFT_126603
493	2	EGT32644.1	Hypothetical protein CAEBREN_29370
494	2	ELU18471.1	Hypothetical protein CAPTEDRAFT_205533
495	2	XP_002648163.1	Hypothetical protein CBG24246
496	2	YP_875591.1	Hypothetical protein censya_0654
497	2	XP_003091566.1	Hypothetical protein CRE_23180
498	2	EPQ05527.1	Hypothetical protein D623_10002145
499	2	EFX77420.1	Hypothetical protein DAPPUDRAFT_106000
500	2	EKX44718.1	Hypothetical protein GUITHDRAFT_152949, partial
501	2	NP_965518.1	Hypothetical protein LJ1711

502	2	XP_001922778.2	Hypothetical protein LOC100000228
503	2	XP_001342589.3	Hypothetical protein LOC100002920
504	2	XP_001344878.1	Hypothetical protein LOC100005990
505	2	XP_001345932.3	Hypothetical protein LOC100007493
506	2	XP_001371662.2	Hypothetical protein LOC100018428
507	2	XP_001922115.2	Hypothetical protein LOC100148674
508	2	XP_001920945.2	Hypothetical protein LOC100150385
509	2	XP_001922042.3	Hypothetical protein LOC100151502
510	2	XP_002663560.2	Hypothetical protein LOC100329499
511	2	XP_002667279.1	Hypothetical protein LOC100330441
512	2	XP_002662656.2	Hypothetical protein LOC100331153
513	2	XP_002667295.2	Hypothetical protein LOC100332327
514	2	XP_002804019.1	Hypothetical protein LOC100424774
515	2	XP_003355087.2	Hypothetical protein LOC100516319
516	2	XP_003200508.1	Hypothetical protein LOC100534899
517	2	XP_003200544.1	Hypothetical protein LOC100535066
518	2	XP_003200965.1	Hypothetical protein LOC100535282
519	2	XP_003200750.1	Hypothetical protein LOC100536356
520	2	XP_003197720.1	Hypothetical protein LOC100536880
521	2	XP_003199622.1	Hypothetical protein LOC100536966
522	2	XP_003199656.1	Hypothetical protein LOC100537292
523	2	XP_003198681.1	Hypothetical protein LOC100537769
524	2	XP_003200691.1	Hypothetical protein LOC100537950
525	2	XP_003197739.1	Hypothetical protein LOC100538131
526	2	XP_003452947.1	Hypothetical protein LOC100690590
527	2	XP_003458186.1	Hypothetical protein LOC100693304
528	2	XP_003453461.1	Hypothetical protein LOC100695593
529	2	XP_003458826.1	Hypothetical protein LOC100696576, partial
530	2	XP_003455041.1	Hypothetical protein LOC100701310
531	2	XP_003459749.1	Hypothetical protein LOC100704041
532	2	XP_003444290.1	Hypothetical protein LOC100705845
533	2	XP_003454068.1	Hypothetical protein LOC100706773
534	2	XP_003451285.1	Hypothetical protein LOC100712215
535	2	XP_001921879.2	Hypothetical protein LOC327226
536	2	XP_003200393.1	Hypothetical protein LOC449979
537	2	XP_003200870.1	Hypothetical protein LOC497165
538	2	XP_686084.4	Hypothetical protein LOC557848
539	2	XP_687443.5	Hypothetical protein LOC559049
540	2	XP_691744.4	Hypothetical protein LOC563288
541	2	XP_003198146.1	Hypothetical protein LOC565612
542	2	XP_696437.5	Hypothetical protein LOC568036
543	2	XP_697383.3	Hypothetical protein LOC568932
544	2	XP_001920157.2	Hypothetical protein LOC568975

545	2	XP_701002.4	Hypothetical protein LOC572217
546	2	XP_001090048.2	Hypothetical protein LOC701749
547	2	XP_001332295.4	Hypothetical protein LOC792710
548	2	XP_001336831.3	Hypothetical protein LOC796490
549	2	XP_001923294.3	Hypothetical protein LOC798565
550	2	ELR52716.1	Hypothetical protein M91_08923
551	2	YP_003010191.1	Hypothetical protein Pjdr2_1430
552	2	EIE80716.1	Hypothetical protein RO3G_05421
553	2	EIE83083.1	Hypothetical protein RO3G_07788
554	2	EIE88243.1	Hypothetical protein RO3G_12954
555	2	EIE88575.1	Hypothetical protein RO3G_13286
556	2	EIE91154.1	Hypothetical protein RO3G_15865
557	2	EFA13300.1	Hypothetical protein tcasga2_TC014830
558	2	XP_002116174.1	Hypothetical protein TRIADDRAFT_60185
559	2	XP_002959926.1	Hypothetical protein VOLCADRAFT_39527
560	2	CCD18005.1	Hypothetical protein, conserved in T. Vivax
561	2	ELW69772.1	Ice nucleation protein
562	2	ABR66915.1	IGFI receptor A
563	2	XP_004567675.1	Ig-like V-type domain-containing protein FAM187A-like isoform X1
564	2	AAH91832.1	Im:7148034 protein, partial
565	2	AAH91841.1	Im:7149628 protein, partial
566	2	AAL01590.1	Immune-type receptor 10
567	2	XP_004557967.1	Immunoglobulin superfamily member 2-like
568	2	XP_004547588.1	Immunoglobulin-like and fibronectin type III domain-containing protein 1-like isoform X1
569	2	EPQ05864.1	Importin subunit alpha-3
570	2	NP_001153133.1	Importin-11 isoform 1
571	2	XP_684163.4	Importin-4
572	2	NP_001037799.2	Importin-8
573	2	XP_003451951.1	Inactive dipeptidyl peptidase 10-like
574	2	XP_004569896.1	Inactive phospholipase C-like protein 1-like
575	2	XP_004067237.1	Inactive serine/threonine-protein kinase VRK3-like
576	2	NP_001077323.1	Indoleamine 2,3-dioxygenase 2
577	2	XP_004541709.1	Inhibitor of Bruton tyrosine kinase-like
578	2	Q4G3H4.1	Inhibitor of nuclear factor kappa-B kinase subunit alpha; Short=I kappa-B kinase alpha; Short=IKK-A; Short=IKK-alpha; Short=ikbka; Short=ikappab kinase; altname: Full=Conserved helix-loop-helix ubiquitous kinase; altname: Full=I-kappa-B kinase 1; Short=IK
579	2	XP_003967192.1	Inosine-5'-monophosphate dehydrogenase 1a-like
580	2	XP_003967352.1	Inositol 1,4,5-trisphosphate receptor type 2-like
581	2	XP_418035.2	Inositol 1,4,5-trisphosphate receptor type 3 isoform X2
582	2	XP_003199958.1	Inositol polyphosphate 5-phosphatase OCRL-1
583	2	XP_002662615.2	Integral membrane protein GPR180-like, partial
584	2	XP_690622.4	Integrator complex subunit 1 isoform 2

585	2	XP_695559.4	Integrin alpha-11
586	2	XP_003199352.1	Integrin alpha-2-like
587	2	XP_004574684.1	Integrin alpha-M-like
588	2	XP_003438689.1	Integrin alpha-X-like
589	2	NP_001177238.1	Integrin, alpha 3b precursor
590	2	XP_003448159.1	Inter-alpha-trypsin inhibitor heavy chain H2-like
591	2	NP_001020335.1	Inter-alpha-trypsin inhibitor heavy chain H4 precursor
592	2	XP_001331201.2	Inter-alpha-trypsin inhibitor heavy chain H5
593	2	XP_004558559.1	Interferon-induced protein 44-like, partial
594	2	NP_001187149.1	Interleukin 1 beta type b
595	2	XP_003440990.1	Interleukin-1 receptor accessory protein-like 1-B-like
596	2	XP_688141.3	Interleukin-17 receptor D
597	2	NP_001107790.1	Interleukin-6 receptor subunit alpha
598	2	NP_001002312.1	Intraflagellar transport protein 172 homolog
599	2	XP_001341428.4	Inward rectifier potassium channel 16-like
600	2	XP_004546426.1	IQ motif and SEC7 domain-containing protein 2-like isoform X3
601	2	NP_963881.1	Islet cell autoantigen 1-like protein
602	2	NP_001134853.1	Isoamyl acetate-hydrolyzing esterase 1 homolog
603	2	CBN80863.1	Isoleucyl-trna synthetase, cytoplasmic
604	2	AAI62468.1	Jag1a protein
605	2	XP_002664097.2	Janus kinase and microtubule-interacting protein 3-like isoform X1
606	2	XP_003199221.1	Kalirin-like
607	2	NP_001007432.1	Katanin p60 atpase-containing subunit A-like 1
608	2	XP_001334231.1	Kelch repeat and BTB domain-containing protein 7
609	2	NP_001070808.1	Kelch-like protein 24
610	2	XP_004550823.1	Kell blood group glycoprotein-like
611	2	NP_001003445.1	Keratin 12
612	2	XP_699336.4	Kinase D-interacting substrate of 220 kda
613	2	XP_001919049.2	Kinase suppressor of Ras 2-like
614	2	XP_004554756.1	Kinesin light chain 1-like isoform X1
615	2	XP_003441094.1	Kinesin light chain 4-like isoform 1
616	2	AAF14560.1	Kinesin-like protein 2
617	2	NP_998353.1	Kinesin-like protein KIF20A
618	2	NP_001004664.2	Kinesin-like protein KIF2A
619	2	NP_001093615.1	Kinesin-like protein KIF3B
620	2	ABK32085.1	Kit receptor tyrosine kinase a
621	2	ADO28375.1	Lactose-binding lectin l-2
622	2	ADO28056.1	Lambda-crystallin-like protein
623	2	XP_004916584.1	Lamina-associated polypeptide 2, isoforms alpha/zeta-like
624	2	XP_004550574.1	Laminin subunit alpha-2-like isoform X7
625	2	XP_003977653.1	Laminin subunit alpha-3-like
626	2	AAI63717.1	Laminin, alpha 4
627	2	XP_005025483.1	La-related protein 6-like isoform X1

628	2	A3KPW9.1	Large proline-rich protein BAG6; altname: Full=BCL2-associated athanogene 6; altname: Full=HLA-B-associated transcript 3
629	2	XP_003450388.1	Latrophilin-3-like
630	2	XP_689096.4	Leishmanolysin-like peptidase
631	2	XP_686927.3	Leucine zipper protein 1
632	2	XP_003200463.1	Leucine-rich repeat and calponin homology domain-containing protein 3
633	2	XP_001337163.3	Leucine-rich repeat and fibronectin type-III domain-containing protein 6-like, partial
634	2	NP_001034730.1	Leucine-rich repeat LGI family member 2 precursor
635	2	XP_002667476.2	Leucine-rich repeat serine/threonine-protein kinase 1-like
636	2	NP_001188385.1	Leucine-rich repeat serine/threonine-protein kinase 2
637	2	XP_001918318.2	Leucine-rich repeat-containing protein 1
638	2	XP_003439802.1	Leucine-rich repeat-containing protein 16B-like
639	2	XP_004570446.1	Leucine-rich repeat-containing protein 45-like
640	2	XP_001921374.3	Leucine-rich repeat-containing protein 56-like
641	2	XP_001920812.3	Leucine-rich repeat-containing protein 7-like
642	2	NP_001070848.1	Leucine-rich repeat-containing protein 8D
643	2	NP_001018537.1	Leucine-rich repeat-containing protein c10orf11
644	2	ABI23570.1	Leukocyte immune-type receptor TS32.15 L2.3c, partial
645	2	ABI16049.1	Leukocyte immune-type receptor TS32.17 L1.1a, partial
646	2	ABI23578.1	Leukocyte immune-type receptor TS32.17 L2.1a, partial
647	2	ACA79941.1	Leukocyte tyrosine kinase protein
648	2	CAG33687.1	Lgmn
649	2	NP_956620.1	Light ear protein
650	2	XP_004633327.1	LIM and senescent cell antigen-like-containing domain protein 1
651	2	XP_004567081.1	Limbic system-associated membrane protein-like isoform X3
652	2	XP_004606161.1	Lipoxygenase homology domain-containing protein 1 isoform X2
653	2	AFO11025.1	Liver carnitine palmitoyltransferase 1A-2
654	2	AAI71666.1	LOC449616 protein
655	2	AAH90185.1	LOC553294 protein, partial
656	2	AAI15338.1	LOC555678 protein
657	2	AAI39714.1	LOC559236 protein
658	2	AAI34902.1	LOC566865 protein
659	2	AAI25873.1	LOC568719 protein
660	2	AAI55325.1	LOC797572 protein
661	2	AAI29405.1	LOC799852 protein
662	2	ACN10954.1	Lon protease homolog, mitochondrial precursor
663	2	NP_001119851.1	Long-chain-fatty-acid--coa ligase ACSBG2
664	2	XP_003975267.1	Long-chain-fatty-acid--coa ligase ACSBG2-like
665	2	XP_001920591.2	Low-density lipoprotein receptor-related protein 1 isoform X1
666	2	XP_004069178.1	Low-density lipoprotein receptor-related protein 1-like
667	2	NP_001128156.1	Low-density lipoprotein receptor-related protein 6 precursor
668	2	XP_002662830.1	Lymphocyte antigen 75

669	2	BAF56885.1	Lymphocyte-specific protein tyrosine kinase
670	2	NP_001121799.1	Lysine (K)-specific demethylase 5Ba
671	2	XP_004869536.1	Lysophospholipid acyltransferase 5 isoform X2
672	2	XP_001921957.1	Lysosomal alpha-glucosidase
673	2	XP_003447883.1	Lysyl oxidase homolog 3-like
674	2	NP_001186799.1	Lysyl oxidase-like 3a precursor
675	2	NP_956613.2	Macoilin-1
676	2	NP_955843.1	Macrophage erythroblast attacher
677	2	XP_001332621.1	Macrophage mannose receptor 1-like protein 1-like isoform 1
678	2	NP_001117075.1	Main olfactory receptor-like protein
679	2	NP_001017667.1	Major facilitator superfamily domain-containing protein 10
680	2	NP_001002315.1	Malcavernin
681	2	NP_001120335.1	Malic enzyme 1, NADP(+)-dependent, cytosolic
682	2	XP_004549663.1	Maltase-glucoamylase, intestinal-like isoform X1
683	2	AAH57463.1	Mars protein
684	2	XP_003965464.1	MARVEL domain-containing protein 2-like
685	2	NP_001187264.1	Matrix metalloproteinase 13 precursor
686	2	NP_001187157.1	Matrix metalloproteinase-9 precursor
687	2	XP_002664410.2	Matrix-remodeling-associated protein 5
688	2	XP_004546388.1	Matrix-remodeling-associated protein 8-like isoform X2
689	2	NP_001164210.1	MAX-interacting protein
690	2	NP_001133848.1	Mediator complex subunit 16
691	2	NP_001116083.1	Membrane-associated guanylate kinase, WW and PDZ domain-containing protein 2
692	2	XP_004546430.1	Membrane-associated guanylate kinase, WW and PDZ domain-containing protein 2-like isoform X4
693	2	XP_690190.5	Membrane-associated guanylate kinase, WW and PDZ domain-containing protein 3
694	2	XP_003197854.1	Meprin A subunit beta
695	2	XP_003454224.1	Merlin-like
696	2	NP_001082308.1	Mesendoderm nuclear factor, gene 1
697	2	NP_955853.1	Mesoderm induction early response 1 homolog a
698	2	XP_004555069.1	Metalloprotease TIKI2-like
699	2	NP_001121710.1	Metaxin 1a
700	2	NP_001188160.1	Methionine adenosyltransferase 2 subunit beta
701	2	XP_004590682.1	Methionine aminopeptidase 1
702	2	NP_956370.1	Methionine--trna ligase, cytoplasmic
703	2	AAH75779.1	Methylenetetrahydrofolate dehydrogenase (NADP+ dependent)
704	2	ACI70132.1	Methyltransferase Mb3374
705	2	AAG29241.1	MHC class I antigen
706	2	XP_004267249.1	Microfibril-associated glycoprotein 4
707	2	AAI35021.1	Microsomal triglyceride transfer protein
708	2	NP_001036187.1	Microtubule-associated protein 4
709	2	XP_001337128.2	Microtubule-associated serine/threonine-protein kinase 1

710	2	XP_003200751.1	Midasin
711	2	XP_003966501.1	Midline-1-like
712	2	XP_003353524.2	Mis18-binding protein 1-like
713	2	NP_001018198.1	Mitochondria-eating protein
714	2	NP_001188000.1	Mitochondrial 28S ribosomal protein s14
715	2	ABD85505.1	Mitochondrial 28S ribosomal protein S32-like
716	2	NP_001188133.1	Mitochondrial ATP synthase f1 complex assembly factor 1
717	2	ADO28404.1	Mitochondrial cytochrome c1 heme protein
718	2	NP_001188156.1	Mitochondrial glycine cleavage system h protein
719	2	NP_001187893.1	Mitochondrial hydroxyacyl-coenzyme a dehydrogenase
720	2	NP_001107910.1	Mitochondrial import receptor subunit TOM40B
721	2	XP_003452865.1	Mitochondrial Rho gtpase 1-A-like
722	2	XP_004551901.1	Mitogen-activated protein kinase kinase kinase 15-like isoform X4
723	2	XP_001343426.4	Mitogen-activated protein kinase kinase kinase kinase 4
724	2	NP_957206.2	Mitotic spindle assembly checkpoint protein MAD1
725	2	XP_004078796.1	MLN64 N-terminal domain homolog
726	2	BAL44665.1	Monocarboxylate transporter 1a
727	2	AEE36662.1	Mucolipin 1.1
728	2	NP_957442.1	Mucolipin-2
729	2	XP_002661250.2	Multidrug resistance-associated protein 1-like
730	2	XP_004545859.1	Multidrug resistance-associated protein 5-like isoform X1
731	2	XP_003969469.1	Multiple C2 and transmembrane domain-containing protein 2-like
732	2	XP_004552705.1	Muscleblind-like protein 2a-like isoform X11
733	2	NP_001187679.1	Muscle-related coiled-coil protein
734	2	NP_001008579.1	Muscle-related coiled-coil protein b
735	2	NP_001071077.1	Muscle-restricted dual specificity phosphatase
736	2	NP_001038364.1	Myelin transcription factor 1-like protein
737	2	XP_004541632.1	Myeloid cell surface antigen CD33-like isoform X1
738	2	XP_687465.4	Myoferlin
739	2	NP_001154812.1	Myomesin 1a
740	2	BAF93222.1	Myosin heavy chain fast skeletal type 1
741	2	NP_001188138.1	Myosin regulatory light chain 2 skeletal muscle isoform
742	2	NP_001108561.1	Myosin, heavy polypeptide 1.1, skeletal muscle
743	2	AAI62152.1	Myosin, heavy polypeptide 11, smooth muscle
744	2	XP_683046.5	Myosin-10
745	2	XP_001921522.1	Myosin-viia-like
746	2	XP_001919593.3	Myosin-XV
747	2	XP_001920958.2	Myosin-XVI
748	2	XP_003974778.1	Myotubularin-related protein 3-like
749	2	NP_956085.1	Myotubularin-related protein 6
750	2	AGI42757.1	Myoxvirus resistance protein B
751	2	AAG27060.1	Na+/K+ atpase alpha subunit isoform 8
752	2	NP_001091660.1	N-acetylated-alpha-linked acidic dipeptidase-like protein

753	2	XP_003454735.1	NACHT, LRR and PYD domains-containing protein 1-like
754	2	XP_003457442.1	NACHT, LRR and PYD domains-containing protein 5-like
755	2	ADO28298.1	N-acylneuraminate cytidylyltransferase
756	2	XP_003438026.1	NAD-dependent deacetylase sirtuin-6-like
757	2	NP_955890.1	NAD-dependent protein deacetylase sirtuin-2
758	2	NP_001003747.1	NADH dehydrogenase [ubiquinone] flavoprotein 1, mitochondrial
759	2	XP_003448033.1	N-alpha-acetyltransferase 15, nata auxiliary subunit-like
760	2	NP_955844.1	N-alpha-acetyltransferase 35, natc auxiliary subunit
761	2	CBN80853.1	Nck-associated protein 1
762	2	NP_001189370.1	Nephrocystin-3
763	2	XP_698040.5	Nesprin-1
764	2	XP_001920330.1	Nesprin-1-like
765	2	XP_004555349.1	Netrin-G1-like isoform X1
766	2	XP_003973781.1	Neural cell adhesion molecule L1-like protein-like
767	2	XP_003967258.1	Neural Wiskott-Aldrich syndrome protein-like
768	2	NP_001073490.1	Neurexin 1a precursor
769	2	ELW48367.1	Neurobeachin
770	2	EMP24328.1	Neurobeachin-like protein 1, partial
771	2	XP_687696.3	Neuroblast differentiation-associated protein AHNAK
772	2	XP_001335151.4	Neuroblast differentiation-associated protein AHNAK-like isoform X1
773	2	NP_001038272.1	Neuroblastoma-amplified sequence
774	2	NP_001131134.1	Neuroendocrine convertase 1 precursor
775	2	XP_001921978.2	Neurofascin-like
776	2	XP_002666871.2	Neuron navigator 2
777	2	NP_001017885.1	Neuronal acetylcholine receptor subunit alpha-5 precursor
778	2	NP_001036149.1	Neuronal acetylcholine receptor subunit alpha-6 precursor
779	2	XP_004572727.1	Neuronal acetylcholine receptor subunit alpha-7-like
780	2	XP_003444756.1	Neuronal acetylcholine receptor subunit beta-2-like
781	2	NP_001038562.1	Neuronal nicotinic acetylocholine receptor alpha-7 subunit-like precursor
782	2	XP_004545613.1	Neutral alpha-glucosidase AB-like isoform X1
783	2	ADO27934.1	Nicotinamide mononucleotide adenylyltransferase 2
784	2	ACO10003.1	Nicotinamide riboside kinase 2
785	2	NP_001230804.1	Niemann-Pick C1 protein precursor
786	2	XP_003445249.1	NKAP-like protein-like
787	2	ABY26544.1	NMDA receptor NR2C subunit
788	2	XP_001344451.2	NMDA receptor-regulated protein 2
789	2	NP_001133241.1	Nogo-B receptor
790	2	NP_001129167.1	Notch 1 precursor
791	2	NP_001108566.1	Notch homolog 2
792	2	NP_001187235.1	Novel immune-type receptor 16 precursor
793	2	CAK04114.1	Novel protein (zgc:56258)
794	2	CAQ15073.1	Novel protein

795	2	CAI21354.1	Novel protein
796	2	CAI11836.1	Novel protein containing multiple scavenger receptor cysteine-rich (SRCR) domains similar to vertebrate deleted in malignant brain tumors 1 (DMBT1)
797	2	CAK10842.1	Novel protein similar to vertebrate adenosine deaminase, RNA-specific, B2 (ADARB2)
798	2	CAQ14846.1	Novel protein similar to vertebrate nuclear transcription factor Y, alpha (NFYA)
799	2	CAE30438.1	Novel protein similar to zebrafish epithelial cadherin 1 (cdh1)
800	2	XP_004564752.1	N-terminal EF-hand calcium-binding protein 2-like
801	2	XP_690435.2	Ntpase KAP family P-loop domain-containing protein 1-like
802	2	EGW00530.1	Nuclear factor of activated T-cells, cytoplasmic 2
803	2	XP_004560368.1	Nuclear factor of activated T-cells, cytoplasmic 2-like isoform X1
804	2	XP_690273.3	Nuclear factor of activated T-cells, cytoplasmic 3
805	2	NP_001025338.1	Nuclear pore complex protein Nup107
806	2	XP_002667606.2	Nuclear pore membrane glycoprotein 210
807	2	XP_004557497.1	Nuclear transcriptional regulator 1-like protein-like isoform X1
808	2	NP_001038928.1	Nucleobindin-1 precursor
809	2	NP_956531.1	Nucleolar protein 10
810	2	XP_844668.1	Nucleoporin (NUP54/57)
811	2	NP_001018431.1	Nucleoredoxin
812	2	XP_004900906.1	Nucleosome-remodeling factor subunit BPTF
813	2	XP_001341205.4	Obscurin
814	2	XP_003206940.1	Obscurin-like
815	2	NP_001124269.1	Odorant receptor, family H, subfamily 132, member 1
816	2	NP_001119944.1	Olfactomedin-like 2A precursor
817	2	XP_004070619.1	Olfactory receptor 13C8-like
818	2	XP_001919585.2	Olfactory receptor 2AJ1
819	2	AEB77798.1	Olfactory receptor family C subfamily 11 member 3
820	2	CAJ90908.1	Opioid growth factor receptor
821	2	NP_001159919.1	ORAI calcium release-activated calcium modulator 2
822	2	BAN16579.1	Organic solute transporter alpha-1
823	2	NP_001017568.2	Ornithine decarboxylase antizyme inhibitor
824	2	AAI62978.1	Osbpl1a protein
825	2	NP_001025283.1	Otoferlin
826	2	XP_687615.2	Oxoglutarate dehydrogenase-like
827	2	XP_004557834.1	Oxysterol-binding protein-related protein 6-like isoform X1
828	2	NP_001152810.1	Oxysterols receptor LXR-alpha
829	2	XP_003454867.1	P2Y purinoceptor 1-like
830	2	XP_004451590.1	Paired mesoderm homeobox protein 1 isoform 1
831	2	NP_001071249.1	Palmitoyltransferase ZDHHC15
832	2	NP_001133464.1	Para-nitrobenzyl esterase
833	2	XP_692276.4	Patatin-like phospholipase domain containing 7
834	2	XP_003198698.1	PDZ domain-containing protein 2

835	2	NP_001177686.1	PDZ domain-containing protein 7
836	2	XP_001340677.2	Pejvakin-like
837	2	NP_001187944.1	Pepsin A precursor
838	2	AAI63549.1	Per2 protein
839	2	ABU96277.1	Perforin-like protein 1
840	2	NP_571659.1	Period circadian protein homolog 3
841	2	NP_001071254.1	Periostin isoform 1 precursor
842	2	ACI33730.1	Peripheral myelin protein 22
843	2	NP_001154805.1	Peroxisome proliferator-activated receptor alpha a
844	2	XP_685305.1	PH and SEC7 domain-containing protein 2-like
845	2	XP_001922171.1	PH and SEC7 domain-containing protein 3, partial
846	2	XP_004568614.1	PHD finger protein 1-like
847	2	XP_004562971.1	Phenylalanine--trna ligase alpha subunit-like
848	2	ACN10549.1	Phosphatase methylesterase 1
849	2	XP_694550.4	Phosphatidylinositol 4,5-bisphosphate 5-phosphatase A-like
850	2	XP_004546609.1	Phosphatidylinositol 5-phosphate 4-kinase type-2 alpha-like isoform X3
851	2	XP_003446864.1	Phosphatidylinositol transfer protein alpha isoform-like
852	2	XP_001923007.2	Phosphatidylinositol-3,4,5-trisphosphate 5-phosphatase 1-like
853	2	NP_001092218.1	Phosphofurin acidic cluster sorting protein 1
854	2	XP_004539992.1	Phosphofurin acidic cluster sorting protein 2-like isoform X3
855	2	XP_003961394.1	Phosphoglycolate phosphatase-like
856	2	XP_003451275.1	Phosphoinositide 3-kinase regulatory subunit 4
857	2	NP_001265769.1	Phospholipase C-beta 2
858	2	NP_001003519.1	Phospholipid transfer protein precursor
859	2	AAI42821.1	Phosphoribosyl pyrophosphate amidotransferase
860	2	Q9W6R1.1	Phosphorylase b kinase regulatory subunit alpha, liver isoform; Short=Phosphorylase kinase alpha L subunit
861	2	AAI54375.1	Pibf1 protein
862	2	NP_001013283.2	Pin2-interacting protein X1
863	2	NP_001070987.1	Pleckstrin homology domain-containing family H member 1
864	2	XP_004571023.1	Pleckstrin homology domain-containing family M member 2-like isoform X1
865	2	XP_003200176.1	Plectin-like
866	2	XP_003365958.1	Plexin-A3
867	2	NP_001108558.1	Poly (ADP-ribose) polymerase family, member 12b
868	2	XP_001338257.2	Poly(ADP-ribose) glycohydrolase
869	2	XP_004555306.1	Poly(U)-specific endoribonuclease-C-like
870	2	XP_004370370.1	Polyadenylate-binding protein 1-like isoform 2
871	2	NP_001038700.1	Polycomb group RING finger protein 5-B
872	2	NP_001076293.2	Polycomb protein suz12-B
873	2	XP_003197787.1	Polycystin-1-like
874	2	NP_001187310.1	Post-GPI attachment to proteins factor 2
875	2	XP_003450492.1	Potassium channel subfamily K member 4-like

876	2	XP_003445454.1	Potassium channel subfamily K member 5-like
877	2	XP_003443804.1	Potassium channel subfamily K member 9-like
878	2	ABX89981.1	Potassium inwardly-rectifying channel Kir2.5
879	2	XP_003451929.1	Potassium voltage-gated channel subfamily S member 1-like
880	2	NP_001072370.1	Potassium voltage-gated channel, shaker-related subfamily, beta member 3
881	2	AAI63642.1	Potassium voltage-gated channel, subfamily H (eag-related), member 2
882	2	XP_004388170.1	Potassium-transporting atpase alpha chain 1
883	2	NP_001038720.2	POU transcription factor isoform 2
884	2	ACN24618.1	PR domain containing 9
885	2	XP_001623045.1	Predicted protein
886	2	XP_001636323.1	Predicted protein
887	2	XP_001633866.1	Predicted protein
888	2	XP_001641379.1	Predicted protein
889	2	XP_001637582.1	Predicted protein
890	2	NP_001265737.1	Pregnancy-associated plasma protein A, pappalysin 1b precursor
891	2	NP_958875.1	Pre-mrna-processing factor 19
892	2	AAK28709.1	Preproinsulin
893	2	NP_956176.1	Probable ATP-dependent RNA helicase DDX23
894	2	XP_003200872.1	Probable E3 ubiquitin-protein ligase HERC1-like
895	2	NP_001038240.1	Probable global transcription activator SNF2L2 isoform 2
896	2	XP_003201802.1	Probable G-protein coupled receptor 116-like, partial
897	2	NP_998431.1	Probable helicase with zinc finger domain
898	2	NP_001159825.2	Probable imidazolonepropionase
899	2	XP_700282.5	Probable jmjc domain-containing histone demethylation protein 2C-like
900	2	XP_004539905.1	Probable palmitoyltransferase ZDHHC1-like isoform X1
901	2	ADO28349.1	Probable palmitoyltransferase zdhhc4
902	2	XP_004541355.1	Probable phospholipid-transporting atpase ID-like isoform X3
903	2	NP_001108144.1	Probable phospholipid-transporting atpase IIA
904	2	XP_003764299.1	Probable ribonuclease ZC3H12C
905	2	XP_004083507.1	Prolyl endopeptidase-like
906	2	XP_003443395.1	Prominin-1-A-like
907	2	ABN03942.1	Prospero-related homeobox protein
908	2	AAD45896.1	Prostaglandin endoperoxide synthase-2
909	2	NP_001155958.1	Proteasome-associated protein ECM29 homolog
910	2	XP_003445828.1	Protein ADP-ribosylarginine hydrolase-like
911	2	XP_699384.4	Protein argonaute-1-like
912	2	NP_001092900.1	Protein asteroid homolog 1
913	2	XP_002664470.1	Protein ATP1B4
914	2	XP_002663838.2	Protein bassoon-like
915	2	XP_004572732.1	Protein capicua homolog isoform X1
916	2	XP_003456156.1	Protein CBFA2T2

917	2	XP_003968317.1	Protein C-ets-1-like
918	2	XP_003198156.1	Protein CLEC16A isoform X1
919	2	XP_001921712.1	Protein delta homolog 2
920	2	XP_004554963.1	Protein Dos-like isoform X1
921	2	XP_004556175.1	Protein eyes shut homolog
922	2	NP_001082796.1	Protein FAM115
923	2	XP_004549615.1	Protein FAM13A-like
924	2	XP_003970687.1	Protein FAM13B-like
925	2	XP_004559756.1	Protein FAM160B1-like
926	2	XP_004835674.1	Protein FAM179B isoform X3
927	2	XP_696880.3	Protein FAM83B isoform X2
928	2	XP_003428305.1	Protein flightless-1 homolog
929	2	XP_004563007.1	Protein Hook homolog 2-like isoform X1
930	2	XP_003504247.1	Protein IWS1 homolog
931	2	XP_687326.2	Protein KIAA1045-like isoform X2
932	2	NP_957283.1	Protein kinase C and casein kinase substrate in neurons 3
933	2	NP_955935.1	Protein kinase C binding protein 1, like
934	2	XP_002664164.1	Protein kinase C epsilon type
935	2	NP_001082929.1	Protein kinase N1
936	2	NP_001019984.1	Protein MCM10 homolog
937	2	NP_956672.1	Protein MIS12 homolog
938	2	XP_004562681.1	Protein NLRC5-like
939	2	XP_003443005.1	Protein numb homolog
940	2	XP_686501.5	Protein osteopotentia
941	2	NP_957128.1	Protein phosphatase 1 regulatory subunit 3C-B
942	2	NP_001094422.1	Protein phosphatase 1, regulatory (inhibitor) subunit 8
943	2	XP_684981.2	Protein phosphatase 1H-like
944	2	NP_919396.1	Protein phosphatase 2, regulatory subunit B', epsilon isoform a
945	2	XP_690522.1	Protein QN1 homolog
946	2	XP_003200596.1	Protein shisa-7 isoform X1
947	2	XP_686686.2	Protein sidekick-1
948	2	XP_693332.4	Protein sidekick-2
949	2	NP_705949.1	Protein spinster homolog 1
950	2	XP_003453562.1	Protein spire homolog 1-like
951	2	XP_004566168.1	Protein spire homolog 2-like
952	2	NP_001186845.1	Protein strawberry notch homolog 1
953	2	XP_001920231.1	Protein TANC1-like
954	2	XP_004556789.1	Protein TANC2-like isoform X10
955	2	XP_001921634.3	Protein transport protein Sec24B isoform X1
956	2	XP_695776.3	Protein turtle homolog B
957	2	XP_004873998.1	Protein unc-13 homolog B isoform X3
958	2	XP_004554612.1	Protein very KIND-like
959	2	XP_688748.3	Protein virilizer homolog

960	2	XP_002691265.1	Protein YIPF4
961	2	XP_694950.3	Protein-glutamine gamma-glutamyltransferase K isoform 2
962	2	NP_997735.1	Protein-tyrosine kinase 2-beta
963	2	NP_998555.1	Prothrombin precursor
964	2	NP_001009589.1	Protocadherin 2 alpha b 10 precursor
965	2	XP_003444714.1	Protocadherin beta-15-like
966	2	XP_003199405.1	Protocadherin Fat 3-like, partial
967	2	XP_002664697.1	Protocadherin-16-like
968	2	NP_956579.1	Proton-coupled folate transporter
969	2	XP_001345065.3	Proto-oncogene serine/threonine-protein kinase pim-1-like
970	2	NP_001123984.1	Putative ATP-dependent RNA helicase TDRD9
971	2	AAL35358.1	Putative GAG protein
972	2	XP_002662576.2	Putative helicase mov-10-B.1
973	2	EFN65029.1	Putative leucine-rich repeat-containing protein DDB0188916
974	2	XP_004082036.1	Putative methyltransferase DDB_G0268948-like
975	2	NP_001187113.1	Putative odorant receptor CF64 precursor
976	2	XP_687318.4	Putative pre-mrna-splicing factor ATP-dependent RNA helicase DHX32 isoform X2
977	2	XP_003201406.1	Putative protein MSS51 homolog, mitochondrial-like isoform X1
978	2	CAA73349.1	Putative RNA helicase (DEAD box)
979	2	XP_003962005.1	Putative sodium-coupled neutral amino acid transporter 11-like
980	2	XP_003974457.1	Putative tyrosine-protein phosphatase auxilin-like
981	2	EPQ06186.1	R3H domain-containing protein 1
982	2	XP_004574271.1	Rab-like protein 6-like isoform X1
983	2	NP_001184130.1	RAC-gamma serine/threonine-protein kinase
984	2	XP_001923770.1	Radial spoke head protein 4 homolog A isoform 1
985	2	XP_002662611.2	Ral gtpase-activating protein subunit beta
986	2	XP_687578.3	Rap guanine nucleotide exchange factor 4 isoform 2
987	2	XP_004017077.1	Rapamycin-insensitive companion of mtor isoform 1
988	2	NP_001264023.1	RAR-related orphan receptor C b
989	2	XP_003441778.1	Ras gtpase-activating protein 2
990	2	XP_001332688.4	Ras gtpase-activating protein ngap
991	2	XP_003200549.1	Ras gtpase-activating protein syngap
992	2	XP_005010595.1	Ras gtpase-activating protein-binding protein 1 isoform X3
993	2	NP_001121812.1	Ras gtpase-activating-like protein IQGAP1
994	2	EGV98126.1	Ras-related protein Rab-13
995	2	XP_003448512.1	Ras-related protein Rab-31-like
996	2	NP_001017761.1	Ras-related protein Rab-3A
997	2	XP_003973253.1	Receptor tyrosine-protein kinase erbb-3-like
998	2	NP_001186992.1	Receptor-type tyrosine-protein phosphatase C precursor
999	2	XP_004562082.1	Receptor-type tyrosine-protein phosphatase delta-like isoform X11
1000	2	XP_003214862.1	Receptor-type tyrosine-protein phosphatase eta-like
1001	2	ELK32846.1	Regulating synaptic membrane exocytosis protein 2

1002	2	XP_003452345.1	Regulating synaptic membrane exocytosis protein 2-like
1003	2	XP_002662224.2	Regulator of G-protein signaling 3-like
1004	2	XP_001919542.3	Relaxin receptor 1
1005	2	XP_004071449.1	Retinal rod rhodopsin-sensitive cgmp 3',5'-cyclic phosphodiesterase subunit gamma-like
1006	2	AFV31625.1	Retinaldehyde dehydrogenase 2, partial
1007	2	XP_003455476.1	Retinaldehyde-binding protein 1-like
1008	2	XP_683123.5	Retinal-specific ATP-binding cassette transporter isoform X2
1009	2	NP_001265735.1	Retinoic acid-inducible protein I
1010	2	BAG72429.1	Retinol dehydrogenase 8
1011	2	XP_684476.1	Retinol dehydrogenase 8-like isoform X2
1012	2	XP_002741079.1	Retrovirus polyprotein, putative-like
1013	2	ADO28323.1	Rhamnose-binding lectin
1014	2	NP_001038715.1	Rho gtpase-activating protein 10
1015	2	NP_001180468.1	Rho gtpase-activating protein 11A isoform 1
1016	2	XP_004572338.1	Rho gtpase-activating protein 12-like isoform X2
1017	2	NP_001122027.1	Rho gtpase-activating protein 15
1018	2	NP_001133974.1	Rho gtpase-activating protein 18
1019	2	XP_004573082.1	Rho gtpase-activating protein 20-like
1020	2	NP_956405.1	Rho gtpase-activating protein 29
1021	2	CBN80691.1	Rho guanine nucleotide exchange factor 10-like protein
1022	2	XP_003966965.1	Rho guanine nucleotide exchange factor 15-like
1023	2	XP_689770.5	Rho guanine nucleotide exchange factor 2 isoform X2
1024	2	NP_957472.1	Rhophilin-2
1025	2	NP_001188131.1	Rho-related gtp-binding protein rhoe
1026	2	NP_001038829.1	Rho-related GTP-binding protein rhon
1027	2	NP_001187596.1	Ribonuclease h2 subunit b
1028	2	NP_001088030.1	Ribosomal protein L3-like
1029	2	XP_004082553.1	Ribosomal protein S6 kinase 2 alpha-like
1030	2	NP_001104620.2	Ribosome biogenesis protein BMS1 homolog
1031	2	ADO28027.1	Ring finger and chy zinc finger domain-containing protein 1
1032	2	NP_001116727.1	RING finger protein 10
1033	2	NP_001004600.1	RING finger protein 103
1034	2	XP_001342512.2	RING finger protein 145-like isoform X1
1035	2	XP_003200094.1	RING finger protein 26
1036	2	XP_696033.4	RING finger protein 31
1037	2	ADO27902.1	Ring finger protein 32
1038	2	XP_003450737.1	RNA binding protein fox-1 homolog 1-like
1039	2	NP_001119888.1	RNA exonuclease 1 homolog
1040	2	NP_001002873.1	RNA polymerase II subunit A C-terminal domain phosphatase
1041	2	XP_005026930.1	RNA-binding motif, single-stranded-interacting protein 1 isoform X2
1042	2	Q6IQ97.1	RNA-binding protein 4.1; altname: Full=RNA-binding motif protein 4.1
1043	2	NP_001120874.1	RNA-binding protein 45

1044	2	NP_998657.1	RNA-binding protein MEX3A
1045	2	XP_683754.2	RNA-binding protein NOB1 isoformx1
1046	2	EHJ71677.1	RNA-directed DNA polymerase
1047	2	CAM60061.1	Robo3
1048	2	CAM60093.1	Robo4
1049	2	XP_003455977.1	Roundabout homolog 1-like
1050	2	NP_777285.1	Ruvb-like 2
1051	2	XP_004570447.1	Ryanodine receptor 2-like
1052	2	NP_001116519.1	S phase cyclin A-associated protein in the endoplasmic reticulum
1053	2	XP_004067487.1	Sal-like protein 1-like
1054	2	XP_004687010.1	SAM domain and HD domain-containing protein 1
1055	2	NP_957259.1	Sarcoplasmic/endoplasmic reticulum calcium atpase 2
1056	2	NP_957422.1	Sarcosine dehydrogenase, mitochondrial
1057	2	NP_001005304.1	Sciellin
1058	2	NP_775351.1	SCL-interrupting locus protein homolog
1059	2	XP_004080924.1	SCO-spondin-like
1060	2	XP_004538842.1	SEC14-like protein 1-like isoform X1
1061	2	NP_001070847.1	SEC23-interacting protein
1062	2	NP_991110.1	Secretory carrier-associated membrane protein 5
1063	2	XP_004568655.1	Secretory phospholipase A2 receptor-like
1064	2	AAI63771.1	Semaphorin 3fa
1065	2	XP_003452457.1	Semaphorin-4F-like
1066	2	XP_001343014.4	Septin-4
1067	2	NP_001133813.1	Sequestosome-1
1068	2	NP_998473.1	Serine/threonine-protein kinase 25
1069	2	XP_004575564.1	Serine/threonine-protein kinase PAK 1-like isoform X1
1070	2	NP_001092715.1	Serine/threonine-protein kinase PLK2
1071	2	XP_004079359.1	Serine/threonine-protein kinase STK11-like
1072	2	XP_004574667.1	Serine/threonine-protein kinase TAO2-like isoform X1
1073	2	XP_003198658.1	Serine/threonine-protein kinase TAO3-like isoform 3
1074	2	NP_998483.1	Serine/threonine-protein phosphatase 2A 56 kda regulatory subunit delta isoform
1075	2	NP_956425.1	Serine/threonine-protein phosphatase 2A regulatory subunit B" subunit gamma
1076	2	XP_003459500.1	Serine/threonine-protein phosphatase 5 isoform 1
1077	2	XP_004072143.1	Serine/threonine-protein phosphatase with EF-hands 2-like
1078	2	XP_002664220.2	SH2 domain-containing protein 4B
1079	2	NP_001103925.1	SH2B adapter protein 1
1080	2	XP_004080428.1	SH3 and cysteine-rich domain-containing protein 2-like
1081	2	XP_003198290.1	SH3 and multiple ankyrin repeat domains protein 3 isoform X1
1082	2	XP_001345357.3	SH3 domain-binding protein 1-like
1083	2	XP_003974212.1	SHC-transforming protein 2-like
1084	2	XP_004563924.1	Short transient receptor potential channel 2-like isoform X1
1085	2	XP_695955.2	Short transient receptor potential channel 7-like

1086	2	NP_001003743.1	Short-chain specific acyl-coa dehydrogenase, mitochondrial
1087	2	AAI34974.1	Si:ch211-105d11.3
1088	2	XP_687974.4	Si:ch211-1n9.7
1089	2	AAI55117.1	Si:ch211-245p7.3 protein
1090	2	XP_696251.4	Si:ch211-267e7.3
1091	2	XP_683027.4	Si:dkey-170m15.2
1092	2	AAH44435.1	Si:dkey-18f23.10 protein
1093	2	NP_001108524.1	Si:dkey-281i8.1
1094	2	XP_001919090.3	Sialoadhesin-like
1095	2	XP_004541757.1	Sialoadhesin-like
1096	2	XP_004546862.1	Sickle tail protein homolog isoform X8
1097	2	NP_001243552.1	Sin3A-associated protein b
1098	2	XP_004934598.1	Single-minded homolog 2 isoform X1
1099	2	XP_004069298.1	Single-strand selective monofunctional uracil DNA glycosylase-like
1100	2	XP_002666900.2	SITS-binding protein-like
1101	2	AAX30323.2	SJCHGC03012 protein
1102	2	AAX26836.2	SJCHGC03019 protein
1103	2	AAX30342.1	SJCHGC03036 protein
1104	2	AAX27441.2	SJCHGC08991 protein
1105	2	AAX25882.2	SJCHGC09227 protein
1106	2	AAX25877.2	SJCHGC09533 protein
1107	2	XP_003457715.1	SLAM family member 7-like
1108	2	ACI70099.1	SLAM family member 8 precursor
1109	2	XP_004077200.1	Slit homolog 1 protein-like isoform 2
1110	2	Q9WVC1.3	Slit homolog 2 protein; Short=Slit-2; Flags: Precursor
1111	2	NP_001187097.1	Small inducible cytokine A3 precursor
1112	2	NP_991230.1	Small nuclear ribonucleoprotein polypeptides B and B1
1113	2	XP_001920500.3	Small subunit processome component 20 homolog, partial
1114	2	NP_001180256.1	SNF-related serine/threonine-protein kinase
1115	2	XP_003969374.1	Sodium bicarbonate cotransporter 3-like
1116	2	XP_004084549.1	Sodium channel protein type 4 subunit alpha B-like
1117	2	NP_571703.1	Sodium channel, voltage-gated, type VIII, alpha a
1118	2	NP_001017549.2	Sodium leak channel non-selective protein precursor
1119	2	XP_003449878.1	Sodium/calcium exchanger 3-like
1120	2	XP_001345853.1	Sodium/glucose cotransporter 5-like
1121	2	XP_003970227.1	Sodium/hydrogen exchanger 6-like
1122	2	NP_956662.1	Sodium-coupled monocarboxylate transporter 2
1123	2	NP_571830.1	Sodium-dependent dopamine transporter
1124	2	XP_004565185.1	Sodium-dependent dopamine transporter-like
1125	2	NP_001108369.1	Sodium-dependent neutral amino acid transporter B(0)AT3
1126	2	XP_003970419.1	Soluble guanylate cyclase 88E-like
1127	2	NP_001157126.1	Solute carrier family 12 member 2 isoform 2
1128	2	XP_701000.4	Solute carrier family 12 member 5-like

1129	2	XP_003457356.1	Solute carrier family 22 member 23-like
1130	2	XP_002663366.1	Solute carrier family 23 member 3-like
1131	2	XP_004542481.1	Solute carrier family 25 member 36-A-like
1132	2	XP_003200421.1	Solute carrier family 35 member F4-like
1133	2	XP_694603.3	Solute carrier family 41 member 2-like isoform X2
1134	2	B0S5Y3.1	Solute carrier family 52, riboflavin transporter, member 3-A; altname: Full=Riboflavin transporter 2-A; Short=RFT2-A
1135	2	NP_001038752.2	Solute carrier family 6 (neurotransmitter transporter, GABA), member 1a
1136	2	NP_001085235.1	Solute carrier family 7 (amino acid transporter light chain, y+L system), member 7
1137	2	XP_004075747.1	Solute carrier organic anion transporter family member 1C1-like
1138	2	NP_998395.1	Sortilin precursor
1139	2	AAI29211.1	Sorting nexin 14
1140	2	NP_001014368.1	Sorting nexin-4
1141	2	NP_001025388.1	Sorting nexin-6
1142	2	XP_004552292.1	Sorting nexin-8-like
1143	2	BAE78525.1	Sox-lz
1144	2	AAT01213.1	Sparc
1145	2	XP_004560650.1	Spartin-like isoform X1
1146	2	NP_571600.1	Spectrin beta chain, erythrocyte
1147	2	NP_001017814.1	Spermatogenesis-associated protein 6
1148	2	NP_001082938.1	Sphingosine-1-phosphate lyase 1
1149	2	XP_003455272.1	Splicing factor 3A subunit 3-like
1150	2	XP_002667683.2	Splicing factor 3B subunit 3-like
1151	2	NP_957209.1	Splicing factor 45
1152	2	NP_571083.4	Spondin-2 precursor
1153	2	XP_686737.2	Sprouty-related, EVH1 domain-containing protein 2-like
1154	2	AAI63594.1	Squamous cell carcinoma antigen recognised by T cells 3
1155	2	XP_003442218.1	SRC kinase signaling inhibitor 1-like
1156	2	AAH90279.1	Srl protein
1157	2	XP_001922696.1	Stonustoxin subunit beta
1158	2	XP_001919441.3	Storkhead-box protein 2-like
1159	2	XP_004575317.1	Stress-70 protein, mitochondrial-like isoform X1
1160	2	NP_001180470.1	Structural maintenance of chromosomes protein 5
1161	2	XP_004075668.1	Succinate receptor 1-like
1162	2	EPQ04722.1	Succinyl-coa ligase [ADP-forming] subunit beta, mitochondrial
1163	2	XP_003974238.1	Sulfhydryl oxidase 1-like
1164	2	NP_001017881.1	Sulfide:quinone oxidoreductase, mitochondrial
1165	2	NP_958871.1	Suppressor of SWI4 1 homolog
1166	2	XP_003966129.1	Suppressor of tumorigenicity 14 protein-like
1167	2	XP_003772918.1	SWI/SNF-related matrix-associated actin-dependent regulator of chromatin subfamily A containing DEAD/H box 1 isoform 2
1168	2	NP_001038451.1	Switch-associated protein 70
1169	2	XP_002916209.1	Switch-associated protein 70-like

1170	2	XP_003964937.1	Synaptonemal complex protein SC65-like
1171	2	XP_686583.4	Synaptoporin-like isoform X2
1172	2	XP_691385.2	Synaptotagmin-6-like
1173	2	XP_002664653.2	Synaptotagmin-like protein 2-like
1174	2	XP_004856263.1	Syntaxin-1B isoform X4
1175	2	XP_004538515.1	Syntaxin-binding protein 1-like isoform X1
1176	2	XP_004068096.1	Syntaxin-binding protein 3-like isoform 1
1177	2	NP_001187790.1	Syntenin-1
1178	2	XP_002939302.1	Tankyrase-1 isoform X1
1179	2	XP_003223838.1	Tankyrase-2-like
1180	2	YP_006591514.1	Tash protein pest motif family
1181	2	XP_003201409.1	TBC1 domain family member 24-like
1182	2	NP_001120987.1	TBC1 domain family member 8B
1183	2	XP_003962264.1	TBC1 domain family member 8-like
1184	2	Q7ZTU9.4	T-box transcription factor TBX2b; Short=T-box protein 2b
1185	2	NP_001004548.1	T-cell activation Rho gtpase-activating protein
1186	2	NP_739571.1	T-cell leukemia homeobox protein 1
1187	2	NP_001133482.1	T-complex protein 1 subunit delta
1188	2	NP_775355.1	T-complex protein 1 subunit eta
1189	2	XP_002667399.2	T-complex-associated testis-expressed protein 1-like
1190	2	XP_004555260.1	Telomerase protein component 1-like isoform X1
1191	2	XP_004602630.1	Telomeric repeat-binding factor 1
1192	2	CAM14027.1	Tenascin R (restrictin, janusin)
1193	2	NP_001107885.1	Tenomodulin
1194	2	XP_003199292.1	Tensin-1
1195	2	XP_001379887.2	Tensin-3
1196	2	XP_003454076.1	Terminal uridylyltransferase 4
1197	2	XP_003448265.1	Tetraspan membrane protein of hair cell stereocilia homolog
1198	2	XP_004569425.1	Tetraspanin-1-like
1199	2	NP_001187654.1	Tetraspanin-9
1200	2	XP_698025.4	Tetratricopeptide repeat protein 17
1201	2	XP_002667710.2	Tetratricopeptide repeat protein 28
1202	2	XP_001921229.3	Tetratricopeptide repeat protein 7A
1203	2	AAH65630.1	Tfg protein
1204	2	NP_963876.1	TFIIH basal transcription factor complex helicase XPB subunit
1205	2	AAH66615.1	Tgds protein
1206	2	NP_001005930.1	Thimet oligopeptidase 1
1207	2	ADO28342.1	Thioredoxin-like protein 1
1208	2	XP_003978153.1	THO complex subunit 3-like
1209	2	NP_001116258.1	Threonyl-trna synthetase, cytoplasmic isoform 1
1210	2	XP_004940194.1	Thrombomodulin
1211	2	NP_001116232.1	Thrombospondin type-1 domain-containing protein 7A precursor
1212	2	NP_775332.1	Thrombospondin-3a precursor

1213	2	XP_004561072.1	Thyroid hormone receptor beta-like isoform X1
1214	2	XP_690889.2	Thyrotropin-releasing hormone-degrading ectoenzyme
1215	2	NP_001188499.1	Tight junction protein ZO-2 isoform 1
1216	2	XP_003974133.1	Tight junction protein ZO-3-like
1217	2	NP_001187154.1	TIR-containing adaptor molecule
1218	2	AEL30371.1	TIR-NBS-LRR type disease resistance protein
1219	2	ABG48500.1	Titin a
1220	2	AEI59671.1	Tlr8-1
1221	2	XP_002665298.1	T-lymphoma invasion and metastasis-inducing protein 2-like
1222	2	XP_003972119.1	Tmc6-related protein 1
1223	2	AAZ66788.1	TNF receptor
1224	2	XP_692341.5	TNF receptor-associated factor 5, partial
1225	2	AAW22592.1	TNF-related apoptosis inducing ligand TRAIL
1226	2	Q7ZV43.1	Toll-interacting protein
1227	2	NP_001186994.1	Toll-like receptor 21
1228	2	XP_001339602.3	Torsin-1A-interacting protein 2
1229	2	DAA31163.1	TPA: dynein, axonemal, light intermediate chain 1-like
1230	2	CAJ00236.1	TPA: endonuclease-reverse transcriptase
1231	2	NP_001076512.1	Trace amine associated receptor 13e
1232	2	NP_001187529.1	Transaldolase
1233	2	XP_003447972.1	Transcription activator BRG1-like
1234	2	NP_001187283.1	Transcription elongation factor b polypeptide 2
1235	2	XP_004580233.1	Transcription elongation regulator 1-like protein
1236	2	XP_689373.2	Transcription factor 20-like
1237	2	XP_002667613.2	Transcription factor AP-4 isoformx2
1238	2	NP_001076493.1	Transcription factor ETV6
1239	2	XP_004564447.1	Transcription factor HIVEP3-like isoform X1
1240	2	ABD77103.1	Transcription factor Nkx6.2
1241	2	NP_001038250.1	Transcription initiation factor TFIID subunit 1
1242	2	XP_692554.4	Transcription initiation factor TFIID subunit 4
1243	2	XP_682933.2	Transcription regulator protein BACH2 isoform X2
1244	2	XP_003452926.1	Transducin beta-like protein 3-like
1245	2	NP_001187249.1	Transferrin precursor
1246	2	XP_003961456.1	Transient receptor potential cation channel subfamily M member 4-like
1247	2	NP_001121711.1	Transient receptor potential cation channel subfamily M member 5
1248	2	NP_001238760.1	Transient receptor potential cation channel, subfamily M, member 6
1249	2	NP_001107069.1	Translation initiation factor eif-2B subunit epsilon
1250	2	XP_004077569.1	Transmembrane 9 superfamily member 3-like
1251	2	XP_001922025.3	Transmembrane and TPR repeat-containing protein 1 isoform X1
1252	2	XP_687955.1	Transmembrane protein 119
1253	2	XP_001923340.1	Transmembrane protein 131 isoform X1
1254	2	XP_001920669.1	Transmembrane protein 158-like

1255	2	NP_997848.1	Transmembrane protein 165 precursor
1256	2	XP_003969728.1	Transmembrane protein 178B-like
1257	2	NP_956786.1	Transmembrane protein 68
1258	2	XP_001336187.3	Trehalase
1259	2	NP_998034.1	Tribbles homolog 3
1260	2	XP_689365.4	Trinucleotide repeat-containing gene 6C protein-like
1261	2	NP_001002870.2	Tripartite motif-containing 24
1262	2	XP_003583620.1	Tripartite motif-containing protein 16-like
1263	2	XP_003198957.1	Tripartite motif-containing protein 39
1264	2	XP_004078838.1	Tripartite motif-containing protein 47-like
1265	2	XP_003444536.1	Tripartite motif-containing protein 69-like
1266	2	Q1LUA6.1	Triple functional domain protein
1267	2	NP_955911.1	Trna-dihydrouridine synthase 1-like
1268	2	NP_001070119.1	Trna-splicing endonuclease subunit Sen2
1269	2	ELV09964.1	Trna-splicing ligase rtcb like protein
1270	2	XP_003458237.1	Tropomodulin-3-like
1271	2	XP_001505895.2	Troponin C, slow skeletal and cardiac muscles-like
1272	2	AAP81159.1	Trypsinogen
1273	2	EPQ14630.1	Tubulin alpha-1C chain
1274	2	NP_998355.1	Tumor necrosis factor receptor superfamily member 1A precursor
1275	2	XP_004562670.1	Type I inositol 3,4-bisphosphate 4-phosphatase-like isoform X7
1276	2	NP_001122019.1	Type I inositol-3,4-bisphosphate 4-phosphatase
1277	2	XP_692000.3	Type II inositol-1,4,5-trisphosphate 5-phosphatase-like, partial
1278	2	AAY43816.1	Type IV collagen alpha 4 chain, partial
1279	2	XP_001337899.1	Tyrosine-protein kinase ABL1-like isoform 2
1280	2	A2BIL7.2	Tyrosine-protein kinase BAZ1B; altname: Full=Bromodomain adjacent to zinc finger domain protein 1B; altname: Full=Williams syndrome transcription factor homolog
1281	2	NP_001025391.1	Tyrosine-protein kinase Blk
1282	2	XP_004549652.1	Tyrosine-protein kinase sgk223-like
1283	2	XP_004342494.1	Tyrosine-protein kinase Src42A
1284	2	XP_003201292.1	Tyrosine-protein kinase Tec-like, partial
1285	2	XP_003965393.1	Tyrosine-protein phosphatase non-receptor type 11-like
1286	2	NP_997819.1	Tyrosine-protein phosphatase non-receptor type 2
1287	2	XP_001920562.1	U2 small nuclear ribonucleoprotein auxiliary factor 35 kda subunit-related protein 2 isoform X1
1288	2	NP_001025127.1	U2 small nuclear RNA auxiliary factor 2a
1289	2	XP_003972978.1	Ubiquitin carboxyl-terminal hydrolase 15-like
1290	2	XP_004071467.1	Ubiquitin carboxyl-terminal hydrolase 22-A-like isoform 2
1291	2	XP_002662556.2	Ubiquitin carboxyl-terminal hydrolase 4 isoform 1
1292	2	NP_001002451.1	Ubiquitin fusion degradation protein 1 homolog
1293	2	NP_001076266.1	Ubiquitin-conjugating enzyme E2L 3b
1294	2	NP_998227.1	Ubiquitin-like modifier-activating enzyme 1
1295	2	NP_001007319.1	Ubiquitin-protein ligase E3A

1296	2	XP_003449112.1	UDP-glcnac:betagal beta-1,3-N-acetylglucosaminyltransferase 3-like
1297	2	Q6GMI9.2	UDP-glucuronic acid decarboxylase 1; altname: Full=UDP-glucuronate decarboxylase 1; Short=UXS-1
1298	2	ELK06239.1	UDP-N-acetylglucosamine/UDP-glucose/GDP-mannose transporter
1299	2	NP_001124103.1	Unc-51-like kinase 1a
1300	2	CBN80807.1	Uncharacterized protein
1301	2	XP_685206.2	Uncharacterized protein c10orf118-like
1302	2	XP_689494.2	Uncharacterized protein c17orf53 isoform X2
1303	2	XP_004014456.1	Uncharacterized protein c20orf195 homolog
1304	2	XP_692250.3	Uncharacterized protein c6orf168-like
1305	2	XP_690269.4	Uncharacterized protein c7orf57 homolog isoformx2
1306	2	XP_001923762.1	Uncharacterized protein c8orf45 homolog
1307	2	XP_001341226.3	Uncharacterized protein KIAA1462-like
1308	2	XP_001340925.2	Uncharacterized protein KIAA1486-like
1309	2	XP_003459696.1	Uncharacterized protein KIAA1797-like
1310	2	XP_001333668.2	Uncharacterized protein KIAA1841-like
1311	2	NP_001082917.1	Uncharacterized protein LOC100000104
1312	2	NP_001139098.1	Uncharacterized protein LOC100004731
1313	2	XP_001344050.1	Uncharacterized protein LOC100004848
1314	2	NP_001076533.1	Uncharacterized protein LOC100034467
1315	2	NP_001096105.1	Uncharacterized protein LOC100124608
1316	2	NP_001108043.1	Uncharacterized protein LOC100136852 precursor
1317	2	NP_001116084.1	Uncharacterized protein LOC100142634 precursor
1318	2	NP_001124091.1	Uncharacterized protein LOC100170780
1319	2	NP_001124128.1	Uncharacterized protein LOC100170821
1320	2	XP_002170374.2	Uncharacterized protein LOC100205880, partial
1321	2	XP_002667014.1	Uncharacterized protein LOC100329955
1322	2	XP_002663158.1	Uncharacterized protein LOC100330250 isoform 1
1323	2	XP_004917457.1	Uncharacterized protein LOC100496129
1324	2	XP_003728921.1	Uncharacterized protein LOC100894010
1325	2	XP_003977182.1	Uncharacterized protein LOC101061985, partial
1326	2	XP_003977599.1	Uncharacterized protein LOC101077301
1327	2	XP_004066208.1	Uncharacterized protein LOC101155010
1328	2	XP_004085103.1	Uncharacterized protein LOC101160622, partial
1329	2	XP_004082770.1	Uncharacterized protein LOC101162800
1330	2	XP_004076002.1	Uncharacterized protein LOC101164665
1331	2	XP_004071748.1	Uncharacterized protein LOC101164807
1332	2	XP_004066535.1	Uncharacterized protein LOC101165212
1333	2	XP_004076963.1	Uncharacterized protein LOC101165752
1334	2	XP_004085509.1	Uncharacterized protein LOC101166189
1335	2	XP_004073499.1	Uncharacterized protein LOC101171842
1336	2	XP_004207495.1	Uncharacterized protein LOC101238179, partial
1337	2	XP_004211737.1	Uncharacterized protein LOC101240276

1338	2	XP_004556909.1	Uncharacterized protein LOC101465870
1339	2	XP_004576347.1	Uncharacterized protein LOC101466341
1340	2	XP_004576292.1	Uncharacterized protein LOC101468398, partial
1341	2	XP_004567438.1	Uncharacterized protein LOC101468928
1342	2	XP_004544105.1	Uncharacterized protein LOC101471150
1343	2	XP_004556396.1	Uncharacterized protein LOC101472750
1344	2	XP_004570995.1	Uncharacterized protein LOC101480471 isoform X1
1345	2	XP_004561853.1	Uncharacterized protein LOC101480605
1346	2	XP_004549715.1	Uncharacterized protein LOC101480956
1347	2	XP_004571522.1	Uncharacterized protein LOC101486585
1348	2	XP_004717948.1	Uncharacterized protein LOC101652918
1349	2	XP_004918303.1	Uncharacterized protein LOC101734130, partial
1350	2	XP_004932995.1	Uncharacterized protein LOC101740474
1351	2	XP_005022161.1	Uncharacterized protein LOC101792521 isoform X2
1352	2	XP_001921014.3	Uncharacterized protein LOC322468
1353	2	NP_001002303.1	Uncharacterized protein LOC432374
1354	2	NP_001002668.1	Uncharacterized protein LOC436941
1355	2	NP_001002704.1	Uncharacterized protein LOC436977
1356	2	NP_001103493.1	Uncharacterized protein LOC556322
1357	2	NP_001188466.1	Uncharacterized protein LOC556673
1358	2	NP_001026840.1	Uncharacterized protein LOC557030 precursor
1359	2	NP_001038388.2	Uncharacterized protein LOC560317
1360	2	NP_001073658.1	Uncharacterized protein LOC567225
1361	2	NP_001116773.2	Uncharacterized protein LOC570267 isoform 1
1362	2	XP_003751748.1	Uncharacterized protein LOC684841
1363	2	XP_001198200.2	Uncharacterized protein LOC762541
1364	2	NP_001071257.1	Uncharacterized protein LOC777745
1365	2	NP_001104678.1	Uncharacterized protein LOC791587
1366	2	XP_001338348.1	Uncharacterized protein LOC797886
1367	2	NP_001082915.1	Uncharacterized protein LOC799770 precursor
1368	2	XP_004556852.1	Unconventional myosin-Id-like isoform X1
1369	2	XP_682849.4	Unconventional myosin-Ie isoform X3
1370	2	XP_004564608.1	Unconventional myosin-ixa-like isoform X4
1371	2	XP_004557494.1	Unconventional myosin-XIX-like
1372	2	AAI57389.1	Unknown (protein for IMAGE:7070805)
1373	2	BAG58901.1	Unnamed protein product
1374	2	CAG05210.1	Unnamed protein product
1375	2	CAF95303.1	Unnamed protein product
1376	2	CAG03184.1	Unnamed protein product
1377	2	CAG11022.1	Unnamed protein product
1378	2	CAG07949.1	Unnamed protein product
1379	2	CAF98580.1	Unnamed protein product
1380	2	CAG06088.1	Unnamed protein product

1381	2	CAG11156.1	Unnamed protein product
1382	2	CAF99999.1	Unnamed protein product
1383	2	CAG08447.1	Unnamed protein product
1384	2	CAF89604.1	Unnamed protein product
1385	2	CAG09021.1	Unnamed protein product
1386	2	CAF96150.1	Unnamed protein product
1387	2	CAG10439.1	Unnamed protein product
1388	2	CAG00982.1	Unnamed protein product
1389	2	CAG04647.1	Unnamed protein product
1390	2	CAG01992.1	Unnamed protein product
1391	2	CAG05920.1	Unnamed protein product
1392	2	CAG10684.1	Unnamed protein product
1393	2	CAG01937.1	Unnamed protein product
1394	2	CAG02417.1	Unnamed protein product
1395	2	CAG01444.1	Unnamed protein product
1396	2	CAG09357.1	Unnamed protein product
1397	2	NP_001014353.1	UPF0364 protein c6orf211 homolog
1398	2	NP_001188061.1	Upf0490 protein c1orf201-like protein
1399	2	NP_001025261.3	UPF0536 protein c12orf66 homolog
1400	2	XP_004002321.1	UPF0577 protein KIAA1324 homolog
1401	2	XP_694960.5	UPF0606 protein c11orf41-like
1402	2	NP_001139876.1	UPF0740 protein c1orf192 homolog
1403	2	XP_001334493.1	Urotensin-2 receptor-like
1404	2	XP_697435.4	Usherin
1405	2	NP_001124071.1	USP6 N-terminal-like protein
1406	2	NP_001091659.1	Vacuolar protein sorting-associated protein 16 homolog
1407	2	XP_684689.2	Vacuolar protein sorting-associated protein 52 homolog isoformx2
1408	2	XP_003443007.1	Vam6/Vps39-like protein-like
1409	2	NP_705960.1	Vang-like protein 2
1410	2	NP_991297.1	Vascular endothelial growth factor C precursor
1411	2	Q5GIT4.2	Vascular endothelial growth factor receptor 2; Short=VEGFR-2; altname: Full=Fetal liver kinase 1b; Short=FLK-1b; altname: Full=Kinase insert domain receptor; altname: Full=Kinase insert domain receptor-B; altname: Full=Protein-tyrosine kinase receptor flk
1412	2	NP_571547.1	Vascular endothelial growth factor receptor kdr-like precursor
1413	2	AAB47406.1	Ventral neural cadherin, partial
1414	2	NP_001187192.1	Vertebrate ancient long opsin
1415	2	NP_001009982.1	Vesicular glutamate transporter 2.2
1416	2	CAH04518.2	Vitamin D receptor I
1417	2	XP_003444364.1	Vitamin K-dependent gamma-carboxylase-like
1418	2	ADO29268.1	Vitelline membrane outer layer protein 1
1419	2	XP_003973333.1	Voltage-dependent calcium channel subunit alpha-2/delta-3-like
1420	2	XP_004566258.1	Voltage-dependent calcium channel subunit alpha-2/delta-4-like
1421	2	NP_571975.1	Voltage-dependent L-type calcium channel subunit alpha-1C

1422	2	XP_004554082.1	Voltage-dependent L-type calcium channel subunit alpha-1D-like isoform X1
1423	2	XP_001333514.4	Voltage-dependent L-type calcium channel subunit alpha-1F-like
1424	2	XP_003452886.1	Voltage-dependent T-type calcium channel subunit alpha-1I
1425	2	NP_001076368.1	Vomeronasal 2 receptor, x1 precursor
1426	2	ACY30364.1	Von Willebrand factor A domain containing 5A
1427	2	XP_696917.5	Von Willebrand factor A domain-containing protein 5A-like
1428	2	XP_692918.4	Von Willebrand factor A domain-containing protein 5B1
1429	2	NP_001121810.1	Von Willebrand factor A domain-containing protein 8
1430	2	AAI63680.1	Vps13a protein
1431	2	XP_003964503.1	V-type proton atpase 16 kda proteolipid subunit-like
1432	2	XP_003447124.1	WASH complex subunit 7-like
1433	2	XP_701288.5	WD repeat- and FYVE domain-containing protein 4
1434	2	NP_001120428.1	WD repeat domain 63
1435	2	NP_001161736.1	WD repeat-containing protein 37
1436	2	BAF75865.1	Winged-helix transcription factor RFX4
1437	2	XP_003200415.1	Wu:fb77g05 isoform 1
1438	2	AAI29246.1	Wu:fc13d03 protein
1439	2	AAI27400.1	Wu:fc51g12 protein
1440	2	AAH90218.1	Xdsh protein
1441	2	AAL32047.1	Xiap
1442	2	NP_001012252.1	XK-related protein 7
1443	2	ADV78230.1	Xylosyltransferase 1
1444	2	XP_002660941.2	YEATS domain-containing protein 2
1445	2	XP_003200727.1	YLP motif-containing protein 1
1446	2	BAF44666.1	Zebrafish target of rapamycin
1447	2	AAH93340.1	Zgc:112491
1448	2	AAI53652.1	Zgc:113385
1449	2	AAI33991.1	Zgc:114114 protein
1450	2	AAI22400.1	Zgc:153725
1451	2	AAI24827.1	Zgc:163073 protein
1452	2	AAI34121.1	Zgc:171352 protein
1453	2	AAI71492.1	Zgc:171724
1454	2	AAI71544.1	Zgc:172040 protein
1455	2	AAH44549.1	Zgc:55941
1456	2	NP_956500.2	Zgc:56194 precursor
1457	2	AAH81655.1	Zgc:92353
1458	2	NP_001122011.1	Zinc finger and BTB domain-containing protein 10
1459	2	XP_004552884.1	Zinc finger and BTB domain-containing protein 11-like
1460	2	XP_003443037.1	Zinc finger and BTB domain-containing protein 24-like
1461	2	XP_003451415.1	Zinc finger and BTB domain-containing protein 7A-like
1462	2	XP_001344107.3	Zinc finger E-box-binding homeobox 1
1463	2	XP_003122336.2	Zinc finger protein 160-like

1464	2	XP_003406484.1	Zinc finger protein 180-like
1465	2	XP_002917645.1	Zinc finger protein 271-like
1466	2	XP_684543.3	Zinc finger protein 280D isoform X5
1467	2	ELK38750.1	Zinc finger protein 300
1468	2	XP_001103464.2	Zinc finger protein 341-like
1469	2	XP_698186.3	Zinc finger protein 366
1470	2	XP_004546057.1	Zinc finger protein 512B-like isoform X13
1471	2	XP_003198717.1	Zinc finger protein 729-like
1472	2	EPQ19413.1	Zinc finger protein interacting with ribonucleoprotein K
1473	2	XP_002662446.1	Zinc finger protein ZXDC
1474	2	XP_002666212.1	Zinc finger RNA-binding protein
1475	2	XP_003457931.1	Zinc-binding protein A33-like
1476	2	XP_002667333.2	Zona pellucida sperm-binding protein 4-like, partial
1477	2	XP_004457235.1	Zonadhesin
1478	2	XP_003448552.1	ZZ-type zinc finger-containing protein 3
1479	3	XP_003905111.1	1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-1-like
1480	3	XP_688270.2	5-hydroxytryptamine receptor 2A-like
1481	3	XP_003198768.1	ADAMTS-like protein 2-like
1482	3	XP_003969795.1	ADAMTS-like protein 5-like
1483	3	XP_687183.1	Adenosine deaminase domain-containing protein 2-like
1484	3	XP_003975351.1	Adenylate cyclase type 1-like
1485	3	XP_688042.3	Aggrecan core protein-like
1486	3	NP_001187825.1	Alcohol dehydrogenase (NADP+) b
1487	3	XP_003443446.1	Alcohol dehydrogenase class-3-like
1488	3	AAH75877.1	Aldh3a2 protein
1489	3	AAF81097.1	Alpha actin
1490	3	ABQ45553.1	Alpha-amylase
1491	3	NP_957187.1	AMP deaminase 1
1492	3	AAI62238.1	Angiopoietin 2
1493	3	XP_004572713.1	Ankyrin repeat domain-containing protein 27-like
1494	3	NP_001122002.1	AP-1 complex subunit beta-1
1495	3	NP_955912.1	Arachidonate 12-lipoxygenase, 12S-type
1496	3	CBN81775.1	Aryl hydrocarbon receptor repressor
1497	3	XP_699642.5	Beta-chimaerin
1498	3	XP_004073292.1	Beta-crystallin B1-like
1499	3	NP_001103766.1	BRCA1 interacting protein C-terminal helicase 1
1500	3	AAI51887.1	Brd3b protein
1501	3	XP_003453615.1	Breast carcinoma-amplified sequence 3
1502	3	NP_998531.2	Bromodomain-containing protein 7
1503	3	XP_002662035.2	Butyrophilin-like protein 2-like
1504	3	XP_683698.2	Calcium/calmodulin-dependent protein kinase type 1-like
1505	3	NP_998712.2	Carbohydrate sulfotransferase 12

1506	3	NP_001002066.1	Cardiac muscle alpha actin 1
1507	3	ACI68739.1	Catechol O-methyltransferase
1508	3	XP_002660581.1	Catenin delta-1-like
1509	3	XP_004078011.1	Cation-dependent mannose-6-phosphate receptor-like isoform 2
1510	3	NP_001107027.1	Cell adhesion molecule 2b precursor
1511	3	NP_001139071.1	Cell division cycle and apoptosis regulator protein 1
1512	3	NP_001188286.1	Cell division protein kinase 17
1513	3	XP_003451154.1	Centrosomal protein of 78 kda-like
1514	3	NP_001187847.1	Charged multivesicular body protein 5
1515	3	XP_695866.3	Chloride channel protein 1-like
1516	3	NP_001017634.1	Choline-phosphate cytidylyltransferase A
1517	3	EGV92380.1	CMRF35-like molecule 7
1518	3	XP_002665305.2	Collagen alpha-1(XII) chain
1519	3	XP_684415.4	Copper-transporting atpase 2
1520	3	NP_001188228.1	Cornifelin-like protein b
1521	3	XP_003200688.1	CUB and sushi domain-containing protein 3
1522	3	XP_004570028.1	CUGBP Elav-like family member 2-like isoform X7
1523	3	AAH45948.1	Cyp2j28 protein
1524	3	NP_001007349.1	Cysteine sulfinic acid decarboxylase
1525	3	XP_003447365.1	Cytoplasmic dynein 1 intermediate chain 2-like
1526	3	XP_001367129.2	Death-associated protein kinase 1
1527	3	XP_003217677.1	Dedicator of cytokinesis protein 3-like
1528	3	XP_001919990.1	DENN domain-containing protein 4C
1529	3	NP_001020621.1	DNA (cytosine-5)-methyltransferase 3B
1530	3	XP_004571720.1	DNA excision repair protein ERCC-6-like
1531	3	XP_694437.5	DNA polymerase theta
1532	3	XP_696859.4	DNA repair protein RAD50 isoform X3
1533	3	AAI63893.1	Dnmt1 protein
1534	3	XP_003443286.1	Double-stranded RNA-specific editase 1-like
1535	3	XP_004545269.1	Dynein intermediate chain 1, axonemal-like
1536	3	NP_001187306.1	Dynein light chain 1 axonemal
1537	3	NP_001103995.1	Ectopic P granules protein 5 homolog
1538	3	XP_003201483.1	EF-hand calcium-binding domain-containing protein 6
1539	3	XP_002660945.2	Ephrin type-B receptor 1
1540	3	P54763.2	Ephrin type-B receptor 2; altname: Full=Neural kinase; altname: Full=Nuk receptor tyrosine kinase; altname: Full=Tyrosine-protein kinase receptor EPH-3; altname: Full=Tyrosine-protein kinase receptor SEK-3; Flags: Precursor
1541	3	XP_003448876.1	Epidermal growth factor receptor substrate 15-like 1
1542	3	XP_001919797.2	Exportin-1
1543	3	NP_001187587.1	Fam183a
1544	3	XP_004673881.1	Fatty acid-binding protein, brain
1545	3	NP_991265.1	Fibroblast growth factor 13b
1546	3	XP_001508334.2	Filamin-A isoform 1

1547	3	XP_004743926.1	Formin-like protein 2 isoform X10
1548	3	NP_001025123.1	Fructose-bisphosphate aldolase C-A
1549	3	NP_001038571.1	Furin A precursor
1550	3	CBN81731.1	Gamma-tubulin complex component 2
1551	3	AAP06146.1	Genbank Accession Number X01064 histone H2A in Oncorhynchus mykiss
1552	3	XP_002934073.2	General transcription factor II-I repeat domain-containing protein 2A-like
1553	3	XP_003458249.1	Glycine amidinotransferase, mitochondrial-like
1554	3	XP_684212.3	Glycosyltransferase 25 family member 3 isoform 1
1555	3	XP_003199883.1	GPI ethanolamine phosphate transferase 2 isoform X1
1556	3	NP_001070082.1	GPI transamidase component PIG-S
1557	3	NP_001001949.2	Granulin-a precursor
1558	3	EAW99553.1	Hcg1990835, isoform CRA_a
1559	3	NP_001187115.1	Hemoglobin-beta
1560	3	NP_997754.1	Heterogeneous nuclear ribonucleoprotein H1, like
1561	3	CAX13609.1	Hexokinase 2
1562	3	Q08BR4.2	Histone-lysine N-methyltransferase SETDB1-B; altname: Full=SET domain bifurcated 1B
1563	3	NP_001074154.1	Hyaluronidase PH-20 precursor
1564	3	EHH30125.1	Hypothetical protein EGK_10721, partial
1565	3	XP_001921896.2	Hypothetical protein LOC100148510
1566	3	XP_002667207.2	Hypothetical protein LOC100329940
1567	3	XP_003198458.1	Hypothetical protein LOC100536930
1568	3	XP_003455916.1	Hypothetical protein LOC100694549
1569	3	XP_003200916.1	Hypothetical protein LOC566704
1570	3	XP_695887.4	Hypothetical protein LOC567501
1571	3	XP_003200543.1	Hypothetical protein LOC794433 isoform 2
1572	3	XP_001338900.4	Hypothetical protein LOC798456
1573	3	ELR55924.1	Hypothetical protein M91_03110, partial
1574	3	ELQ35716.1	Hypothetical protein OOU_y34scaffold00692g19
1575	3	EEC71856.1	Hypothetical protein osi_04556
1576	3	XP_002111687.1	Hypothetical protein TRIADDRAFT_55968
1577	3	EGW13722.1	Ig heavy chain V region 441
1578	3	XP_003442444.1	Immunoglobulin superfamily DCC subclass member 3
1579	3	XP_002664353.2	Inositol-trisphosphate 3-kinase B-like
1580	3	NP_001166098.1	Integrin, alpha 11a precursor
1581	3	NP_001187175.1	Intelectin 1 precursor
1582	3	AAV97700.1	Interferon 4
1583	3	AAN41456.1	Interleukin-8 variant 4
1584	3	A7MBP4.2	Intraflagellar transport protein 46 homolog
1585	3	XP_001922323.3	IQ motif and SEC7 domain-containing protein 1-like isoform X1
1586	3	NP_001082833.1	Junctophilin-2
1587	3	XP_690835.4	Kanadaptin

1588	3	AER42657.1	Keratin type II E3
1589	3	XP_001920258.3	Kinesin family member 21A
1590	3	XP_696785.3	Kinesin family member C3 isoform X3
1591	3	NP_001116747.1	Kinesin heavy chain isoform 5C
1592	3	XP_004552913.1	Kinesin-like protein KIF1A-like
1593	3	XP_699677.4	Leucine-rich repeat and death domain-containing protein
1594	3	XP_003201382.1	Leucine-rich repeat transmembrane neuronal protein 2-like
1595	3	XP_003966077.1	Leucine-rich repeat-containing protein 16A-like
1596	3	XP_003452408.1	Leucine-rich repeat-containing protein 7
1597	3	ADO27861.1	Leukocyte elastase inhibitor
1598	3	NP_001136035.1	Lipase maturation factor 2
1599	3	NP_001158846.1	Lysosomal-associated membrane protein 1 precursor
1600	3	XP_004542154.1	MAM domain-containing glycosylphosphatidylinositol anchor protein 2-like
1601	3	NP_001123869.1	MAM domain-containing protein 2 precursor
1602	3	NP_001073429.2	Matrix-remodelling associated 8b precursor
1603	3	EDL23431.1	Mcg18249, isoform CRA_b
1604	3	ACM41864.1	Microfibril-associated 4
1605	3	XP_001920578.3	Mitogen-activated protein kinase kinase kinase 5
1606	3	ABC55064.1	Myocyte-specific enhancer factor 2A
1607	3	NP_001099057.1	Myosin light chain kinase 3
1608	3	NP_694514.2	Myosin, heavy polypeptide 2, fast muscle specific
1609	3	NP_001039230.1	Myotubularin related protein 1
1610	3	XP_004574848.1	Myotubularin-related protein 1-like isoform X2
1611	3	AFB82538.1	Na-bicarbonate cotransporter nbcn1-N
1612	3	XP_001919428.3	NACHT, LRR and PYD domains-containing protein 1
1613	3	CAM60097.1	Ncam2
1614	3	XP_004558643.1	Negative elongation factor B-like
1615	3	NP_001013518.1	Neurobeachin a
1616	3	XP_003443009.1	Neutral alpha-glucosidase C
1617	3	XP_695071.5	NHS-like 1a
1618	3	AAH76490.1	Nptx2 protein
1619	3	NP_001243593.1	NT-3 growth factor receptor precursor
1620	3	NP_571814.1	Odorant receptor, family C, subfamily 102, member 2
1621	3	NP_001187196.1	Olfactory receptor 1
1622	3	AAK14719.1	Olfactory receptor 32D
1623	3	NP_001153312.1	Olfactory receptor C family, g1 precursor
1624	3	NP_001093536.2	Optineurin
1625	3	NP_001077310.1	Pantothenate kinase 1
1626	3	ACF18588.1	Pepsinogen A1 precursor
1627	3	NP_001122138.1	Peptidyl-prolyl cis-trans isomerase FKBP10
1628	3	XP_004570004.1	Peroxidasin-like isoform X2
1629	3	XP_003975400.1	Piezo-type mechanosensitive ion channel component 2-like

1630	3	ACM09649.1	Pleiotrophic factor-alpha-2 precursor
1631	3	NP_001155072.1	Plexin-B2 precursor
1632	3	XP_003197822.1	Polymeric immunoglobulin receptor
1633	3	XP_004573910.1	Potassium voltage-gated channel subfamily H member 3-like
1634	3	XP_004561826.1	Probable ATP-dependent RNA helicase DDX46-like
1635	3	XP_003206612.1	Probable D-tyrosyl-trna(Tyr) deacylase 2-like
1636	3	XP_004072733.1	Probable E3 ubiquitin-protein ligase HECTD4-like
1637	3	XP_004074388.1	Protein 4.1-like
1638	3	NP_001121834.1	Protein FAM117B
1639	3	NP_001020721.1	Protein FAM167A
1640	3	XP_692886.5	Protein kinase C-binding protein 1
1641	3	NP_998313.1	Protein LSM12 homolog A
1642	3	XP_001921217.3	Protein piccolo
1643	3	XP_694356.4	Protein SON-like isoform X3
1644	3	XP_697257.2	Protein TANC2 isoform X2
1645	3	NP_001019272.1	Protocadherin 1 alpha 6 precursor
1646	3	EHB08104.1	Protocadherin Fat 4, partial
1647	3	XP_003199877.1	Protocadherin gamma-A8-like
1648	3	CBN81648.1	Putative all-trans-retinol 13,14-reductase
1649	3	Q6DN03.3	Putative histone H2B type 2-C; altname: Full=Histone H2B.t; Short=H2B/t
1650	3	XP_001920508.1	Ral gtpase-activating protein subunit alpha-2-like
1651	3	NP_001138713.1	Rapunzel2
1652	3	NP_001187761.1	Ras-related protein rab-27b
1653	3	XP_004386501.1	Receptor-type tyrosine-protein phosphatase beta isoform 3
1654	3	XP_003459480.1	Receptor-type tyrosine-protein phosphatase mu-like, partial
1655	3	NP_001017781.1	Rho gtpase-activating protein 1
1656	3	XP_004563633.1	Rho guanine nucleotide exchange factor 40-like
1657	3	NP_001074132.1	RIB43A-like with coiled-coils protein 1
1658	3	XP_003455118.1	Roundabout homolog 2-like
1659	3	NP_956165.1	S-adenosylmethionine synthase isoform type-1
1660	3	XP_004539650.1	Semaphorin-4B-like
1661	3	XP_003909157.1	Septin-10 isoform 1
1662	3	XP_003969117.1	Serine/threonine-protein kinase pdik1l-like
1663	3	XP_002661633.2	Serine/threonine-protein kinase pim-3-like
1664	3	XP_002665971.2	Serine/threonine-protein kinase ULK1 isoform X1
1665	3	XP_004562212.1	Serine-rich adhesin for platelets-like
1666	3	NP_991103.1	SET and MYND domain containing 1
1667	3	AAI53394.1	Si:ch211-154o6.6 protein
1668	3	XP_696782.5	Small conductance calcium-activated potassium channel protein 3-like
1669	3	XP_001923596.1	Sodium- and chloride-dependent neutral and basic amino acid transporter B(0+)
1670	3	XP_003763944.1	Sodium-driven chloride bicarbonate exchanger

1671	3	XP_423637.3	Solute carrier family 2, facilitated glucose transporter member 6 isoform X3
1672	3	NP_001091726.2	Solute carrier family 9 (sodium/hydrogen exchanger), member 6a precursor
1673	3	XP_004565556.1	Solute carrier organic anion transporter family member 4A1-like isoform X1
1674	3	XP_698075.3	Spectrin beta chain, brain 4-like
1675	3	CBN80839.1	Sterol O-acyltransferase 2
1676	3	XP_001338475.1	Synaptotagmin-15-like isoform X1
1677	3	NP_956146.1	TAF5-like RNA polymerase II p300/CBP-associated factor-associated factor 65 kda subunit 5L
1678	3	XP_003440476.1	Talin-2-like
1679	3	NP_001077325.1	Taste receptor, type 1, member 2, tandem duplicate 2 precursor
1680	3	NP_999961.1	TATA box-binding protein-like protein 2
1681	3	XP_004554604.1	TATA-binding protein-associated factor 172-like
1682	3	ACI68751.1	T-cell leukemia translocation-altered gene protein homolog
1683	3	CAD29787.1	TCR-alpha V segment II-73
1684	3	XP_003975627.1	Tenascin-R-like
1685	3	XP_004554617.1	Tetratricopeptide repeat protein 40-like isoform X3
1686	3	XP_003442708.1	Thioredoxin reductase 3
1687	3	NP_001104618.1	Tonsoku-like protein
1688	3	NP_001076372.1	Trace amine-associated receptor 10b
1689	3	NP_001070068.1	TRAF family member-associated NF-kappa-B activator
1690	3	XP_004562380.1	Transformation/transcription domain-associated protein-like isoform X1
1691	3	NP_998502.1	Transmembrane protein 59-like precursor
1692	3	XP_003199318.1	Transmembrane protein ENSP00000382582 homolog
1693	3	NP_001188203.1	Ubiquitin cross-reactive protein
1694	3	XP_691141.2	Ubiquitin-like modifier activating enzyme 5
1695	3	XP_002938353.2	Uncharacterized protein LOC100490320
1696	3	XP_003910210.1	Uncharacterized protein LOC101008912
1697	3	XP_004086241.1	Uncharacterized protein LOC101174731
1698	3	XP_004083839.1	Uncharacterized protein LOC101175121
1699	3	XP_004916564.1	Uncharacterized protein LOC101730351
1700	3	XP_003199962.1	Uncharacterized protein LOC368429 isoform X1
1701	3	NP_956749.1	Uncharacterized protein LOC393427
1702	3	NP_991205.1	Uncharacterized protein LOC402939
1703	3	NP_001017689.1	Uncharacterized protein LOC550384
1704	3	NP_001077293.2	Uncharacterized protein LOC556440 precursor
1705	3	NP_001073433.1	Uncharacterized protein LOC556929
1706	3	NP_001099060.1	Uncharacterized protein LOC565650
1707	3	NP_001038685.1	Uncharacterized protein LOC571482
1708	3	XP_001199932.1	Uncharacterized protein LOC763830, partial
1709	3	NP_001070783.1	Uncharacterized protein LOC768172 precursor
1710	3	XP_004079015.1	Unconventional myosin-viia-like

1711	3	BAE00445.1	Unnamed protein product
1712	3	BAE30426.1	Unnamed protein product
1713	3	CAG09314.1	Unnamed protein product
1714	3	CAF92719.1	Unnamed protein product
1715	3	CCD15206.1	Unnamed protein product, partial
1716	3	NP_001158709.1	UPF0360 protein c11orf60 homolog
1717	3	NP_001032797.1	Vacuolar protein sorting-associated protein 11 homolog
1718	3	XP_682807.5	Valine--trna ligase
1719	3	XP_003444255.1	Vigilin-like
1720	3	XP_004559815.1	Vitrin-like isoform X3
1721	3	NP_990072.1	Voltage-dependent anion-selective channel protein 2
1722	3	XP_002665969.1	Voltage-dependent P/Q-type calcium channel subunit alpha-1A
1723	3	XP_001920550.3	Voltage-dependent T-type calcium channel subunit alpha-1G
1724	3	NP_001116219.1	V-type proton atpase 116 kda subunit a isoform 2
1725	3	EMP39841.1	WD repeat-containing protein 49
1726	3	XP_002931652.2	WD repeat-containing protein 65
1727	3	ADO28249.1	WD repeat-containing protein 82
1728	3	XP_696898.4	Wu:fb77e12
1729	3	AAI39611.1	Zgc:162621 protein
1730	3	AAI53507.1	Zgc:174162 protein
1731	3	AAH85639.1	Zgc:92122
1732	3	NP_001098574.1	Zona pellucida glycoprotein 2, like 1 precursor
1733	3	XP_685613.3	Zona pellucida sperm-binding protein 3 isoform 1
1734	4	XP_004086821.1	Actin filament-associated protein 1-like
1735	4	XP_684508.3	AMP deaminase 3-like
1736	4	NP_001091653.1	Ankyrin repeat domain 34B
1737	4	XP_003967136.1	Aryl hydrocarbon receptor nuclear translocator-like protein 1-like
1738	4	XP_003970226.1	Aryl-hydrocarbon-interacting protein-like 1-like
1739	4	XP_004539424.1	Asialoglycoprotein receptor 2-like
1740	4	AAI55125.1	Atl2 protein
1741	4	XP_003968306.1	ATP-sensitive inward rectifier potassium channel 1-like
1742	4	XP_004540719.1	Autophagy-related protein 2 homolog B-like
1743	4	XP_003198101.1	Butyrophilin subfamily 1 member A1-like
1744	4	XP_002661517.2	Cadherin EGF LAG seven-pass G-type receptor 1
1745	4	XP_005000797.1	Caseinolytic peptidase B protein homolog isoform X4
1746	4	NP_001152844.1	Catenin alpha-2
1747	4	XP_004564635.1	Chromodomain-helicase-DNA-binding protein 1-like
1748	4	XP_001922046.3	Collagen alpha-1(XIV) chain
1749	4	XP_001340839.4	Contactin-3-like
1750	4	NP_001187380.1	Creatine kinase b-type
1751	4	CBN81753.1	Cullin-2
1752	4	XP_003975124.1	Death-associated protein kinase 1-like
1753	4	XP_003312844.1	Deleted in malignant brain tumors 1 protein-like isoform 3

1754	4	NP_001038734.1	DEP domain-containing protein 7
1755	4	XP_691731.5	Disintegrin and metalloproteinase domain-containing protein 12
1756	4	NP_956526.1	DNA-directed RNA polymerase III subunit RPC3
1757	4	XP_698507.4	Dynein heavy chain 10, axonemal
1758	4	XP_002663728.1	E3 ubiquitin-protein ligase DTX3L
1759	4	NP_001002504.2	E3 ubiquitin-protein ligase HECTD1
1760	4	XP_689534.4	Elongator complex protein 1
1761	4	XP_002741080.1	Enzymatic polyprotein, putative-like, partial
1762	4	ADH21447.1	Excitatory amino acid transporter SLC1A6
1763	4	AAI63545.1	Family with sequence similarity 65, member A
1764	4	XP_001344855.1	F-box/LRR-repeat protein 7
1765	4	NP_001092114.1	Fig-Hepta protein
1766	4	XP_003768925.1	Forkhead box protein K2, partial
1767	4	XP_004569838.1	Germ cell-specific gene 1-like protein-like
1768	4	NP_001116760.1	Glutamate receptor interacting protein 2
1769	4	XP_002665330.2	Glycerophosphodiester phosphodiesterase domain-containing protein 5-like
1770	4	XP_002662587.1	Growth/differentiation factor 5-like
1771	4	ACO08828.1	GSK-3-binding protein
1772	4	ABD83903.1	Gtpase IMAP family member 8-like
1773	4	NP_001188201.1	Hemoglobin subunit alpha
1774	4	ACO14339.1	HERV-H LTR-associating protein 2 precursor
1775	4	AAH97075.1	Hiat1b protein
1776	4	EHH58911.1	Hypothetical protein EGM_08878
1777	4	XP_002662026.2	Hypothetical protein LOC100333032
1778	4	XP_003340863.1	Hypothetical protein LOC100617363
1779	4	XP_003449587.1	Hypothetical protein LOC100697617
1780	4	XP_003450494.1	Hypothetical protein LOC100698500
1781	4	XP_003441353.1	Hypothetical protein LOC100700717
1782	4	XP_003446549.1	Hypothetical protein LOC100709248
1783	4	XP_003459981.1	Hypothetical protein LOC100709753, partial
1784	4	XP_701610.1	Hypothetical protein LOC368413
1785	4	XP_685984.5	Hypothetical protein LOC557772
1786	4	EFA12620.1	Hypothetical protein tcasga2_TC010247
1787	4	ADO28294.1	Ictacalcin
1788	4	T18537	Ig heavy chain - channel catfish
1789	4	AAA56683.1	Ig heavy chain V region, partial
1790	4	ABD85545.1	Immune-type receptor 5
1791	4	NP_001038275.1	Immunity-related gtpase family, q2
1792	4	AAA49336.1	Immunoglobulin heavy chain V-region, partial
1793	4	XP_001922368.1	Inositol-tetrakisphosphate 1-kinase
1794	4	ACN41862.1	Insulin-like growth factor binding protein 2
1795	4	ACN11353.1	Interferon-induced guanylate-binding protein 1

1796	4	XP_001339650.4	Kinesin heavy chain isoform 5A-like
1797	4	XP_004544787.1	Kinetochore-associated protein 1-like
1798	4	XP_004554409.1	Leucine-rich repeat and fibronectin type-III domain-containing protein 2-like isoform X1
1799	4	NP_001181916.1	Low-density lipoprotein receptor-related protein 2 precursor
1800	4	XP_003962888.1	MAP/microtubule affinity-regulating kinase 3-like isoform 4
1801	4	NP_001077307.1	Mediator of RNA polymerase II transcription subunit 13-like
1802	4	NP_001038234.2	Methyltransferase-like protein 13
1803	4	AAB67871.1	MHC class II beta chain
1804	4	XP_001923807.2	Mitochondrial glutamate carrier 1 isoform X1
1805	4	XP_689424.4	Mitogen-activated protein kinase kinase kinase 10
1806	4	XP_001335920.3	NACHT, LRR and PYD domains-containing protein 3
1807	4	ADO27823.1	Natterin-like protein
1808	4	XP_002660697.2	Neoverrucotoxin subunit beta-like
1809	4	CAH19212.1	NILT1 leucocyte receptor
1810	4	XP_002664688.1	Nuclear fragile X mental retardation-interacting protein 2
1811	4	XP_005026871.1	Nuclear transcription factor Y subunit alpha isoform X4
1812	4	NP_001030339.1	Odorant receptor, family F, subfamily 115, member 7
1813	4	NP_001121860.1	Odorant receptor, family F, subfamily 117, member 1
1814	4	NP_001074079.1	Olfactory receptor C family, d2 precursor
1815	4	ABD75543.1	Orf1ab polyprotein
1816	4	BAE46428.1	ORF1-encoded protein
1817	4	XP_688274.4	Otogelin-like
1818	4	XP_001921651.2	P2y purinoceptor 4-like
1819	4	ABX44849.1	Pepsinogen C
1820	4	BAJ07835.1	Perforin 1
1821	4	XP_700029.4	Phosphatidylinositol 4-phosphate 3-kinase C2 domain-containing subunit alpha isoform X3
1822	4	NP_001018438.1	Phosphatidylinositol-4-phosphate 5-kinase type-1 gamma
1823	4	AAH45836.1	Pim1 protein
1824	4	NP_001121714.1	Plasma membrane calcium-transporting atpase 3
1825	4	XP_003460138.1	Polymeric immunoglobulin receptor-like
1826	4	EGW14493.1	Polysialoglycoprotein
1827	4	XP_684855.3	Potassium voltage-gated channel subfamily KQT member 5
1828	4	XP_222205.6	Probable E3 ubiquitin-protein ligase c12orf51-like
1829	4	XP_004550004.1	Proprotein convertase subtilisin/kexin type 5-like
1830	4	EMP41675.1	Proprotein convertase subtilisin/kexin type 6
1831	4	AEU08947.1	Protein arginine methyltransferase, partial
1832	4	XP_001343106.4	Protein FAM135B-like
1833	4	NP_001071224.1	Protein FAM188A
1834	4	XP_003970733.1	Protocadherin Fat 4-like
1835	4	NP_001038428.1	Rab-like protein 2B
1836	4	NP_001103594.1	Rapunzel 5
1837	4	XP_003453243.1	Receptor tyrosine-protein kinase erbb-4-like

1838	4	XP_001919055.3	Receptor-type tyrosine-protein phosphatase gamma
1839	4	XP_005023660.1	Regulator of G-protein signaling 7
1840	4	XP_001918968.2	Rho gtpase-activating protein 21-B
1841	4	XP_004570243.1	Rho guanine nucleotide exchange factor 18-like isoform X4
1842	4	XP_706962.2	RNA polymerase I-specific transcription initiation factor RRN3 isoformx2
1843	4	XP_003201673.1	Semaphorin-7A-like, partial
1844	4	AAH97041.1	Sept7a protein
1845	4	NP_001038590.1	Serine/threonine-protein kinase MRCK beta
1846	4	AAI55144.1	Si:dkey-218f9.5 protein
1847	4	XP_004541758.1	Sialic acid-binding Ig-like lectin 8-like
1848	4	XP_003962411.1	Son of sevenless homolog 2-like
1849	4	XP_002663445.1	Sterol 26-hydroxylase, mitochondrial isoform 1
1850	4	CCH50976.1	T4.15
1851	4	AEL12156.1	T-cell receptor delta
1852	4	ABD93353.1	T-cell receptor gamma TS32.17
1853	4	NP_853634.1	Transcription activator BRG1
1854	4	XP_004071250.1	Transmembrane protease serine 9-like
1855	4	XP_003223238.1	Tubulin alpha-1A chain-like isoform 1
1856	4	E7F6T8.1	Ubiquitin carboxyl-terminal hydrolase 37; altname: Full=Deubiquitinating enzyme 37; altname: Full=Ubiquitin thioesterase 37; altname: Full=Ubiquitin-specific-processing protease 37
1857	4	NP_001188116.1	Uncharacterized protein c1orf43-like protein
1858	4	XP_002940938.2	Uncharacterized protein LOC100493960
1859	4	XP_003723921.1	Uncharacterized protein LOC100891635
1860	4	XP_004085839.1	Uncharacterized protein LOC101169091
1861	4	XP_004080047.1	Uncharacterized protein LOC101173507
1862	4	XP_004085732.1	Uncharacterized protein LOC101174564
1863	4	XP_004572435.1	Uncharacterized protein LOC101465255
1864	4	XP_004560288.1	Uncharacterized protein LOC101483591
1865	4	XP_004916525.1	Uncharacterized protein LOC101733395
1866	4	NP_001093472.1	Uncharacterized protein LOC561255 precursor
1867	4	NP_001124250.1	Uncharacterized protein LOC563446
1868	4	XP_003753034.1	Uncharacterized protein LOC679994
1869	4	NP_001070110.1	Uncharacterized protein LOC767704 precursor
1870	4	CAG10804.1	Unnamed protein product
1871	4	CAG08667.1	Unnamed protein product
1872	4	CAG06226.1	Unnamed protein product
1873	4	CAG10910.1	Unnamed protein product
1874	4	NP_956532.1	Villin-1
1875	4	XP_003456619.1	Voltage-dependent T-type calcium channel subunit alpha-1H-like
1876	4	XP_003438001.1	WD repeat-containing protein 47-like
1877	4	AAH90743.1	Zgc:110752

1878	4	AAH65891.1	Zgc:136713 protein
1879	4	AAH55609.1	Zgc:77174 protein, partial
1880	4	XP_685966.5	Zinc finger CCCH domain-containing protein 7B
1881	4	EPQ09840.1	Zinc finger protein 850
1882	4	XP_004078886.1	Zinc finger protein Xfin-like
1883	5	XP_003452630.1	Camp-dependent protein kinase catalytic subunit alpha-like isoform 3
1884	5	ABA54967.1	CC chemokine SCYA120
1885	5	AAV65851.1	CD45 precursor
1886	5	ADJ80991.1	Cr1-3
1887	5	XP_002662052.1	C-X-C chemokine receptor type 2-like
1888	5	XP_004539352.1	E3 ubiquitin-protein ligase TRIM21-like
1889	5	XP_004540665.1	E3 ubiquitin-protein ligase UBR1-like
1890	5	NP_001038735.2	Fintrim family, member 14
1891	5	JE0324	Gammas-crystallin 2 - catfish
1892	5	XP_700069.5	Glutamate [NMDA] receptor subunit 3A-like
1893	5	XP_004552194.1	Hemicentin-1-like isoform X2
1894	5	EKC18365.1	Histone-lysine N-methyltransferase PRDM9
1895	5	XP_001345924.3	Hypothetical protein LOC100007481
1896	5	XP_003199358.1	Hypothetical protein LOC100534768
1897	5	XP_003443479.1	Hypothetical protein LOC100692970
1898	5	AAA82596.1	Immunoglobulin light chain F class
1899	5	XP_004552247.1	Inactive rhomboid protein 1-like isoform X3
1900	5	XP_693683.4	Interferon-inducible gtpase 5
1901	5	AAP42145.1	Interferon-inducible protein IFI58
1902	5	XP_004544524.1	Kinesin-like protein KIF2A-like isoform X4
1903	5	AAI71344.1	LATS, large tumor suppressor, homolog 1 (Drosophila)
1904	5	XP_004573311.1	Lymphocyte antigen 75-like
1905	5	XP_002664258.2	Lysyl oxidase homolog 4
1906	5	XP_693782.3	Macrophage-capping protein isoform X2
1907	5	XP_001922055.1	Metabotropic glutamate receptor 8-like
1908	5	NP_001187832.1	Metallophosphoesterase 1
1909	5	AAD08648.1	MHC class I alpha chain
1910	5	XP_002936537.2	Mucin-5AC-like
1911	5	XP_684539.4	Myosin-Ih
1912	5	NP_001018557.1	Myotubularin related protein 1a
1913	5	NP_998116.1	Neurofibromin 2b (merlin)
1914	5	CAD87780.1	Novel protein similar to human titin (TTN)
1915	5	NP_001187199.1	Olfactory receptor 8
1916	5	XP_001345586.3	Otoancorin-like
1917	5	NP_001187813.1	Pancreatic triacylglycerol lipase precursor
1918	5	XP_003459115.1	Perilipin-2-like
1919	5	XP_004547459.1	Plasma membrane calcium-transporting atpase 1-like isoform X4
1920	5	XP_002338344.1	Predicted protein

1921	5	NP_001232875.1	Prostate stem cell antigen precursor-like
1922	5	BAB18860.1	Protein kinase raf 1
1923	5	NP_001037809.1	Protein Wnt-2b
1924	5	AGL94530.1	Putative general transcription factor II-I repeat domain containing protein 2-like protein
1925	5	AAW38958.1	Putative pheromone receptor cppr9
1926	5	XP_003200207.1	RNA-binding protein 12B
1927	5	XP_005022027.1	Serine/threonine-protein kinase MRCK alpha-like isoform X3
1928	5	XP_689331.3	Serine/threonine-protein kinase N1 isoform X3
1929	5	NP_001096096.2	Slow myosin heavy chain 2
1930	5	XP_005008025.1	Tigger transposable element-derived protein 1-like
1931	5	AEI59668.1	Tlr5-1
1932	5	XP_001923717.3	Trace amine-associated receptor 7a-like
1933	5	NP_001177979.1	UDP glucuronosyltransferase 2 family, polypeptide A1 precursor
1934	5	XP_002660711.1	Uncharacterized protein LOC100334929
1935	5	XP_003970342.1	Uncharacterized protein LOC101075470
1936	5	XP_004562696.1	Uncharacterized protein LOC101464921
1937	5	NP_001013353.1	Uncharacterized protein LOC503757
1938	5	XP_003971469.1	Unconventional myosin-VI-like
1939	5	CAF99183.1	Unnamed protein product
1940	5	CAG06945.1	Unnamed protein product
1941	6	XP_004074489.1	Adenylate cyclase type 2-like
1942	6	XP_685605.5	A-kinase anchor protein 13
1943	6	AAH93239.1	Apolipoprotein A-IV
1944	6	XP_003811686.1	Calcium-dependent secretion activator 1
1945	6	XP_004085552.1	Coiled-coil domain-containing protein 109B-like, partial
1946	6	XP_003200902.1	C-type mannose receptor 2-like
1947	6	XP_003457703.1	Dynein heavy chain 2, axonemal-like
1948	6	XP_001345641.1	Endonuclease domain-containing 1 protein-like
1949	6	XP_004562487.1	Epidermal growth factor receptor kinase substrate 8-like protein 1-like isoform X2
1950	6	ABB29997.1	Fbp32
1951	6	ADO27993.1	Heme oxygenase
1952	6	XP_002931732.2	Hydrocephalus-inducing protein homolog
1953	6	XP_002663377.2	Hypothetical protein LOC100331483
1954	6	XP_003200011.1	Hypothetical protein LOC100536704, partial
1955	6	XP_003458758.1	Hypothetical protein LOC100700692
1956	6	XP_003453997.1	Hypothetical protein LOC100710874
1957	6	XP_003200593.1	Hypothetical protein LOC497432
1958	6	XP_687197.4	Hypothetical protein LOC558838
1959	6	NP_001038406.1	Kruppel-like factor 11a
1960	6	XP_004553025.1	Meprin A subunit beta-like
1961	6	BAE16584.1	Myosin heavy chain
1962	6	XP_001924014.2	Myosin heavy chain, fast skeletal muscle

1963	6	NP_001187028.1	Novel immune-type receptor 10 precursor
1964	6	XP_001920983.3	Plexin-B1
1965	6	NP_001004966.1	Protein FAM206A
1966	6	XP_004568408.1	Protein kinase C beta type-like isoform X1
1967	6	NP_958450.1	Putative adenosylhomocysteinase 2
1968	6	CBN82168.1	Putative nuclease HARBI1
1969	6	NP_001075100.1	Rho guanine nucleotide exchange factor (GEF) 1a
1970	6	XP_004077688.1	Sodium- and chloride-dependent transporter XTRP3-like isoform 1
1971	6	XP_691651.5	Teneurin-2
1972	6	NP_001076524.1	Trace amine associated receptor 14g
1973	6	XP_002161008.2	Uncharacterized protein LOC100207270
1974	6	XP_004084042.1	Uncharacterized protein LOC101164525
1975	6	CAF87187.1	Unnamed protein product
1976	6	AAH66463.1	Vat1 protein, partial
1977	6	XP_695006.1	WD repeat-containing protein 67
1978	6	XP_003460684.1	Zinc finger BED domain-containing protein 5-like, partial
1979	6	ELK25545.1	Zinc finger protein 256
1980	6	AGK38253.1	Zona pellucida 3
1981	7	NP_001120951.1	Bloodthirsty-related gene family, member 6
1982	7	BAK32934.1	C3 complement component
1983	7	XP_697137.2	Cystine/glutamate transporter-like
1984	7	ACO09168.1	Fumarylacetoacetase
1985	7	XP_683950.5	Galactose-3-O-sulfotransferase 2-like, partial
1986	7	XP_003445604.1	Gamma-glutamyltranspeptidase 1-like
1987	7	ELU03397.1	Hypothetical protein CAPTEDRAFT_219998
1988	7	EFA06499.1	Hypothetical protein tcasga2_TC009398
1989	7	XP_004622049.1	Ig heavy chain Mem5-like
1990	7	NP_001187136.1	Leukocyte immune-type receptor 3 precursor
1991	7	XP_004025562.1	Microtubule-actin cross-linking factor 1, isoforms 1/2/3/5
1992	7	XP_696367.4	NACHT, LRR and PYD domains-containing protein 14
1993	7	AAI63724.1	Nuclear receptor coactivator 2
1994	7	XP_001335383.3	Olfactomedin-4-like
1995	7	NP_001191383.1	P2Y purinoceptor 11
1996	7	XP_688322.5	Protein sel-1 homolog 3 isoform X3
1997	7	XP_696060.5	Solute carrier family 12 member 5
1998	7	CBN80638.1	Uncharacterized protein
1999	7	XP_003970351.1	Uncharacterized protein LOC101077501
2000	7	XP_004087031.1	Uncharacterized protein LOC101169431
2001	7	XP_004554031.1	Uncharacterized protein LOC101484061
2002	7	XP_003458962.1	Vomeronasal type-2 receptor 1-like
2003	7	XP_004575946.1	Zinc finger CCHC domain-containing protein 3-like
2004	8	XP_002660654.1	B-cell receptor CD22-like
2005	8	ABN49522.1	Complement component C9

2006	8	XP_001921252.3	Dynein heavy chain 6, axonemal
2007	8	BAM65012.1	Egg envelope protein
2008	8	CBA11992.1	Endonuclease-reverse transcriptase hmrte-e01
2009	8	XP_004542413.1	G-protein coupled receptor family C group 6 member A-like
2010	8	AAP06505.1	Hypothetical protein
2011	8	XP_002414393.1	Hypothetical protein iscw_ISCW012389
2012	8	XP_003200649.1	Hypothetical protein LOC100333514
2013	8	XP_003457921.1	Hypothetical protein LOC100707135
2014	8	XP_697912.5	Hypothetical protein LOC569436
2015	8	EIE86569.1	Hypothetical protein RO3G_11280
2016	8	XP_004086428.1	Laminin subunit alpha-5-like
2017	8	XP_004569737.1	Myocardin-like isoform X1
2018	8	AEN55534.1	Piwi-like 1
2019	8	XP_001921941.3	Polymeric immunoglobulin receptor-like
2020	8	XP_003447874.1	Protein FAM179B-like
2021	8	XP_003198264.1	Secretory phospholipase A2 receptor
2022	8	XP_689656.5	Serine/threonine-protein kinase WNK1
2023	8	AEM44667.1	T cell receptor alpha
2024	8	XP_003458433.1	Titin
2025	8	CBN80683.1	Transcription factor HES-5
2026	8	XP_001201605.2	Uncharacterized protein LOC764941
2027	8	AAH81616.1	Yrk protein
2028	9	XP_004066520.1	Alpha-2-macroglobulin-like protein 1-like
2029	9	XP_004545892.1	COP9 signalosome complex subunit 6-like
2030	9	XP_002666734.2	Dynein heavy chain 5, axonemal
2031	9	AFC88124.1	Immunoglobulin light chain variable region
2032	9	NP_001187140.1	Interferon-induced GTP-binding protein Mx2
2033	9	XP_001346001.1	Interferon-induced protein 44 isoform X1
2034	9	NP_956831.1	Lysophospholipid acyltransferase 7
2035	9	XP_003442267.1	Myosin heavy chain, fast skeletal muscle-like
2036	9	NP_001187170.1	Novel immune-type receptor 13
2037	9	P10272.1	Pol polyprotein; Contains:
2038	9	XP_004227302.1	Putative nuclease HARBI1-like
2039	9	XP_003794978.1	Sodium-dependent phosphate transport protein 2C
2040	9	XP_004067849.1	Uncharacterized protein LOC101156397
2041	10	NP_001092213.1	C-C chemokine receptor type 7 precursor
2042	10	XP_004686738.1	Fibrillin-2
2043	10	XP_002662769.1	Hypothetical protein LOC100330014
2044	10	XP_003198150.1	Hypothetical protein LOC100535957
2045	10	XP_003201143.1	Hypothetical protein LOC100537953
2046	10	XP_004570794.1	NACHT, LRR and PYD domains-containing protein 3-like
2047	10	AAC64076.1	Putative odorant receptor
2048	10	NP_001002648.1	Solute carrier family 38, member 5b

2049	10	XP_004570672.1	Spectrin beta chain, non-erythrocytic 1-like
2050	10	XP_004552155.1	Uncharacterized protein LOC101481353
2051	10	XP_003458350.1	Zinc finger C3H1 domain-containing protein-like
2052	11	XP_003198303.1	Deleted in malignant brain tumors 1 protein
2053	11	XP_001919828.2	Dynein beta chain, ciliary
2054	11	XP_003444287.1	Guanine nucleotide exchange factor DBS-like
2055	11	XP_003458913.1	Hypothetical protein LOC100699620
2056	11	XP_001919987.3	Hypothetical protein LOC559221
2057	11	ABI16036.1	Leukocyte immune-type receptor TS32.15 L1.1a, partial
2058	11	NP_001187172.1	Novel immune-type receptor 15a precursor
2059	11	XP_002337912.1	Predicted protein
2060	11	AAD56895.1	TCR alpha variable region
2061	11	CAF90645.1	Unnamed protein product
2062	12	XP_003451558.1	Hypothetical protein LOC100693979
2063	12	XP_003786411.1	Kelch-like protein 10 isoform 1
2064	12	NP_001120779.1	Protein deltex-2
2065	12	XP_004917444.1	SCAN domain-containing protein 3-like
2066	12	XP_004914305.1	Uncharacterized protein LOC101733602
2067	13	XP_001344220.3	C-C chemokine receptor type 5-like
2068	13	XP_003197744.1	Hypothetical protein LOC100536106, partial
2069	13	XP_003201051.1	Hypothetical protein LOC568716
2070	13	XP_002922168.1	Kelch-like protein 10-like
2071	13	CAF89010.1	Unnamed protein product
2072	14	XP_004541012.1	Kinesin-like protein KIF3A-like isoform X2
2073	14	XP_004074780.1	Macrophage mannose receptor 1-like
2074	14	XP_003706152.1	Uncharacterized protein LOC100881508
2075	14	XP_004539372.1	Zinc finger BED domain-containing protein 1-like
2076	15	XP_002660539.2	Hypothetical protein LOC100333395
2077	15	XP_001921142.3	Hypothetical protein LOC799338
2078	15	XP_003454888.1	Protein jagged-1b-like
2079	15	YP_073558.1	RNA-dependent DNA polymerase
2080	15	XP_002166082.2	Sacsin-like
2081	15	XP_003340812.1	Zinc finger protein 420-like
2082	16	NP_001070611.1	5'-nucleotidase domain-containing protein 2
2083	16	NP_001187539.1	Gamma-crystallin m2
2084	16	AAI63627.1	Mib2 protein
2085	17	XP_003979358.1	General transcription factor II-I repeat domain-containing protein 2-like
2086	18	XP_003445273.1	PDZ domain-containing protein 11-like isoform 1
2087	18	XP_003341174.1	Zinc finger protein 850-like
2088	19	NP_998581.1	Elongation of very long chain fatty acids (FEN1/Elo2, SUR4/Elo3, yeast)-like 1b
2089	19	NP_957225.1	KDEL motif-containing protein 1 precursor
2090	19	YP_443922.1	Polyprotein

2091	20	XP_001921590.3	E3 ubiquitin-protein ligase UBR2 isoform X1
2092	20	XP_003457483.1	Hypothetical protein LOC100695698
2093	20	XP_004550746.1	Interferon-induced very large gtpase 1-like
2094	21	XP_690143.3	Dynein heavy chain 7, axonemal
2095	21	AAA16654.1	Immunoglobulin light chain
2096	21	XP_003978992.1	NACHT, LRR and PYD domains-containing protein 12-like, partial
2097	22	AAL35359.2	Lambda-recombinase-like protein
2098	23	XP_004575427.1	Importin-5-like
2099	23	XP_004211780.1	Zinc finger BED domain-containing protein 1-like, partial
2100	24	XP_003198393.1	Hypothetical protein LOC100331523
2101	24	NP_001187173.1	Novel immune-type receptor 15b precursor
2102	26	EHJ78503.1	Endonuclease-reverse transcriptase
2103	26	XP_003444806.1	Hypothetical protein LOC100710310
2104	26	NP_001187103.1	Interferon-induced GTP-binding protein Mx1
2105	26	XP_002121094.1	Uncharacterized protein K02A2.6-like, partial
2106	29	XP_001336197.4	NACHT, LRR and PYD domains-containing protein 12
2107	29	XP_001919012.2	Putative uncharacterized transposon-derived protein F54H12.3-like
2108	29	BAB83841.1	Reo_6
2109	29	XP_003198594.1	RNA-directed DNA polymerase from mobile element jockey-like
2110	30	NP_001189381.1	Chromodomain-helicase-DNA-binding protein 8
2111	30	BAC82609.1	Pol-like protein
2112	31	NP_955925.1	Rac gtpase-activating protein 1
2113	32	AAD19348.1	Reverse transcriptase-like protein
2114	33	ACD38461.1	Immunoglobulin heavy chain variable region
2115	34	XP_004567015.1	Phosphatidylinositol 4-kinase alpha-like isoform X1
2116	36	XP_697781.2	UDP-glucose:glycoprotein glucosyltransferase 2
2117	39	XP_002667924.1	Hypothetical protein LOC100333442
2118	40	XP_002931684.2	Uncharacterized protein F54H12.2-like
2119	41	XP_001334567.1	Extracellular calcium-sensing receptor-like
2120	41	CAP08937.1	Fish virus induced TRIM protein
2121	47	XP_003978902.1	Dynein heavy chain 9, axonemal-like
2122	54	XP_003459773.1	Protein NLRC3-like
2123	58	CAD60788.1	Novel protein similar to DNA polymerases
2124	59	AAW24907.1	SJCHGC06398 protein
2125	85	XP_002942463.2	Vomeronasal type-2 receptor 26-like
2126	89	XP_003460241.1	Hypothetical protein LOC100705656
2127	121	XP_003459734.1	CD2-associated protein-like
2128	124	XP_003199629.1	Hypothetical protein LOC100535052, partial
2129	131	BAE46429.1	ORF2-encoded protein
2130	303	BAD72127.1	Reverse transcriptase

Supplement table 2. The list of catfish species-specific duplicated genes

Group	Copy number	NR id	Gene name
1	2	EFN62203.1	120.7 kda protein in NOF-FB transposable element
2	2	ADF97628.1	14 kda apolipoprotein
3	2	NP_001156671.1	2'-deoxynucleoside 5'-phosphate N-hydrolase 1
4	2	ACO09534.1	ACN9 protein homolog, mitochondrial precursor
5	2	XP_001919948.2	A-kinase anchor protein 6
6	2	NP_001020705.2	Alpha globin regulatory element containing-like
7	2	NP_001166124.1	Amylo-1, 6-glucosidase, 4-alpha-glucanotransferase
8	2	NP_001119867.1	Ankyrin repeat and LEM domain-containing protein 2
9	2	NP_955834.2	Apoptosis inhibitor 5
10	2	AAI62155.1	Bardet-Biedl syndrome 1
11	2	AAH55578.1	Blzf1 protein
12	2	XP_003973271.1	Bystin-like
13	2	F1Q8W0.1	Cactin
14	2	NP_570987.1	Catalase
15	2	NP_001082810.1	CCR4-NOT transcription complex subunit 11
16	2	NP_001187191.1	CD3 zeta precursor
17	2	XP_004556684.1	CDP-diacylglycerol--glycerol-3-phosphate 3-phosphatidyltransferase, mitochondrial-like isoform X1
18	2	NP_001107081.1	Centromere protein H
19	2	XP_002663091.2	Centrosomal protein of 120 kda
20	2	BAA78329.1	Chmadrin (short type)
21	2	NP_001038478.2	Chromatin assembly factor 1 subunit A
22	2	NP_001082851.1	Coiled-coil domain-containing protein 79
23	2	NP_001035414.1	Coiled-coil domain-containing protein 93
24	2	XP_003783998.1	Coproporphyrinogen-III oxidase, mitochondrial
25	2	XP_004573747.1	Cubilin-like
26	2	NP_001018407.1	Cystinosin precursor
27	2	NP_001074047.2	Cytosolic endo-beta-N-acetylglucosaminidase
28	2	ELK32047.1	DDB1- and CUL4-associated factor 15
29	2	NP_001121995.1	DNA polymerase epsilon catalytic subunit A
30	2	NP_001133624.1	DNA polymerase subunit alpha B
31	2	NP_001006081.1	DNA replication complex GINS protein PSF1
32	2	XP_001340400.3	E3 ubiquitin-protein ligase listerin
33	2	XP_693770.3	ER membrane protein complex subunit 1 isoform X2
34	2	NP_956039.2	ESF1 homolog
35	2	NP_001187548.1	Ester hydrolase c11orf54-like protein
36	2	XP_003441279.1	Eukaryotic translation initiation factor 3 subunit A-like
37	2	XP_003442557.1	Eukaryotic translation initiation factor 3 subunit D-like
38	2	AAH44132.1	Exportin 6
39	2	XP_004563576.1	FACT complex subunit SPT16-like

40	2	NP_001070215.1	F-box only protein 36
41	2	BAC82608.1	Gag-like protein
42	2	NP_001243171.1	General transcription factor 3C polypeptide 3
43	2	NP_997981.1	Glycine N-methyltransferase
44	2	XP_004546800.1	Glycine--trna ligase-like
45	2	NP_957116.2	G-protein coupled receptor 143
46	2	NP_001187735.1	Guanylin precursor
47	2	WP_019505193.1	Hypothetical protein
48	2	WP_009782202.1	Hypothetical protein
49	2	PC1123	Hypothetical protein - bloodfluke planorb (fragment)
50	2	EKX44718.1	Hypothetical protein GUITHDRAFT_152949, partial
51	2	XP_002667279.1	Hypothetical protein LOC100330441
52	2	XP_003200965.1	Hypothetical protein LOC100535282
53	2	XP_003199656.1	Hypothetical protein LOC100537292
54	2	XP_003200691.1	Hypothetical protein LOC100537950
55	2	XP_003197739.1	Hypothetical protein LOC100538131
56	2	XP_003455041.1	Hypothetical protein LOC100701310
57	2	XP_003451285.1	Hypothetical protein LOC100712215
58	2	ELR52716.1	Hypothetical protein M91_08923
59	2	EIE88575.1	Hypothetical protein RO3G_13286
60	2	EFA13300.1	Hypothetical protein tcasga2_TC014830
61	2	XP_002959926.1	Hypothetical protein VOLCADRAFT_39527
62	2	CCD18005.1	Hypothetical protein, conserved in T. Vivax
63	2	XP_004567675.1	Ig-like V-type domain-containing protein FAM187A-like isoform X1
64	2	XP_002662615.2	Integral membrane protein GPR180-like, partial
65	2	XP_690622.4	Integrator complex subunit 1 isoform 2
66	2	NP_001187149.1	Interleukin 1 beta type b
67	2	NP_001134853.1	Isoamyl acetate-hydrolyzing esterase 1 homolog
68	2	XP_004550823.1	Kell blood group glycoprotein-like
69	2	XP_004916584.1	Lamina-associated polypeptide 2, isoforms alpha/zeta-like
70	2	XP_001921374.3	Leucine-rich repeat-containing protein 56-like
71	2	NP_001018537.1	Leucine-rich repeat-containing protein c10orf11
72	2	NP_956620.1	Light ear protein
73	2	AAH90185.1	LOC553294 protein, partial
74	2	AAI55325.1	LOC797572 protein
75	2	NP_001128156.1	Low-density lipoprotein receptor-related protein 6 precursor
76	2	NP_001133848.1	Mediator complex subunit 16
77	2	XP_004555069.1	Metalloprotease TIKI2-like
78	2	NP_001188160.1	Methionine adenosyltransferase 2 subunit beta
79	2	AAI35021.1	Microsomal triglyceride transfer protein
80	2	XP_003353524.2	Mis18-binding protein 1-like

81	2	NP_001018198.1	Mitochondria-eating protein
82	2	NP_001188000.1	Mitochondrial 28S ribosomal protein s14
83	2	ABD85505.1	Mitochondrial 28S ribosomal protein S32-like
84	2	ADO28404.1	Mitochondrial cytochrome c1 heme protein
85	2	NP_957206.2	Mitotic spindle assembly checkpoint protein MAD1
86	2	ADO28298.1	N-acylneuraminate cytidylyltransferase
87	2	NP_001003747.1	NADH dehydrogenase [ubiquinone] flavoprotein 1, mitochondrial
88	2	NP_955844.1	N-alpha-acetyltransferase 35, natc auxiliary subunit
89	2	NP_001038272.1	Neuroblastoma-amplified sequence
90	2	XP_003445249.1	NKAP-like protein-like
91	2	NP_001133241.1	Nogo-B receptor
92	2	NP_001025338.1	Nuclear pore complex protein Nup107
93	2	XP_002667606.2	Nuclear pore membrane glycoprotein 210
94	2	NP_956531.1	Nucleolar protein 10
95	2	BAN16579.1	Organic solute transporter alpha-1
96	2	XP_004562971.1	Phenylalanine--trna ligase alpha subunit-like
97	2	AAI54375.1	Pibf1 protein
98	2	XP_001637582.1	Predicted protein
99	2	NP_001159825.2	Probable imidazolonepropionase
100	2	XP_700282.5	Probable jmjc domain-containing histone demethylation protein 2C-like
101	2	XP_003198156.1	Protein CLEC16A isoform X1
102	2	NP_001082796.1	Protein FAM115
103	2	XP_687326.2	Protein KIAA1045-like isoform X2
104	2	NP_001019984.1	Protein MCM10 homolog
105	2	XP_690522.1	Protein QN1 homolog
106	2	XP_688748.3	Protein virilizer homolog
107	2	XP_004574271.1	Rab-like protein 6-like isoform X1
108	2	XP_001923770.1	Radial spoke head protein 4 homolog A isoform 1
109	2	NP_001187596.1	Ribonuclease h2 subunit b
110	2	ADO28027.1	Ring finger and chy zinc finger domain-containing protein 1
111	2	ADO27902.1	Ring finger protein 32
112	2	XP_683754.2	RNA-binding protein NOB1 isoformx1
113	2	NP_001116519.1	S phase cyclin A-associated protein in the endoplasmic reticulum
114	2	NP_775351.1	SCL-interrupting locus protein homolog
115	2	NP_001133813.1	Sequestosome-1
116	2	AAI34974.1	Si:ch211-105d11.3
117	2	AAX30323.2	SJCHGC03012 protein
118	2	ACI70099.1	SLAM family member 8 precursor
119	2	NP_991230.1	Small nuclear ribonucleoprotein polypeptides B and B1
120	2	XP_001920500.3	Small subunit processome component 20 homolog, partial

121	2	NP_001082938.1	Sphingosine-1-phosphate lyase 1
122	2	NP_958871.1	Suppressor of SWI4 1 homolog
123	2	YP_006591514.1	Tash protein pest motif family
124	2	AAH65630.1	Tfg protein
125	2	NP_963876.1	TFIIH basal transcription factor complex helicase XPB subunit
126	2	Q7ZV43.1	Toll-interacting protein
127	2	CAJ00236.1	TPA: endonuclease-reverse transcriptase
128	2	NP_001187529.1	Transaldolase
129	2	NP_997848.1	Transmembrane protein 165 precursor
130	2	XP_001336187.3	Trehalase
131	2	XP_003444536.1	Tripartite motif-containing protein 69-like
132	2	NP_001070119.1	Trna-splicing endonuclease subunit Sen2
133	2	NP_001002451.1	Ubiquitin fusion degradation protein 1 homolog
134	2	XP_004014456.1	Uncharacterized protein c20orf195 homolog
135	2	XP_690269.4	Uncharacterized protein c7orf57 homolog isoformx2
136	2	XP_001923762.1	Uncharacterized protein c8orf45 homolog
137	2	XP_003459696.1	Uncharacterized protein KIAA1797-like
138	2	XP_001333668.2	Uncharacterized protein KIAA1841-like
139	2	NP_001096105.1	Uncharacterized protein LOC100124608
140	2	XP_002170374.2	Uncharacterized protein LOC100205880, partial
141	2	XP_002667014.1	Uncharacterized protein LOC100329955
142	2	XP_002663158.1	Uncharacterized protein LOC100330250 isoform 1
143	2	XP_003977182.1	Uncharacterized protein LOC101061985, partial
144	2	XP_004076963.1	Uncharacterized protein LOC101165752
145	2	XP_004085509.1	Uncharacterized protein LOC101166189
146	2	XP_004211737.1	Uncharacterized protein LOC101240276
147	2	XP_004556396.1	Uncharacterized protein LOC101472750
148	2	XP_004549715.1	Uncharacterized protein LOC101480956
149	2	XP_004571522.1	Uncharacterized protein LOC101486585
150	2	XP_004918303.1	Uncharacterized protein LOC101734130, partial
151	2	XP_004932995.1	Uncharacterized protein LOC101740474
152	2	XP_005022161.1	Uncharacterized protein LOC101792521 isoform X2
153	2	NP_001002704.1	Uncharacterized protein LOC436977
154	2	NP_001116773.2	Uncharacterized protein LOC570267 isoform 1
155	2	XP_001198200.2	Uncharacterized protein LOC762541
156	2	NP_001071257.1	Uncharacterized protein LOC777745
157	2	CAG09357.1	Unnamed protein product
158	2	NP_001014353.1	UPF0364 protein c6orf211 homolog
159	2	NP_001188061.1	Upf0490 protein c1orf201-like protein
160	2	NP_001025261.3	UPF0536 protein c12orf66 homolog
161	2	NP_001139876.1	UPF0740 protein c1orf192 homolog
162	2	NP_001091659.1	Vacuolar protein sorting-associated protein 16 homolog

163	2	XP_684689.2	Vacuolar protein sorting-associated protein 52 homolog isoformx2
164	2	XP_003444364.1	Vitamin K-dependent gamma-carboxylase-like
165	2	XP_003447124.1	WASH complex subunit 7-like
166	2	XP_003200727.1	YLP motif-containing protein 1
167	2	AAI33991.1	Zgc:114114 protein
168	2	EPQ19413.1	Zinc finger protein interacting with ribonucleoprotein K
169	2	XP_004457235.1	Zonadhesin
170	2	XP_003448552.1	ZZ-type zinc finger-containing protein 3
171	3	XP_003453615.1	Breast carcinoma-amplified sequence 3
172	3	XP_002662035.2	Butyrophilin-like protein 2-like
173	3	XP_004078011.1	Cation-dependent mannose-6-phosphate receptor-like isoform 2
174	3	NP_001139071.1	Cell division cycle and apoptosis regulator protein 1
175	3	XP_696859.4	DNA repair protein RAD50 isoform X3
176	3	NP_001103995.1	Ectopic P granules protein 5 homolog
177	3	NP_001187587.1	Fam183a
178	3	NP_001070082.1	GPI transamidase component PIG-S
179	3	XP_003455916.1	Hypothetical protein LOC100694549
180	3	AAN41456.1	Interleukin-8 variant 4
181	3	A7MBP4.2	Intraflagellar transport protein 46 homolog
182	3	XP_004558643.1	Negative elongation factor B-like
183	3	XP_003206612.1	Probable D-tyrosyl-trna(Tyr) deacylase 2-like
184	3	ACI68751.1	T-cell leukemia translocation-altered gene protein homolog
185	3	XP_004554617.1	Tetratricopeptide repeat protein 40-like isoform X3
186	3	XP_003199318.1	Transmembrane protein ENSP00000382582 homolog
187	3	XP_002938353.2	Uncharacterized protein LOC100490320
188	3	XP_004083839.1	Uncharacterized protein LOC101175121
189	3	XP_003199962.1	Uncharacterized protein LOC368429 isoform X1
190	3	NP_001073433.1	Uncharacterized protein LOC556929
191	3	XP_001199932.1	Uncharacterized protein LOC763830, partial
192	3	CCD15206.1	Unnamed protein product, partial
193	3	XP_696898.4	Wu:fb77e12
194	3	AAI53507.1	Zgc:174162 protein
195	4	NP_956526.1	DNA-directed RNA polymerase III subunit RPC3
196	4	XP_689534.4	Elongator complex protein 1
197	4	NP_001092114.1	Fig-Hepta protein
198	4	XP_003449587.1	Hypothetical protein LOC100697617
199	4	XP_003450494.1	Hypothetical protein LOC100698500
200	4	XP_701610.1	Hypothetical protein LOC368413
201	4	EFA12620.1	Hypothetical protein tcasga2_TC010247
202	4	XP_002664688.1	Nuclear fragile X mental retardation-interacting protein 2
203	4	CCH50976.1	T4.15

204	4	NP_001188116.1	Uncharacterized protein c1orf43-like protein
205	4	XP_003723921.1	Uncharacterized protein LOC100891635
206	4	XP_004085839.1	Uncharacterized protein LOC101169091
207	4	XP_004085732.1	Uncharacterized protein LOC101174564
208	4	XP_004916525.1	Uncharacterized protein LOC101733395
209	4	CAG10804.1	Unnamed protein product
210	5	XP_001345924.3	Hypothetical protein LOC100007481
211	5	XP_003443479.1	Hypothetical protein LOC100692970
212	5	XP_693683.4	Interferon-inducible gtpase 5
213	5	XP_001345586.3	Otoancorin-like
214	5	NP_001232875.1	Prostate stem cell antigen precursor-like
215	6	XP_003458758.1	Hypothetical protein LOC100700692
216	6	NP_001004966.1	Protein FAM206A
217	6	XP_002161008.2	Uncharacterized protein LOC100207270
218	6	XP_004084042.1	Uncharacterized protein LOC101164525
219	7	ELU03397.1	Hypothetical protein CAPTEDRAFT_219998
220	7	EFA06499.1	Hypothetical protein tcasga2_TC009398
221	7	XP_004087031.1	Uncharacterized protein LOC101169431
222	7	XP_004554031.1	Uncharacterized protein LOC101484061
223	7	XP_004575946.1	Zinc finger CCHC domain-containing protein 3-like
224	8	AAP06505.1	Hypothetical protein
225	8	XP_002414393.1	Hypothetical protein iscw_ISCW012389
226	10	XP_002662769.1	Hypothetical protein LOC100330014
227	10	XP_004552155.1	Uncharacterized protein LOC101481353
228	11	XP_003458913.1	Hypothetical protein LOC100699620
229	12	XP_004914305.1	Uncharacterized protein LOC101733602
230	13	XP_003197744.1	Hypothetical protein LOC100536106, partial
231	15	YP_073558.1	RNA-dependent DNA polymerase
232	15	XP_002166082.2	Sacsin-like
233	26	EHJ78503.1	Endonuclease-reverse transcriptase
234	29	XP_003198594.1	RNA-directed DNA polymerase from mobile element jockey-like
235	30	BAC82609.1	Pol-like protein
236	47	XP_003978902.1	Dynein heavy chain 9, axonemal-like
237	58	CAD60788.1	Novel protein similar to DNA polymerases
238	89	XP_003460241.1	Hypothetical protein LOC100705656
239	124	XP_003199629.1	Hypothetical protein LOC100535052

Reference

1. Bridges, C.B. **The Bar "gene" a duplication.** *Science* 1936, **83**: 210–211
2. Ohno S: **Evolution by gene duplication**. *New York*: Springer Verlag. 1970
3. Li, W.H. et al. **Evolutionary analyses of the human genome**. *Nature* 2001, **409**, 847–849
4. Cannon SB, Mitra A, Baumgarten A, Young ND, May G: **The roles of segmental and tandem gene duplication in the evolution of large gene families in Arabidopsis thaliana**. *BMC Plant Biol* 2004, **4**:10.
5. Sankoff D: **Gene and genome duplication**. *Curr Opin Genet Dev* 2001, **11**:681-684.
6. Zhang J.: **Evolution by gene duplication: an update.** *Trends Ecol Evol*, 2003.**18**:292-298.
7. Li, W.H: **Molecular Evolution**, Sinauer1997
8. Long M, Betrán E, Thornton K, Wang W: **The origin of new genes: glimpses from the young and old**. *Nat Rev Genet*. 2003, **4**:865-75.
9. Long, M.: **Evolution of novel genes**. *Curr. Opin. Genet. Dev.* 2001, **11**, 673–680
10. Eichler EE, Sankoff D.: **Structural dynamics of eukaryotic chromosome evolution**. *Science* 2003, **301**: 793-797.
11. Pan D, Zhang LQ: **Tandemly arrayed genes in vertebrate genomes.** *Comparative and Functional Genomics*. 2008: 545269
12. Bailey, J.A., Yavor, A.M., Massa, H.F., Trask, B.J., and Eichler, E.E.: **Segmental duplications: Organization and impact within the current human genome project assembly**. *Genome Res*. 2001, **11**:1005-1017.
13. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R, Oseroff VV, Albertson DG, Pinkel D, Eichler EE: **Segmental duplications and copy-number variation in the human genome**. *Am J Hum Genet*, 2005, **77**:78-88
14. Samonte, R.V. and Eichler, E.E.: **Segmental duplications and the evolution of the primate genome**. *Nat. Rev. Genet*. 2002, **3**, 65–72
15. Bailey JA, Church DM, Ventura M, Rocchi M, Eichler EE: **Analysis of segmental duplications and genome assembly in the mouse**. Genome Res. 2004, **14**:789-801.
16. Liu GE, Ventura M, Cellamare A, Chen L, Cheng Z, Zhu B, Li C, Song J, Eichler EE. **Analysis of recent segmental duplications in the bovine genome**. *BMC Genomics*. 2009 , **1**;10:571.
17. Nicholas TJ, Cheng Z, Ventura M, Mealey K, Eichler EE, Akey JM. T**he genomic architecture of segmental duplications and associated copy number variants in dogs.** *Genome Res.* 2009 , **19**:491-9
18. Lynch, M., Conery, J.S.: **The origins of genome complexity**. *Science*. 2003 **302**,1401–1404.

19. Maere, S., et al.: **Modeling gene and genome duplications in eukaryotes.** *Proc. Natl. Acad. Sci. U. S. A.* 2005, **102**, 5454–5459.

20. Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., Postlethwait, J., **Preservation of duplicate genes by complementary, degenerative mutations.** *Genetics* 1999, **151**:1531-1545.

21. Rastogi S, Liberles DA: **Subfunctionalization of duplicated genes as a transition state to neofunctionalization**. *BMC Evol Biol* 2005, **5**:28.

22. Nowak, M.A. et al.: **Evolution of genetic redundancy.** *Nature* 1997, **388**, 167–171

23. Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, Werner J, Villanea FA, Mountain JL, Misra R, Carter NP, Lee C, Stone AC: **Diet and the evolution of human amylase gene copy number variation**. *Nat Genet* 2007, **39**:1256-1260.

24. Fitch,W.M.,. **Distinguishing homologous from analogous protein**. *Syst. Zoo*l. 1970, **19**, 99–113.

25. De Grassi A, Lanave C, Saccone C. **Genome duplication and gene-family evolution: the case of three OXPHOS gene families.** *Gene*. 2008, **421**:1-6.

26. Shoja V, Zhang L. **A roadmap of tandemly arrayed genes in the genomes of human, mouse, and rat**. *Mol Biol Evol*. 2006, **23**:2134-41.

27. Lu J, Peatman E, Tang H, Lewis J, Liu Z. **Profiling of gene duplication patterns of sequenced teleost genomes: evidence for rapid lineage-specific genome expansion mediated by recent tandem duplications. BMC Genomics**. 2012, **13**:246.

28. Ferraresso S, Kuhl H, Milan M, Ritchie DW, Secombes CJ, Reinhardt R, Bargelloni L. **Identification and characterisation of a novel immune-type receptor (NITR) gene cluster in the European sea bass, Dicentrarchus labrax, reveals recurrent gene expansion and diversification by positive selection**. *Immunogenetics*. 2009, **61**:773-88.

29. Song H, Yan YL, Titus T, He X, Postlethwait JH.. **The role of stat1b in zebrafish hematopoiesis.** *Mech Dev*. 2011, **128**:442-56

30. Wiens GD, Glenney GW. **Origin and evolution of TNF and TNF receptor superfamilies**. *Dev Comp Immunol*. 2011.

31. Niu D, Peatman E, Liu H, Lu J, Kucuktas H, Liu S, Sun F, Zhang H, Feng T, Zhou Z, Terhune J, Waldbieser G, Li J, Liu Z. **Microfibrillar-associated protein 4 (MFAP4) genes in catfish play a novel role in innate immune responses.** *Dev Comp Immunol*. 2011.**35**:568-79.

32. Nomiyama H, Hieshima K, Osada N, Kato-Unoki Y, Otsuka-Ono K, Takegawa S, Izawa T, Yoshizawa A, Kikuchi Y, Tanase S, Miura R, Kusuda J, Nakao M, Yoshie O. **Extensive expansion and diversification of the chemokine gene family in zebrafish: identification of a novel chemokine subfamily CX**. *BMC Genomics*. 2008. **15**;9:222.

33. Peatman E, Liu Z. **Evolution of CC chemokines in teleost fish: a case study in gene duplication and implications for immune diversity**. *Immunogenetics*. 2007.**59**:613-23.

34. Odom EW, Vasta GR.. **Characterization of a binary tandem domain F-type lectin from striped bass (Morone saxatilis).** *J Biol Chem*. 2006, **281**:1698-713.

35. Reusch TB, Schaschl H, Wegner KM..**Recent duplication and inter-locus gene conversion in major histocompatibility class II genes in a teleost, the three-spined stickleback.** *Immunogenetics*. 2004,**56**:427-37.

36. Hordvik I, Thevarajan J, Samdal I, Bastani N, Krossøy B. **Molecular cloning and phylogenetic analysis of the Atlantic salmon immunoglobulin D gene**. *Scand J Immunol*. 1999.**50**:202-10.

37. Venkatesh B, Brenner S. **Structure and organization of the isotocin and vasotocin genes from teleosts**. *Adv Exp Med Biol*. 1995, **395**:629-38.

38. Novak AE, Jost MC, Lu Y, Taylor AD, Zakon HH, Ribera AB. **Gene duplications and evolution of vertebrate voltage-gated sodium channel**s. *J Mol Evol*. 2006. **63**:208-21.

39. Conlon JM, Larhammar D. **The evolution of neuroendocrine peptides**. *Gen Comp Endocrinol*. 2005. **142**:53-9.

40. Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W., and Eichler, E.E. **Recent segmental duplications in the human genome**. *Science*. 2002, **297:** 1003-1007

41. Sharp, A.J. et al.: **Segmental duplications and copy number variation in the human genome.** *Am J Hum Gene*t 2005, **77**, 78–88

42. Iafrate, A.J. et al. **Detection of large-scale variation in the human genome.** *Nat Genet* 2004, **36**, 949–951

43. Tuzun, E. et al. **Fine-scale structural variation of the human genome**. *Nat Genet* 2005, **37**, 727–732

44. Marques-Bonet T, Girirajan S, Eichler EE. **The origins and impact of primate segmental duplications**. *Trends Genet*. 2009, **25**:443-54.

45. Estivill, X. et al. **Chromosomal regions containing high-density and ambiguously mapped putative single nucleotide polymorphisms (SNPs) correlate with segmental duplications in the human genome.** *Hum. Mol. Genet*. 2002, **11**: 1987–1995.

46. Fredman D, White SJ, Potter S, Eichler EE, Den Dunnen JT, Brookes AJ: **Complex SNP-related sequence variation in segmental genome duplications**. *Nat Genet*. 2004

47. Cheung, J. et al. **Genome-wide detection of segmental duplications and potential assembly errors in the human genome sequence**. *Genome Biol*. 2003, **4**, R25

48. Liu S, Zhang Y, Zhou Z, Waldbieser G, Sun F, Lu J, Zhang J, Jiang Y, Zhang H, Wang X, Rajendran K, Khoo L, Kucuktas H, Peatman E, Liu Z: **Efficient assembly and annotation of the transcriptome of catfish by RNA-Seq analysis of a doubled haploid homozygote**. *BMC Genomics*. 2012, **13**:595.

49. Gidskehaug L, Kent M, Hayes BJ, Lien S: **Genotype calling and mapping of multisite variants using an Atlantic salmon iSelect SNP array**. *Bioinformatics*. 2011, **27**:303-10.

50. Cheng, Z. et al. **A genome-wide comparison of recent chimpanzee and human segmental duplications**. *Nature* 2005, **437**: 88–93

51. Bailey JA, Eichler EE. **Primate segmental duplications: crucibles of evolution, diversity and disease**. *Nat Rev Genet*. 2006, **7**:552-64.

52. Koszul, R. et al. (**Eukaryotic genome evolution through the spontaneous duplication of large chromosomal segments**. *EMBO J.* 2004, **23**, 234–243

53. Linardopoulou, E.V. et al. **Human subtelomeres are hotspots of interchromosomal recombination and segmental duplication**. *Nature* 2005, **437**, 94–100

54. Bailey, J.A. et al. **An Alu transposition model for the origin and expansion of human segmental duplications**. *Am. J. Hum. Gene*t. 2003, **73**, 823–834

55. Horvath, J.E. et al. **Punctuated duplication seeding events during the evolution of human chromosome 2p11.** *Genome Res*. 2005, **15**, 914–927

56. She X, Cheng Z, Zöllner S, Church DM, Eichler EE. **Mouse segmental duplication and copy number variation.** *Nat Genet.* 2008, **40**:909-14.

57. Liu GE, Bickhart DM. **Copy number variation in the cattle genome**. *Funct Integr Genomics*. 2012,

58. Mills RE, Walter K, Stewart C, Handsaker RE, Chen K, Alkan C, Abyzov A, Yoon SC, Ye K, Cheetham RK, Chinwalla A, Conrad DF, Fu Y, Grubert F, Hajirasouliha I, Hormozdiari F, Iakoucheva LM, Iqbal Z, Kang S, Kidd JM, Konkel MK, Korn J, Khurana E, Kural D, Lam HYK, Leng J, Li R, Li Y, Lin CY, Luo R, Mu XJ, Nemesh J, Peckham HE, Rausch T, Scally A, Shi X, Stromberg MP et al.: **Mapping copy number variation by population-scale genome sequencing**. *Nature* 2011, **470**:59–65

59. Stankiewicz P, Lupski JR: S**tructural variation in the human genome and its role in disease**. *Annu Rev Med*, 2010 **61**:437–455

60. Fadista J, Thomsen B, Holm LE, Bendixen C: **Copy number variation in the bovine genome**. *BMC Genomics*. 2010, **11**:284

61. Bae JS, Cheong HS, Kim LH, NamGung S, Park TJ, Chun JY, Kim JY, Pasaje CF, Lee JS, Shin HD: **Identification of copy number variations and common deletion polymorphisms in cattle.** *BMC Genomics*. 2010, **11**:232

62. Liu GE, Hou Y, Zhu B, Cardone MF, Jiang L, Cellamare A, Mitra A, Alexander LJ, Coutinho LL, Dell'Aquila ME, Gasbarre LC, Lacalandra G, Li RW, Matukumalli LK, Nonneman D, Regitano LCD, Smith TPL, Song J, Sonstegard TS, Van Tassell CP, Ventura M, Eichler EE, McDaneld TG, Keele JW: **Analysis of copy number variations among diverse cattle breeds.** *Genome Res.*2010, **20**:693–703

63. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, Yamrom B, Yoon S, Krasnitz A, Kendall J, Leotta A, Pai D, Zhang R, Lee YH, Hicks J, Spence SJ, Lee AT, Puura K, Lehtimaki T, Ledbetter D, Gregersen PK, Bregman J, Sutcliffe JS, Jobanputra V, Chung W, Warburton D, King MC, Skuse D, Geschwind DH, Gilliam TC, Ye K, Wigler M: **Strong association of de novo copy number mutations with autism**. *Science* 2007, **316**:445–449

64. Tam GW, Redon R, Carter NP, Grant SG. T**he role of DNA copy number variation in schizophrenia**. *Biol Psychiatry*. 2009, **66**:1005-12.

65. Brown KH, Dobrinski KP, Lee AS, Gokcumen O, Mills RE, Shi X, Chong WW, Chen JY, Yoo P, David S, Peterson SM, Raj T, Choy KW, Stranger BE, Williamson RE, Zon LI, Freeman JL, Lee C. **Extensive genetic diversity and substructuring among zebrafish**

strains revealed through copy number variant analysis. *Proc Natl Acad Sci U S A*. 2012, **109**:529-34.

66. Pinto D, Darvishi K, Shi XH, Rajan D, Rigler D, Fitzgerald T, Lionel AC, Thiruvahindrapuram B, MacDonald JR, Mills R, Prasad A et al. **Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants**. *Nat Biotechnol* 2011, **29**:512–520

67. Sudmant PH, Kitzman JO, Antonacci F, Alkan C, Malig M, Tsalenko A, Sampas N, Bruhn L, Shendure J, Eichler EE: **Diversity of human copy number variation and multicopy genes**. *Science* 2010, **330**:641–646

68. Kidd JM, Cooper GM, Donahue WF, Hayden HS, Sampas N, Graves T, Hansen N, Teague B, Alkan C, Antonacci F, Haugen E, Zerr T, Yamada NA, Tsang P, Newman TL, Tuzun E, Cheng Z, Ebling HM, Tusneem N, David R, Gillett W, Phelps KA, Weaver M et al.: **Mapping and sequencing of structural variation from eight human genomes**. *Nature*. 2008, **453**:56–64

69. Kidd JM, Sampas N, Antonacci F, Graves T, Fulton R, Hayden HS, Alkan C, Malig M, Ventura M, Giannuzzi G, Kallicki J, Anderson P, Tsalenko A, Yamada NA, Tsang P, Kaul R, Wilson RK, Bruhn L, Eichler EE: **Characterization of missing human genome sequences and copy-number polymorphic insertions.** *Nat Methods*. 2010, **7**:365–371

70. Alkan C, Kidd JM, Marques-Bonet T, Aksay G, Antonacci F, Hormozdiari F, Kitzman JO, Baker C, Malig M, Mutlu O, Sahinalp SC, Gibbs RA, Eichler EE: **Personalized copy number and segmental duplication maps using nextgeneration sequencing**. *Nat Genet.* 2009, **41**:1061–1067

71. Hach F, Hormozdiari F, Alkan C, Hormozdiari F, Birol I, Eichler EE, Sahinalp SC. **mrsFAST: a cache-oblivious algorithm for short-read mapping.** *Nat methods* 2010, **7**:576–577

72. Karakoc E, Alkan C, O'Roak BJ, Dennis MY, Vives L, Mark K, Rieder MJ, Nickerson DA, Eichler EE: **Detection of structural variants and indels within exome data.** *Nat Meth* 2011, **9**:176–178

73. Hajirasouliha I, Hormozdiari F, Alkan C, Kidd JM, Birol I, Eichler EE, Sahinalp SC: **Detection and characterization of novel sequence insertions using paired-end next-generation sequencing.** *Bioinformatics*. 2010, **26**:1277–1283

74. Agam A, Yalcin B, Bhomra A, Cubin M, Webber C, Holmes C, Flint J, Mott R: **Elusive copy number variation in the mouse genome.** *PLoS One*. 2010, **5**:9

75. Bickhart DM, Hou Y, Schroeder SG, Alkan C, Cardone MF, Matukumalli LK, Song J, Schnabel RD, Ventura M, Taylor JF, Garcia JF, Van Tassell CP, Sonstegard TS, Eichler EE, Liu GE. **Copy number variation of individual cattle genomes using next-generation sequencing.** *Genome Res*. 2012, **22**:778-90.

76. McLysaght A, Hokamp K, Wolfe KH: **Extensive genomic duplication during early chordate evolution**. *Nature Genetics*. 2002, **31**: 200-204.

77. Dehal P, Boore JL.: **Two rounds of whole genome duplication in the ancestral vertebrate**. *Plos Biology* 2005, **3**: 1700-1708.

78. Christoffels A, Koh EG, Brenner S, Aparicio S, Venkatesh B.. **Fugu genome analysis provides evidence for a whole-genome duplication early during the evolution of ray-finned fishes.** *Mol Biol Evol.* 2004, **21**:1146– 1151.

79. Jaillon O, Aury JM, Brunet F, Petit JL, Stange-Thomann N, et al. **Genome duplication in the teleost fish Tetraodon nigroviridis reveals the early vertebrate proto-karyotype.** *Nature* 2004, **431**: 946-957.

80. Meyer A, Van de Peer Y. **From 2R to 3R: evidence for a fish-specific genome duplication (FSGD).** *Bioessays* 2005, **27**: 937-945.

81. Ravi V, Venkatesh B.. **Rapidly evolving fish genomes and teleost diversity.** *Curr Opin Genet Dev.* 2008, **18**: 544-550.

82. Prohaska SJ, Stadler PF. **The duplication of the Hox gene clusters in teleost fishes.** *Theory in BioSciences* 2004, ***23***:89–110.

83. Crow KD, Stadler PF, Lynch VJ, Amemiya C, Wagner GP. **The "fish-specific" Hox cluster duplication is coincident with the origin of teleosts.** *Mol Biol Evol.* 2006, **23**:121-36.

84. McGinnis, W., Krumlauf, R: **Homeobox genes and axial patterning.** *Cell* 1992, **68**, 283–302.

85. Lewis EB: **The 1991 Albert Lasker Medical Awards. Clusters of master control genes regulate the development of higher organisms**. *JAMA* 1992, **267**:1524–1531

86. Seo HC, Edvardsen RB, Maeland AD, Bjordal M, Jensen MF, Hansen A, Flaat M, Weissenbach J, Lehrach H, Wincker P, Reinhardt R, Chourrout D: **Hox cluster disintegration with persistent anteroposterior order of expression in Oikopleura dioica.** *Nature* 2004, **431**:67–71

87. Ferrier, D.E., Minguillon, C., Holland, P.W.H., Garcia-Fernandez, J: **The amphioxus Hox cluster: deutrostome posterior flexibility and Hox14**. *Evol. Dev.* 2000, **2**, 284–293.

88. Ruddle, F.H., et al.: **Evolution of chordate Hox gene clusters**. *Ann. N.Y. Acad. Sci.* 1999, **870**, 238–248.

89. Kurosawa G, Takamatsu N, Takahashi M, Sumitomo M, Sanaka E, Yamada K, Nishii K, Matsuda M, Asakawa S, Ishiguro H, Miura K, Kurosawa Y, Shimizu N, Kohara Y, Hori H. **Organization and structure of hox gene loci in medaka genome and comparison with those of pufferfish and zebrafish genomes.** *Gene.* 2006, **370**:75-82.

90. Amores A, Suzuki T, Yan YL, Pomeroy J, Singer A, Amemiya C, Postlethwait JH: **Developmental roles of pufferfish Hox clusters and genome evolution in ray-fin fish**. *Genome Res.* 2004, **14**:1–10

91. Moghadam HK, Ferguson MM, Danzmann RG. **Evolution of Hox clusters in Salmonidae: a comparative analysis between Atlantic salmon (Salmo salar) and rainbow trout (Oncorhynchus mykiss).** *J Mol Evol.* 2005, **61**:636-49.

92. Crow KD, Smith CD, Cheng JF, Wagner GP, Amemiya CT. **An independent genome duplication inferred from Hox paralogs in the American paddlefish—a representative basal ray-finned fish and important comparative reference**. *Genome Biol Evol.* 2012, **4**:937-53.

93. Henkel CV, Burgerhout E, de Wijze DL, Dirks RP, Minegishi Y, Jansen HJ, Spaink HP, Dufour S, Weltzien FA, Tsukamoto K, van den Thillart GE. **Primitive duplicate Hox clusters in the European eel's genome.** *PLoS One.*2012, **7**:e32231.

94. Panopoulou, G. et al. **New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes.** *Genome Res.* 2003, **13**:1056–1066

95. Durand D, Hoberman R: **Diagnosing duplications--can it be done?** *Trends Genet.* 2006, **22**:156-64.

96. Van de Peer Y. **Computational approaches to unveiling ancient genome duplications**. *Nat Rev Genet.* 2004, **5**:752-63

97. Samonte RV, Eichler EE: **Segmental duplications and the evolution of the primate genome.** *Nat Rev Genet.* 2002, **3**:65-72.

98. Sanzol J. **Dating and functional characterization of duplicated genes in the apple (Malus domestica Borkh.) by analyzing EST data**. *BMC Plant Biol.* 2010 **14**;10:87.

99. Panopoulou, G., Poustka, A.J: **Timing and mechanism of ancient vertebrate genome duplications**. **The adventure of a hypothesis.** *Trends Genet.* 2005, **21**: 559–567.

100. Marques-Bonet, T. et al.: **A burst of segmental duplications in the genome of the African great ape ancestor.** *Nature.* 2009, **457**: 877–881

101. Copley, R.R., Goodstadt, L., and Ponting, C. **Eukaryotic domain evolution inferred from genome comparisons.** *Curr. Opin. Genet.* 2003, **13:** 623-628

102. Waldbieser, G. C., Bosworth, B. G., & Quiniou, S. M. **Production of viable homozygous, doubled haploid channel catfish (Ictalurus punctatus).***Marine Biotechnology*. 2010, **12**: 380-385.

103. Fu L, Niu B, Zhu Z, Wu S, Li W**: CD-HIT: accelerated for clustering the next-generation sequencing data**. *Bioinformatics* 2012, **28**:3150-3152.

104. Larkin M, Blackshields G, Brown N, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R: **Clustal W and Clustal X version 2.0.** *Bioinformatics* 2007, **23:**2947-2948.

105. Zhang Z, Xiao J, Wu J, Zhang H, Liu G, Wang X, Dai L: **ParaAT: a parallel tool for constructing multiple protein-coding DNA alignments.** *Biochemical and Biophysical Research Communications* 2012, **419:**779-781.

106. Zhang Z, Li J, Zhao X-Q, Wang J, Wong GK-S, Yu J: **KaKs_Calculator: calculating Ka and Ks through model selection and model averaging.** *Genomics, Proteomics & Bioinformatics* 2006, **4:**259-263.

107. Posada D: **Using Modeltest and PAUP* to select a model of nucleotide substitution.** *Current protocols in bioinformatics* 2003**:**6.5. 1-6.5. 14.

108. Bauer S, Grossmann S, Vingron M, Robinson PN: **Ontologizer 2.0—a multifunctional tool for GO term enrichment analysis and data exploration**. *Bioinformatics* 2008, **24**:1650-1651.

109. Grossmann S, Bauer S, Robinson PN, Vingron M: **Improved detection of overrepresentation of Gene-Ontology annotations with parent–child analysis.** *Bioinformatics* 2007, **23**:3024-3031.

110. Ninwichian P, Peatman E, Liu H, et al**. Second-generation genetic linkage map of catfish and its integration with the BAC-based physical map**. *G3: Genes Genomes Genetics*, 2012, **2**: 1233-1241.

111. Kondrashov F A, Rogozin I B, Wolf Y I, et al. **Selection in the evolution of gene duplications**. *Genome Biol*, 2002, **3**: 8.1-8.9.

112. Chen L, DeVries A L, Cheng C H C**. Evolution of antifreeze glycoprotein gene from a trypsinogen gene in Antarctic notothenioid fish**. *Proceedings of the National Academy of Science*s, 1997, **94**: 3811-3816.

113. Riehle M M, Bennett A F, Long A D. **Genetic architecture of thermal adaptation in Escherichia coli**. *Proceedings of the National Academy of Sciences*, 2001, **98**: 525-530.

114. Sonti R V, Roth J R. **Role of gene duplications in the adaptation of Salmonella typhimurium to growth on limiting carbon sources**. *Genetics*, 1989, **123**: 19-28.

115. Brown C J, Todd K M, Rosenzweig R F**. Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment**. *Molecular Biology and Evolution*, 1998, **15**: 931-942.

116. Maroni G, Wise J, Young J E, et al. **Metallothionein gene duplications and metal tolerance in natural populations of Drosophila melanogaster**. *Genetics*, 1987, **117**: 739-744.

117. van Hoof N A L M, Hassinen V H, Hakvoort H W J, et al. **Enhanced copper tolerance in Silene vulgaris (Moench) Garcke populations from copper mines is associated with increased transcript levels of a 2b-type metallothionein gene**. Plant Physiology, 2001, **126**: 1519-1526.

118. Romero D, Palacios R**. Gene amplification and genomic plasticity in prokaryotes**. *Annual review of genetics*, 1997, **31**: 91-111.

119. Guillemaud T, Raymond M, Tsagkarakou A, et al. **Quantitative variation and selection of esterase gene amplification in Culex pipiens**. Heredity, 1999, **83**: 87-99.

120. Wang Y, Wang Q, Baoprasertkul P, et al. **Genomic organization, gene duplication, and expression analysis of interleukin-1β in channel catfish (*Ictalurus punctatus*).** *Molecular immunology*, 2006, **43**: 1653-1664.

121. Peatman E, Li C, Peterson BC, Straus DL, Farmer BD, Beck BH. **Basal polarization of the mucosal compartment in Flavobacterium columnare susceptible and resistant channel catfish (*Ictalurus punctatus*)**. *Mol Immunol*. 2013, **56**:317-327

122. Li C, Zhang Y, Wang R, Lu J, Nandi S, Mohanty S, Terhune J, Liu ZJ, Peatman E: **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 & Shellfish Immunology* 2012, **32**:816-827.

123. Hirokawa, Nobutaka, et al. **Kinesin superfamily motor proteins and intracellular transport**. *Nature reviews Molecular cell biology* 2009, **10**: 682-696.

124. Ikushima, Hideto, et al. **Internalization of CD26 by mannose 6-phosphate/insulin-like growth factor II receptor contributes to T cell activation**. *Proceedings of the National Academy of Sciences* 2000, **97**: 8439-8444.

125. Muppirala, Madhavi, Vijay Gupta, and Ghanshyam Swarup. **Syntaxin 17 cycles between the ER and ERGIC and is required to maintain the architecture of ERGIC and Golgi**. *Biology of the Cell* 2011, **103**:333-350.

126. Short, Benjamin, et al. "**A GRASP55-rab2 effector complex linking Golgi structure to membrane traffic**." *The Journal of cell biology* 2001, **155**: 877-884.

127. Bird, S., Zou, J., Wang, T., Munday, B., Cunningham, C., Secombes, C.J. **Evolution of interleukin-1 beta**. *Cytokine Growth Factor Rev*. 2002, **13**, 483–502.

128. Chen, Liqiao, et al. **Analysis of a catfish gene resembling interleukin-8: cDNA cloning, gene structure, and expression after infection with *Edwardsiella ictaluri***. *Developmental & Comparative Immunology* 2005, **29**: 135-142.

129. Yoshimoto, Makoto, et al. **Stabilization of quaternary structure and activity of bovine liver catalase through encapsulation in liposomes**. *Enzyme and Microbial Technology* 2007, **41**: 849-858.

130. Klotz, Martin G., and Peter C. Loewen. **The molecular evolution of catalatic hydroperoxidases: evidence for multiple lateral transfer of genes between prokaryota and from bacteria into eukaryota**. *Molecular biology and evolution,* 2003,**20**: 1098-1112.

131. Augoustides-Savvopoulou, P., et al. **Glycine N-methyltransferase deficiency: a new patient with a novel mutation**. *Journal of inherited metabolic disease,* 2003, **26**:745-759.

132. Lee, Na-Rae, et al. **Negative regulation of RIG-I-mediated antiviral signaling by TRK-fused gene (TFG) protein**. *Biochemical and biophysical research communications* 2013.

133. Kim, Sunyoung, and David S. Gross. **Mediator Recruitment to Heat Shock Genes Requires Dual Hsf1 Activation Domains and Mediator Tail Subunits Med15 and Med16**. *Journal of Biological Chemistry* 2013, **288**: 12197-12213.