Terrestrial adaptation: Expansion of myoglobin and up-regulation of hemoglobin genes in the walking catfish (*Clarias batrachus*)

by

Ning Li

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
August 5, 2017

Keywords: *Clarias batrachus*, genome, adaptation, evolution, duplication, air-breathing organ

Approved by

Zhanjiang Liu, Chair, Professor of School of Fisheries, Aquaculture, and Aquatic Sciences
Rex Dunham, Professor of School of Fisheries, Aquaculture, and Aquatic Sciences
Charles Y. Chen, Professor of Crop, Soil and Environmental Sciences
Aaron M. Rashotte, Associate Professor of Biological Sciences

Abstract

Walking catfish (*Clarias batrachus*), a freshwater fish capable of locomotion on land, serves as a great model for understanding adaptations to terrestrial life. Through comparative genome analysis, here I report that no genes were found within the walking catfish genome specifically related to the air-breathing capability. Instead, adaptive evolution was found in the *C. batrachus* genome to be involved in gene expression and nitrogenous waste metabolic process. Furthermore, myoglobin, olfactory receptor related to class A G protein-coupled receptor 1, and sulfotransferase 6b1 gene families were found specifically expanded in the air-breathing walking catfish genome with 15, 12, and 15 copies, respectively, while non-air-breathing fishes possess only 1-2 copies of these genes. Additionally, comparative transcriptome analysis between the air-breathing organ and the gill demonstrates the mechanism of aerial respiration involved in elastic fiber formation, oxygen binding and transport, and angiogenesis. Especially hemoglobin genes were expressed dramatically higher in the air-breathing organ than in the gill of *C. batrachus*. It is apparent that the coupling of enhanced abilities in oxygen transport and oxygen storage through genomic expansion of myoglobin and transcriptomic up-regulation of hemoglobin is an important component of the molecular basis for the adaptation of aquatic species to terrestrial life.

Acknowledgements

I would like to express my sincere gratitude to my major advisor. Dr. Zhanjiang Liu, for his constant guidance and support throughout my Ph.D. study. I am also grateful to my committee members: Dr. Rex Dunham, Dr. Charles Y. Chen and Dr. Aaron M. Rashotte and my university reader Dr. Joseph Kloepper for their assistance and time on my dissertation work. I would also like to thank Ludmilla Kaltenboeck for her technical assistance on my experiments. Many thanks also go to Dr. Shikai Liu, Dr. Jun Yao, Dr. Yun Li, Dr. Chen Jiang, Dr. Lisui Bao, Dr. Xiaoyan Xu, Dr. Yang Liu, Dr. Chenglong Wu, Dr. Sen Gao, Dr. Qifan Zeng, Dr. Xiaozhu Wang, Tao Zhou, Qiang Fu, Changxu Tian, Wendy Prabowo, Yulin Jin, Yujia Yang, Zihao Yuan, Suxu Tan, Huitong Shi, Wenwen Wang and all the other colleagues in the laboratory for their help, collaboration and friendship. I would like to thank Chinese Scholarship Council for the financial support during four years. Finally, I am grateful to my parents and my girlfriend Xinmi Zhang for their love and support.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

BUSCO	Benchmarking Universal Single-Copy Orthologs
CEG	Core Eukaryotic Gene
DEG	Differentially Expressed Gene
FPKM	Fragments Per Kilobase per Million mapped fragments
GO	Gene Ontology
LINE	Long Interspersed Elements
LRT	Likelihood Ratio Test
LTR	Long Terminal Repeats
MOE	Main Olfactory Epithelium
NR	Non-Redundant
OUC	Ornithine Urea Cycle
qRT-PCR	Quantitative Real-Time PCR
ROS	Reactive Oxygen Species
RPKM	Reads Per Kilobase per Million mapped reads
rRNA	Ribosomal RNA
SINE	Short Interspersed Elements
VNO	Vomeronasal Organ

LITERATURE REVIEW

Air-breathing fishes and their air-breathing organs

Air breathing function first evolved in fishes, retained across their evolutionary history and also played an important role in their evolutions (Graham 1997; Zvaritch, et al. 2007). It is believed that air-breathing evolved to cope with aquatic hypoxia for freshwater species and to adapt to periodic emersion for marine fish (Graham 1997; Graham and Lee 2004; Randall 1981). All of those air breathing fishes are actually bimodal breathers with gills respiring aquatically and also with air-breathing organs breathing aerially (Graham 2011; Randall 1981). The latter situation usually occurs when the water environments are hypoxic. This ability to obtain atmospheric oxygen enables fishes to develop their special adaptations to living in hypoxic environments. At present, about 400 fish species in 50 families of bony fishes (Osteichthyes) were reported to be air breathers, however, the number was thought to be an underestimate of the true total due to some poorly described families such as the Neotropical Loricariids (Graham 2011; Lefevre, et al. 2014a).

Based on the behavior and ecology, air breathing fishes are categorized into amphibious and aquatic air breathing fishes (Graham 1997). Amphibious air breathers could be divided into three grades based on activeness on land: species that volitionally emerge from water (e.g., mudskippers *Scartelaos bistophorus* and *Boleophthalmus pectinirostris*); species that endure brief exposure on land due to receding tide (e.g., the sculpins *Clinocottus globiceps* and *Oligocottus snyderi*); and species that endure seasonal or longer exposure on land (e.g., some clariid catfish *Clarias* and some snakeheads *Channa*) (Graham 2011; Graham 1997). On the other hand, aquatic air breathers include facultative type and continuous type. Facultative air breathers (i.e., swamp eel *Synbanchus marmoratus*) normally respire aquatically unless forced to

breathe atmospheric air by the unfavorable water conditions (hypoxia) or increased oxygen requirements, while continuous air breathers regularly consume atmospheric oxygen even in normoxic environments (Graham 2011; Graham 1997). Based on necessity of respiring aerial oxygen for survival, continuous air breathers could be divided into obligatory and non-obligatory types. Obligatory air breathers (i.e., the silurid *Pangasius sutchi*) need aerial oxygen to keep alive even in oxygen-rich water, on the contrary, non-obligatory air breathers do not require air-breathing in well-oxygenated water to keep alive (Graham 2011; Graham 1997).

Different types of air breathers have developed different kinds of structural adaptations for air-breathing. Commonly, air-breathing fishes have an air chamber modified from an already existing structure, which is also supplied with a large number of blood vessels and fine capillaries (Johansen 1970). These contain some modified spaces including suprabranchial chamber, pharyngeal sacs, pneumatic duct, lungs and respiratory swim bladders, as well as some modified epithelial membranes which are situated in buccal chambers, pharyngeal chambers, opercular chambers, stomach and intestine (Graham 1997). Furthermore, the interface between air and blood is always specifically modified to expand the gas exchange surface, such as foldings, papillations and arborizations of the diffusion exchange membrane (Johansen 1970). These consist of some modified structures including labyrinth, dendrites and gill fans (Graham 1997). Finally, the skin functions as an auxiliary air-respiring organ in many species (Graham 1997).

The adaptive mechanisms corresponding to the bimodal respiration lifestyle

The accessory air-breathing organs not only offer fishes extra oxygen to cope with hypoxia, but also allow them to develop corresponding mechanisms to adapt to the aerial

respiration periods. First for the air-breathing organs, they always have air chambers to provide good ventilation during air-breathing periods and they are usually highly vascularized to facilitate gas exchange between the blood and the atmospheric air (Jiang, et al. 2016; Johansen 1970; Luo, et al. 2016; Martin 2014). Due to the existence of the air-breathing organ and the large amount of oxygen obtained from the air-breathing activities, the gill surface size was reduced to decrease the oxygen loss to the hypoxic water for aquatic air-breathers (Farrell 2011; Graham 1997), however the gill still functions as an respiratory organ in the water. With these structural modifications, air-breathing fishes could survive well in the hypoxic environments. For some air-breathing species, the hemoglobin affinity for oxygen is increased and the metabolic rate is decreased to cope with hypoxia as well (Damsgaard, et al. 2014; Martin 2014; Turko, et al. 2014).

It is noteworthy that the reduced gills not only serve to exchange gas in the water, but also function to regulate ion homeostasis and blood acid-base balance (Cameron 1989; Goss, et al. 1998). However, the regulation of ion and acid-base arises as a challenge for air breathers due to the respiratory media differences and the reduced interactions between gill and water (Shartau and Brauner 2014). In water-breathing fishes, acid-base regulation is coupled to carbon dioxide (CO_2) excretion through the reversible hydration/dehydration reactions of CO_2 and the acid-base equivalents H^+ and HCO_3^- by carbonic anhydrase (Gilmour and Perry 2009). In most fish species, a disturbance of acid-base is usually caused by severe hypoxia, exhaustive exercise or increased environmental CO_2 concentration (Shartau and Brauner 2014). The majority experience reductions in extracellular pH and intracellular pH, whereas the intracellular pH decreases less than extracellular pH owing to the buffering ability of the tissues (Heisler 1984; Shartau and Brauner 2014). Recent studies in five bimodal fish species have demonstrated that air-breathing

fishes regulate disturbed acid-base in a different way where intracellular pH is tightly regulated in spite of large reduction of extracellular pH, termed preferential intracellular pH regulation (Brauner and Baker 2009; Brauner, et al. 2004b; Heisler 1982; Shartau and Brauner 2014).

Acid-base balance is always linked to ion regulation because acid-base compensation depends on the transfer of H^+ and HCO_3^- in exchange for Na^+ and Cl^- across the gill, respectively (Gilmour and Perry 2009). Due to the reduced gill surface and the limited interactions between gill and water during air-breathing intervals, the gill is not sufficient in modulating ions for bimodal breathers, and therefore it has been assumed that the gut and kidney may play critical roles (Brauner, et al. 2004a; Graham 1997; Hochachka, et al. 1978; Johansen 1970; Shartau and Brauner 2014). Additionally, bimodal breathers were reported to possess more mitochondrial-rich cells, which are believed to be responsible for ionoregulation, on the filament and lamellae while aquatic breathers possess fewer mitochondrial-rich cells only on the filament (Evans, et al. 2005; Lin and Sung 2003; Perry 1997). Both strategies are well-known to maintain ion homeostasis, however, ion regulation is not always a challenge. For some air-breathers which may inhabit terrestrial ion-free environments for a long period such as African lungfish (Perry, et al. 2008), ion conservation becomes more important (Shartau and Brauner 2014). In this case, kidney takes the place of gill to play a critical role in avoiding ion loss (Perry, et al. 2008).

In addition to gas exchange, ion and acid-base regulation, ammonia excretion is another function of the gill. Ammonia is generated from amino acid catabolism. As the main nitrogenous waste and a toxic chemical, its excretion or elimination becomes another challenge for air-breathing fishes due to the reduced gill surface size (Chew and Ip 2014; Graham 1997). However, they develop various strategies to cope with this ammonia problem. Generally, these strategies could be summarized into four categories: (1) reduction of exogenous ammonia entry

and active ammonia excretion; (2) conversion of ammonia to less toxic chemicals such as urea and amino acids; (3) reduction of amino acid catabolism for decreased production of ammonia; and (4) increase of tolerance to ammonia (Chew and Ip 2014). When dealing with hyper-ammonia stress, air-breathers may apply one or more strategies, which depends greatly on their behavior and the nature of their environments (Chew and Ip 2014).

Although different mechanisms are developed for bimodal breathers to adapt to stressful environments, emergence and terrestrial locomotion may be triggered by a series of environmental degradation and biotic factors such as dried habitat and reproductive process (Pace and Gibb 2014; Sayer and Davenport 1991). For instance, some intertidal fishes emerge during the low tide (Martin 2014), and Mangrove rivulus fish (*Kryptolebias marmoratus*) incubates embryos on land (Wells, et al. 2015). It was reported that 26 air-breathing fish genera are able to get out of the water and move around (Graham 1997). Undoubtedly, their musculoskeletal systems got evolved to be suitable not only for swimming in the water but also for moving about on land, and it was demonstrated that the pectoral fins and the axial body play key roles during terrestrial locomotion (Pace and Gibb 2014). In general, three modes of movements were reported: axial-based, appendage-based and axial-appendage-based (Pace and Gibb 2014). Axial-based locomotion was described in air-breathing fishes with an elongate body, such as the ropefish *Erpetoichthys calabaricus* (Pace and Gibb 2011); appendage-based locomotion was well studied in mudskippers, displaying a slow halting motion with the help of pectoral fins (HARRIS 1960); and axial-appendage-based locomotion was found to be more effective with the combination of pectoral fins and tail fin, such as the snake-like movements in walking catfishes of the Clariidae (Johnels 1957).

Walking catfish (*Clarias batrachus*) and its air breathing organ

The walking catfish is a freshwater air-breathing teleost species native to Southeast Asia where it is widely distributed as an aquaculture species due to its high economic value as food (Srivastava, et al. 2016). Unfortunately, this species has been categorized as endangered species because of over-exploitation and habitat alterations in India and Bangladesh (Dahanukar, et al. 2004; Islam, et al. 2007; Khedkar, et al. 2010). On the contrary, it is an invasive species in the United States, expanding to over 10 states on the eastern and western coasts (http://maps.iucnredlist.org/map.html?id=166613) (J. 2011). It was imported into Florida, reportedly from Thailand, in the early 1960s for the aquarium trade (Courtenay Jr, et al. 1986). It has been thought to be damaging to native fish populations, however, there is little evidence to support this except they do invade aquaculture facilities and can cause severe damage to cultured fish populations (Florida Museum, University of Florida 2017; https://www.floridamuseum.ufl.edu/fish/discover/species-profiles/clarias-batrachus/). Their air breathing capability allows them to establish in environments that native fish cannot live or occupy, which tends to segregate walking catfish and they cause little damage to wild populations of native fishes (Florida Fish and Wildlife Conservation Commission, 2017; http://myfwc.com/wildlifehabitats/profiles/freshwater/nonnatives/walking-catfish/). The actual ecological and economic impacts of its introduction are still not clear. One specific example of an observed economic impacts is the cost associated with the installation of barrier fences to keep walking catfish out of their ponds (Courtenay and Stauffer 1990; Nico 2005).

A combination of traits such as high fecundity, adaptation to adverse ecological conditions, the ability to "walk" between isolated water bodies make this fish a successful invader. It is able to inhabit various low-oxygen habitats such as swamps and wetlands, and

burrows inside the mudflat during summer periods (Das 1928; Islam, et al. 2007; Saha and Ratha 2007). When the original habitat dries up or after a heavy rainfall, this walking catfish could make snake-like movements from one pond to another using pectoral fins (Courtenay, et al. 1974; Das 1928; Islam, et al. 2007).

Such walking ability on terrestrial environment is attributed to its accessory respiratory organs derived from the gill tissue (Lewis 1979; Munshi 1961). These organs can breathe air as well as gills for respiring in water (Das 1928). Munshi made a detailed study of the structure of the accessory organ and described that it comprises (1) the two-compartment suprabranchial chambers situated dorsally to the gill cavities and lined by the respiratory membrane, (2) the gill fan derived from each gill arch and (3) the respiratory tree or arborescent organ derived from the second and fourth gill arches (Munshi 1961). The arborescent organ on the second gill arch is smaller while the one on the fourth arch is larger and better developed, both of which are situated in the suprabranchial chambers (Lewis 1979; Munshi 1961). During each air-breathing activity, the air from the pharynx goes into the dorsal chamber through the slit between the second and the third gill arches, and then enters the posterior chamber to escape into the opercular cavity through the slit between the third and the fourth arches (Munshi 1961). The gas exchange between the blood and the atmospheric air occurs through the highly vascularized surface of the accessory organ (Chandra and Banerjee 2003; Munshi 1961). Generally, the arborescent organs are situated in the center of the air reservoir of formed by the suprabranchial chamber and close to the inhalant aperture, providing maximum air-flow for gaseous exchange (Chandra and Banerjee 2003; Lewis 1979; Munshi 1961). Also, arborescent organs have larger relative surface than the other two components of the accessory respiratory organs (Olson, et al. 1995). Collectively, the arborescent organ may play a major role in breathing oxygen from the air.

Overall, walking catfish is an efficient bimodal breather and a successful survivor in the stressful environments, exhibiting remarkable adaptations to the extreme environmental challenges such as high environmental ammonia and hypoxic and desiccation stresses (Saha and Ratha 1998). This makes it a perfect study model for understanding of evolution of terrestrial migration, aerial respiration and high tolerance to hypoxia and ammonia.

Genetic and genomic studies of air-breathing fish species

There are many publications of air-breathing fish species, but the majority are not focused on their air-breathing function. Instead, their economic and evolutionary importance attracted more attention (Lefevre, et al. 2014a). For air-breathing part, there is one book written by Jeffrey Graham in 1997 discussing the evolution, diversity and physiological adaptation of air-breathing fishes in detail (Graham 1997). Then there are not many literature published on mechanisms and adaptations to bimodal respiration until recently more reviews on diverse aspects were published, including comprehensive descriptions for intertidal air-breathing fish species (Martin 2014), the terrestrial locomotion of air-breathing fishes (Pace and Gibb 2014), how air-breathing functions in meeting metabolic demands of swimming (Lefevre, et al. 2014b), the physiology of fish breathing air through gut (Nelson 2014), defense system against ammonia toxicity (Chew and Ip 2014) and acid-base and ion balance in air-breathing fishes (Shartau and Brauner 2014).

However, few concerns were put on the genetic and genomic studies related to air-breathing function. One study on the compromise between nutrient uptake and gas exchange in one gut-based air-breathing fish *Misgurnus anguillicaudatus* was reported in 2007 (Gonçalves, et al. 2007). The expression levels of several nutrient uptake marker genes (SGLT1, SLC5A1,

SLC15A1 and PEPT1a) were examined in the hindgut which is the respiratory zone, suggesting that it retains some potential nutrient uptake function (Gonçalves, et al. 2007). Based on this study, one Chinese group used histological observation and compared transcriptome of normal and air-breathing inhibited groups to indicate the overlap of air-breathing and nutrient uptake functions of hindgut (Huang, et al. 2016). They also compared the transcriptomes of hindgut from different time points after hatch to explore the formation of the air-breathing function, identifying 15 key pathways and 25 key genes mainly involved in development, angiogenesis and cytoskeleton (Luo, et al. 2016). In addition to *Misgurnus anguillicaudatus*, *Channa argus* was also studied for aerial respiration by comparative transcriptome analysis between gill and suprabranchial chamber, indicating that angiogenesis and elastic fibre formation pathways were up-regulated in the air-breathing organ while ion uptake and transport and acid-base balance pathways were up-regulated in the gill to produce a homeostasis for air-breathing function (Jiang, et al. 2016). Furthermore, genome sequences of two air-breathing fish species, namely mangrove rivulus fish (*Kryptolebias marmoratus*) (Kelley, et al. 2016) and four types of mudskippers (*Bolelphthalmus pectinirostris*, *Scartelaos histophorus*, *Periophthalmodon schlosseri* and *Periophthalmus magnuspinnatus*) (You, et al. 2014), were published, whereas mangrove rivulus paper only presented its draft genome assembly and the sex determination loci. However, the mudskipper paper generated the draft genome sequences and also provided insights into the genetic basis of their terrestrial adaptation in immune, DNA metabolism, ammonia excretion, vision modification, olfaction and responses to desiccation and hypoxia through comparative genome and transcriptome analyses (You, et al. 2014). Lungfish, as an air-breathing fish which was thought to be the closest relative to tetrapods, was deeply sequenced only at the transcriptomic level because the large genome size makes it unfeasible to perform

genome sequencing, assembly and annotation (Biscotti, et al. 2016). Fortunately, the evolutionary retentions and innovations of several genes related to fin to limb transition, pulmonary surfactants and urea production were determined from the reference transcriptome sequences, providing the genetic basis of adaptation of lungfish to transition from water to land.

In our case, because walking catfish *Clarias batrachus* is an aquaculture species, most of the publications are focused on its culture characters, biochemistry, toxicology, host parasite interaction and conservation studies (Debnath 2011). There are no studies researching on the genetic and genomic basis of its aerial respiration and its adaptation to terrestrial life.

INTRODUCTION

The walking catfish (*Clarias batrachus*) is a freshwater teleost species with air-breathing capability native to Southeast Asia where it is widely used as an aquaculture species due to its high economic value as food (Srivastava, et al. 2016). Unfortunately, this species has also been categorized as endangered because of over-exploitation and habitat alterations in India and Bangladesh (Dahanukar, et al. 2004; Islam, et al. 2007; Khedkar, et al. 2010). On the contrary, it is an invasive species in the United States, observed in over 10 states on the eastern and western coasts (http://maps.iucnredlist.org/map.html?id=166613), but with established populations likely found only in Florida (J. 2011). *C. batrachus* was imported into Florida from Thailand in the early 1960s (Courtenay Jr, et al. 1986). It has been thought to be damaging to native fish populations, however, there is little evidence to support this except they do invade aquaculture facilities and can cause severe damage to cultured fish populations (Florida Museum, University of Florida 2017; https://www.floridamuseum.ufl.edu/fish/discover/species-profiles/clarias-batrachus/). In actuality, their air breathing capability allows them to establish in environments that native fish cannot live or occupy, which tends to segregate walking catfish and they cause little damage to wild populations of native fishes (Florida Fish and Wildlife Conservation Commission, 2017; http://myfwc.com/wildlifehabitats/profiles/freshwater/nonnatives/walking-catfish/).

A combination of traits such as high fecundity, adaptation to adverse ecological conditions, the ability to "walk" between isolated water bodies make this fish a successful invader. It is able to inhabit various low-oxygen habitats such as swamps and wetlands, and burrows inside the mudflat during summer periods (Das 1928; Islam, et al. 2007; Saha and Ratha 2007). When the original habitat dries up or after a heavy rainfall, the walking catfish can make

snake-like movements to move from one body of water to another pulling with its pectoral fins (Courtenay, et al. 1974; Das 1928; Islam, et al. 2007). The accessory air-breathing organ is the key to survival as the *Clarias* "walks" terrestrially to the next aquatic environment. Although derived from the gill tissue, the air-breathing organ consists of suprabranchial chambers, gill fans and arborescent organs (Lewis 1979; Munshi 1961). They can breathe air as well as using gills for respiration in water (Das 1928). The walking abilities allow *Clarias* to cope with respiration challenges without a lung in the terrestrial environment, as well as adaptation to extreme environmental challenges such as high ammonia as well as hypoxic and desiccation stresses (Saha and Ratha 1998). This makes *C. batrachus* a perfect model for studying the evolution of terrestrial migration, aerial respiration and high tolerance to hypoxia and ammonia.

Recent genome projects demonstrated that comparative genomic analysis combined with transcriptomic analysis allow the elucidation of the genomic basis associated with the adaptation to terrestrial life in mangrove rivulus fish (*Kryptolebias marmoratus*) and mudskippers (*Bolelphthalmus pectinirostris*, *Scartelaos histophorus*, *Periophthalmodon schlosseri* and *Periophthalmus magnuspinnatus*) (Kelley, et al. 2016; You, et al. 2014). Mangrove rivulus fish mainly utilizes its skin and mudskipper mainly utilizes its buccal cavity to breathe air (Randall, et al. 2004; Wright 2012), while *C. batrachus* possesses an accessory air-breathing organ. It is of great interest to determine the genomic basis of adaptation of aquatic species to terrestrial environment. Here I present the draft genome sequence assembly of *C. batrachus* and the characterization of its genomic features potentially for terrestrial adaptation by comparative genomic and transcriptomic analyses.

MATERIALS AND METHODS

C. batrachus samples and genome sequencing

One wild walking catfish (*Clarias batrachus*) was collected from Florida, USA in June of 2014. Genomic DNA was extracted from the blood cells using DNeasy Blood and Tissue kit (Qiagen, CA). One short insert size (180bp) paired-end library and one long insert-size (3 kb) library were constructed. Each library was subjected to 2 x 100 bp read length PE runs on an Illumina HiSeq 2500 sequencer at HudsonAlpha (Huntsville, AL, USA).

Genome assembly and assessment

After raw reads were evaluated in FastQC v0.11.4 (Andrews 2010), low-quality bases and adapter sequences were trimmed from the raw sequences using cutadapt v1.8.1 (Martin 2011), and then short reads with length shorter than 30 bases after trimming were removed. The genome sequences were assembled by ABySS v1.5.2 (Simpson, et al. 2009) with k-mers ranging from 40 to 70 in size and ALLPATHS-LG (Gnerre, et al. 2011). Finally, k-mer size of 61 yielded the best assembly results using ABySS. To increase scaffold length, I selected assembled sequences with longer contig N50 from ALLPATHS-LG for scaffolding by SSPACE v3.0 (Boetzer, et al. 2011). Finally, the paired-end reads were utilized to fill the gaps in the scaffolds with Gapfiller v1.10 (Boetzer and Pirovano 2012). Genome size was estimated in the ALLPATHS-LG using trimmed paired-end reads.

To assess the quality of the assembly results, CEGMA v2.5 (Core Eukaryotic Genes Mapping Approach) (Parra, et al. 2007) was first employed to evaluate the completeness of *C. batrachus* genome sequences. In other words, 248 highly conserved core eukaryotic genes

(CEGs) in six typical genomes (*Homo sapiens*, *Drosophila melanogaster*, *Caenorhabditis elegans*, *Arabidopsis thaliana*, *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe*) (Parra, et al. 2007) were mapped to the assemblies to display the percentage of CEGs present in the *C. batrachus* genome. Another assessment procedure, BUSCO v1.22 (Benchmarking Universal Single-Copy Orthologs) (Simão, et al. 2015), was used to evaluate the completeness of genome assembly by 3,023 genes selected from orthologous groups with single-copy orthologs in > 90% of available vertebrate genomes. Then, the 5 longest scaffolds of assembly resulting from another assembler (ABySS) were mapped against the genome sequence using NUCmer in MUMmer v3.23 (Delcher, et al. 2003) to evaluate the aligned identity. Lastly, RNA-Seq reads were assembled from gill and the air-breathing organ in this study by Trinity v2.0.6 (Haas, et al. 2013) and then the assembled fragments were mapped to the genome assembly using BLAT v35 (Kent 2002) to assess the gene coverage of the assembly.

Genome annotation

A *de novo* repeat library was constructed using RepeatModeler v1.0.8 (http://www.repeatmasker.org/RepeatModeler.html) which contains two *de novo* repeats finding programs, RECON (Bao and Eddy 2002) and RepeatScout (Price, et al. 2005). Next RepeatMasker v4.0.6 (http://www.repeatmasker.org/) was used to predict and categorize repeat sequences in the *C. batrachus* genome with the repeat library constructed from RepeatModeler. The Jukes-Cantor formula (Jukes and Cantor 1969) was used to calculate the average number of substitutions per site for each fragment based on the divergence levels from the results of RepeatMasker. For the subsequent genome annotation, the genome sequences were masked with "N" in the repeat regions except low complexity DNA or simple repeats.

AUGUSTUS v3.2.1 (Stanke, et al. 2004) was used for the *ab initio* predictions of genes on the repeat-masked genome. Gene model parameter sets for AUGUSTUS were trained from genes in *Danio rerio*. The predicted genes with length less than 30 amino acids were removed. The remaining predicted amino acid sequences were aligned to NCBI non-redundant (nr) protein database and SwissProt and TrEMBL subsets of the UniProt database (Boeckmann, et al. 2003) by BLASTP with an e-value cut-off of 1e – 5 to identify homologous genes. Functional categories of GO terms were determined by Blast2GO version 4.0.7 (Conesa, et al. 2005), and KEGG Automatic Annotation Server (KAAS) (http://www.genome.jp/tools/kaas/) BBH (bi-directional best hit) method (Moriya, et al. 2007) was used to perform biological pathways analysis.

Comparative genomic analysis

The protein sequences of channel catfish (*Ictalurus punctatus*; NCBI version IpCoco_1.2) were downloaded for comparison with those of the *C. batrachus* to determine *C. batrachus* specific genes in the catfish lineage. The methodology was based on the one used in the channel catfish genome paper (Liu, et al. 2016). First, the proteins from both catfish were sent to OrthoFinder v1.0.2 (Emms and Kelly 2015) for an all-to-all BLASTP with an e-value threshold of 1e – 5 and subsequent clustering into orthogroups based on MCL algorithm. Next, a further round of BLASTP was performed using the genes not included in the orthogroups to query against the genes in the orthogroups within the same species with an e-value threshold of 1e – 10. In the end, a reciprocal BLASTP between them with an e-value threshold of 1e – 5 was performed using genes with no hits from last step as queries. The remaining genes in *C.*

batrachus were considered as specific genes and kept for further GO term overrepresentation test by PANTHER version 11 (Mi, et al. 2016) with the best homologous gene ID from zebrafish.

Protein sequences of an additional 10 teleost fish species including zebrafish (*Danio rerio*; Ensembl version GRCz10), stickleback (*Gasterosteus aculeatus*; Ensembl version BROAD S1), tetraodon (*Tetraodon nigroviridis*; Ensembl version TETRAODON8.0), fugu (*Takifugu rubripes*; Ensembl version FUGU4.0), medaka (*Oryzias latipes*; Ensembl version HdrR), cod (*Gadus morhua*; Ensembl version fadMor1), cave fish (*Astyanax mexicanus*; Ensembl version AstMex102), Nile tilapia (*Oreochromis niloticus*; Ensembl version Orenil1.0), platyfish (*Xiphophorus maculatus*; Ensembl version Xipmac4.4.2) and amazon molly (*Poecilia formosa*; Ensembl version Poecilia_formosa-5.1.2) were downloaded for inferring orthologues. The longest protein sequence was selected for each gene in the 11 sequenced fish species (channel catfish included). After combination with *C. batrachus* protein sequences, all the sequences were sent to OrthoFinder v1.0.2 (Emms and Kelly 2015) to identify orthologues and orthogroups among these species. Genes present in the *C. batrachus* genome but absent from the non-air-breathing fishes were obtained. Next these genes specific for *C. batrachus* were searched with all existing sequences from non-air-breathing fishes in the NCBI databases to determine the genes present only in the *C. batrachus* genome.

The single-copy genes were extracted from all the species to construct a phylogenetic tree. Multiple sequence alignments were performed using MUSCLE v3.8.31 (Edgar 2004) for protein alignments and PAL2NAL (Suyama, et al. 2006) for codon alignments. I used Gblock v0.91b (Talavera and Castresana 2007) to eliminate poorly aligned positions and divergent regions of the alignments. Final alignments with length shorter than 50 amino acids for protein alignments and 150 bp for codon alignments were removed. AMAS (Borowiec 2016) was

performed to combine all the refined alignments into a super alignment. PartitionFinder v2.0.0 was used to determine the best substitution model for each gene with the parameters of -rcluster-percent = 20.0 (Lanfear, et al. 2016). Then I used the rapid bootstrap algorithm with a thorough ML search (-f a) and 100 bootstrap replicates in RAxML v8.2.9 (Stamatakis 2014) to construct a maximum likelihood tree from those single-copy genes.

To determine positively selected genes in *C. batrachus*, the single-copy genes were collected for analyzing the dN/dS ratio. The values of dN and dS and dN/dS ratio were estimated using codeml program in the PAML package version 4.9 (Yang 2007). Sequence alignments with dS value greater than 2 were removed to avoid distortion of the dN/dS ratio by saturation of synonymous substitutions (Gojobori 1983). The values of dN/dS between each species branch and the ancestral branch from 150 randomly picked genes were estimated with 10,000 bootstrap replicates to evaluate the magnitude of natural selection acting on each species. Then branch-site model was used to designate *C. batrachus* as a "foreground" branch and the rest of the species as "background". A likelihood ratio test (LRT) was computed to compare a model that allows sites to be under positive selection ($\omega > 1$) on the foreground branch with the null model that allows sites to be under negative selection ($\omega < 1$) and evolve neutrally ($\omega = 1$) with a posterior probability greater than 0.95 based on the Bayes Empirical Bayes (BEB) results. After an FDR multiple testing correction (FDR < 0.05), the positively selected genes were selected for further GO term enrichment analysis by Blast2GO version 4.0.7 (Conesa, et al. 2005) with the whole reference gene set as the background for statistical analysis. After assigning *C. batrachus* genes with the best homologous zebrafish genes by BLASTP and Ensembl BioMart (Kinsella, et al. 2011), Reactome pathway database v60 (Fabregat, et al. 2016) was used for further pathway enrichment analysis.

Gene family analysis

Orthologous genes were sent to CAFÉ v3.0 (Han, et al. 2013) program to assess gene family expansion and contraction (-r 1000 -s). A family-wide *P*-value less than 0.01 and a branch-specific *P*-value less than 0.001 was utilized to identify gene family expansion in the *C. batrachus* genome. The expanded families in the *C. batrachus* genome were also searched in the NCBI database to exclude false positive expansions due to the limited number of species in previous analyses, during which the number of genes in tetraploid species were divided by two for comparison.

For those significantly expanded genes, phylogenetic trees were constructed to display the gene expansions. The accession numbers of all the protein sequences used in the phylogenetic analyses were listed in the Supplementary Table 10. Multiple sequence alignments were performed using ClustalW (Thompson, et al. 2002) in MEGA6 (Tamura, et al. 2013) and ProtTest v3.4 (Darriba, et al. 2011) was utilized to select the best model for constructing phylogenetic trees (Supplementary Table 10). Phylogenetic analysis was conducted using MEGA6 with the maximum likelihood method. The bootstrapping with 1,000 replications was conducted to evaluate the phylogenetic tree. RNA-Seq data from gill and the air-breathing organ in this study and also from brain (SRR2057993), head kidney (SRR2057995) and liver (SRR2052655) in the public database were mapped to genome sequences to estimate the Fragments Per Kilobase Per Million mapped fragments (FPKM) of those expanded genes respectively by TopHat 2.0.10 and Cufflinks 2.1.1 (Trapnell, et al. 2009; Trapnell, et al. 2012; Trapnell, et al. 2010).

Comparative transcriptomic analysis between gill and the air-breathing organ

Wild *C. batrachus* individuals (70-136 g) were collected from Miami, Florida, USA in October of 2015. The tissue samples of gill and air-breathing organ were kept in the RNAlater solution (Ambion) to prevent RNA degradation. Total RNAs were extracted from tissues of five individuals using RNeasy Plus Universal Mini kit (Qiagen, CA) according to manufacturer's instructions, and then the RNA from five samples were mixed in equal amounts for RNA-Seq at HudsonAlpha (Huntsville, AL, USA). Standard Poly A libraries were prepared and 125 paired-end reads were generated using Illumina HiSeq 2500 sequencing platform. Raw reads were filtered with the parameters of base quality ≥ 20 and trimmed length ≥ 36 bp by Trimmomatic v0.32 (Bolger, et al. 2014). All the trimmed reads from both tissues were mapped to 22,914 coding sequences predicted from genome assembly by the CLC Genomics Workbench software package. The parameters for mapping were set as 90% or greater sequence identity with a maximum of 2 mismatches. The number of total mapped reads on each contig and Reads Per Kilobase per Million mapped reads (RPKM) were collected. After normalization of RPKM values, fold changes were estimated to exhibit differentially expressed patterns between the air-breathing organ and gill transcriptomes with a P-value < 0.05 using proportions-based Kal's test in the CLC Genomics Workbench software package. Transcripts with fold change values greater than 2 were determined as differentially expressed genes for subsequent analysis. Blast2GO version 4.0.7 (Conesa, et al. 2005) was used with default settings for the over-representation analysis of GO term among the differentially expressed genes in the air-breathing organ and gill, and Reactome pathway database v60 (Fabregat, et al. 2016) was used for further pathway enrichment analysis to indicate the functional differences between air-breathing organ and gill.

Quantitative real-time PCR validation of differentially expressed genes

To confirm the accuracy of RNA-Seq analysis, quantitative real-time PCR (qRT-PCR) analysis was conducted in this study. Total RNAs were extracted from the gill and air-breathing organ using RNeasy Plus Universal Mini kit (Qiagen, CA) following the manufacturer's instructions. After quantification with Nanodrop (Thermo Scientific), cDNA was synthesized with a final concentration of 50 ng/μL using the iScript cDNA Synthesis Kit (Quanta BioSciences) based on the manufacturer's protocol. The primers used in qRT-PCR are listed in Table 1. Amplification was performed on a CFX96 real-time PCR Detection System (Bio-Rad, CA). The thermal cycling profile consists of an initial denaturation at 95 °C for 30s, 40 cycles

Table 1. Primer sequences used for qRT-PCR.

Gene name	Gene ID	Forward primer	Reverse primer
28S rRNA		5' TTTCAGGGCTAGTTGATTCGGCAG	5' GGTTGATATAGACAGCAGGACGGT
hba	g20835.t1	5' ACAAGGCCGTTGTGAAAGAC	5' GCCAGTGAGCGAAGTAGGTC
hba	g20837.t1	5' ACAAGGCCGTTGTGAAGGAG	5' CCCAGTGAGCGAAGTAGGTC
lyg	g3626.t1	5' GTCACCCAAGGCACTGAAAT	5' TAGGCTGAGATTCCCCCTTT
tgfb2	g2430.t1	5' GCTTTGCAGGGATTGATGAT	5' TCTGTCTGTGGCTCTTGTGC
fbln1	g4468.t1	5' ACAGGCGAGCAAGTGAAACT	5' ACAGGGATGTTGATGGGAAA
tgfb3	g14668.t1	5' CATCGATTTCCGACAGGACT	5' AGCGAACTGTGTGTGGTGTC
s1pr1	g5232.t1	5' GAAATTCCGCAAAGTGGAAA	5' TCGTTCGCCAAATAGTGAGA
sem3e	g5331.t1	5' CGCCTCATGGCTTATTGTTT	5' CTCTTCCAGCACCTCCAAAG
bmp4	g189.t1	5' TCTGCATCTGAACCGAACAC	5' TCATGACCGAAAGTGACCAA
smad3	g3356.t1	5' AGCTGGACGAGTTGGAGAAA	5' TCACATGAGGGAGACCCTTC
at233	g9311.t1	5' GAGCCTGCTCAGGCATAGAT	5' GCGACACATTGCTTTTGAGA
cahz	g8816.t1	5' AGTGACCTTCACGGATGAGG	5' TGTGCTCAGACCCTTTGTTG
cah2	g3019.t1	5' GGCTCAGAGCACACAGTTGA	5' GGAGCCATCGTATGTCCAGT
mb	g8414.t1	5' CACCCTGACACCCAGAAACT	5' TTTGCCTTGAGGATCTCACC
cah6	g456.t1	5' AACCTGGGTCATCCAATGAA	5' TGAGCTCGTTTTCGTGTTTG
slc9a3	g3908.t1	5' TTGCTGTGTTCGAAGAGGTG	5' CCAGAGACACAAAGGCATCA
slc4a1	g12078.t1	5' TGTAGGCCTGTCCATCATCA	5' TATGCGATCCCACAGTTGAA

of denaturation at 94 °C for 5s and an appropriate annealing/extension temperature at 60 °C for 10s, and 72 °C for 5s, followed by dissociation curve analysis to validate the specificity of

amplified products. The gene 28S ribosomal RNA (rRNA) (Mohindra, et al. 2014) (accession number JK488212) was set as the reference gene. Relative fold changes for each gene were calculated in the Relative Expression Software Tool (REST) version 2009 (Pfaffl, et al. 2002) based on the values of cycle threshold from real-time PCR.

RESULTS AND DISCUSSION

Genome assembly and annotation

The statistics of the draft genome sequence assembly are shown in Table 2. The final assembly contained 10,041 scaffolds, with a scaffold N50 of 361.2 kb. The assembly covered a total of 821 Mb, similar to the genome size of 854 Mb estimated from ALLPATHS-LG, but slightly smaller than the 900 Mb based on feulgen densitometry method (Jianxun, et al. 1991) and the 1.17 Gb based on bulk fluorometric assay method (Hinegardner and Rosen 1972).

Table 2. Summary statistics of walking catfish (*Clarias batrachus*) genome sequencing, assembly and annotation.

Genome sequencing				
Library	# of reads	Read length	Trimmed data	Genome coverage
Paired-end 180 bp	426M	100 bp	39.2 Gb	46 X
Mate-pair 3 Kb	483M	100 bp	32.1 Gb	38 X
Genome assembly				
	Total	N50	Longest	Assembled size
Contigs	87,962	19.0 Kb	194 Kb	747.7 Mb
Scaffolds	10,041	361.2 Kb	2,843 Kb	821.8 Mb
Genome annotation				
Number of genes	Repetitive elements content			
22,914	30.3%			

The completeness of the genome assembly was assessed by mapping the 248 core eukaryotic genes (CEGs) from CEGMA v2.5 (Parra, et al. 2007) to the genome sequences. The draft genome sequences covered over 95.2% of the CEGs (Table 3). When the 3,023 genes from vertebrate BUSCO (Simão, et al. 2015) were mapped to the genome assembly, the draft genome sequences included 83.9% of these genes (Table 3). As our objective was to identify any additional genes in the walking catfish not found in non-air-breathing fishes, which may account

for its adaptation on land, this level of completeness is reasonable, although a small percentage of missing genes may reduce the full capacity of identifying such genes. The assembly was assessed to be also accurate. The 5 longest scaffolds (1.3 Mb-2.2 Mb) assembled using a second software ABySS had 99.4% alignments with the genome sequence assembled using ALLPATH-LG (Table 4). At last, the *de novo* assembled transcriptomes from the gill and the air-breathing organ were mapped to the draft genome sequences, showing that our assembly covered 81.7% and 83.4% of transcribed regions with a nucleotide sequence similarity threshold of 80%, respectively (Table 5). Overall, these results indicated that the assembled genome of *C. batrachus* had high coverage and was of high quality for further analysis.

Table 3. Completeness of genome assembly assessed by CEGMA and BUSCO.

	Complete orthologues (%)	Partial orthologues (%)	# Total orthologues analyzed
CEGMA	73.4	21.8	248
BUSCO	72.8	11.1	3,023

Table 4. Mapping of five longest ABySS scaffolds to genome assembly.

ABySS scaffold ID	ABySS aligned length (bp)	scaffold ID	Aligned length (bp)	Aligned identity (%)
3829531	2,275,460	scaffold7; scaffold192	2,273,636	99.77
3833592	1,609,354	scaffold107; scaffold129	1,607,455	99.78
3829461	1,444,754	scaffold2	1,442,715	99.54
3821019	1,565,771	scaffold8	1,564,193	99.75
3829340	1,371,683	scaffold17	1,370,520	99.43

Table 5. Assessment of gene coverage based on transcriptomes of gill and air-breathing organ.

Gill							
Assembled transcript length	Number	Covered by assembly	Percentage (%)	With ≥80% of sequence in one scaffold		With ≥50% of sequence in one scaffold	
				Number	Percentage	Number	Percentage
≥200 bp	489,021	472,484	96.6	399,317	81.7	444,825	91.0
≥500 bp	196,940	193,809	98.4	160,117	81.3	182,354	92.6
≥1000 bp	98,961	98,433	99.5	78,782	79.6	92,616	93.4
≥2000 bp	43,546	43,499	99.9	33,366	76.6	40,730	93.5
Air-breathing organ							
Assembled transcript length	Number	Covered by assembly	Percentage (%)	With ≥80% of sequence in one scaffold		With ≥50% of sequence in one scaffold	
				Number	Percentage	Number	Percentage
≥200 bp	430,765	421,754	97.9	359,217	83.4	397,903	92.4
≥500 bp	175,272	173,712	99.1	145,089	82.8	163,873	93.5
≥1000 bp	89,879	89,626	99.7	72,845	81.0	84,475	94.0
≥2000 bp	40,333	40,312	99.9	31,632	78.4	37,845	93.8

The *C. batrachus* genome had a GC content of 39.2%, similar to other fish species (Tarallo, et al. 2016; Vinogradov 1998). The repetitive elements comprise 30.3% of the genome size (Table 1, Table 6). Although the types of repetitive elements in the *C. batrachus* genome are

Table 6. Repetitive elements in the *C. batrachus* genome.

Repetitive elements	Number of elements	Length occupied (bp)	Percentage of genome (%)
Interspersed repeats			
SINEs	61,630	9,458,516	1.15
LINEs	110,499	27,867,994	3.39
LTR elements	113,379	30,149,843	3.67
DNA elements	577,729	126,278,979	15.37
Unclassified	292,027	54,288,506	6.61
Subtotal	*1,155,264*	*248,043,838*	*30.18*
Tandem repeats			
Satellites	2,097	624,424	0.08
Simple repeats	1,516	135,706	0.02
Subtotal	*3,613*	*760,130*	*0.1*
Total	**1,158,877**	**248,803,960**	**30.28**

similar to those in the channel catfish genome (Liu, et al. 2016), it appeared that the number of substitutions per site for *C. batrachus* repetitive elements exhibited a peak at about 16% (Figure 1), higher than that of channel catfish repetitive elements (~10%, Figure 1), indicating that its repetitive elements had a longer evolutionary history and/or are more active in evolution than those of channel catfish genome (Li and Graur 1991).

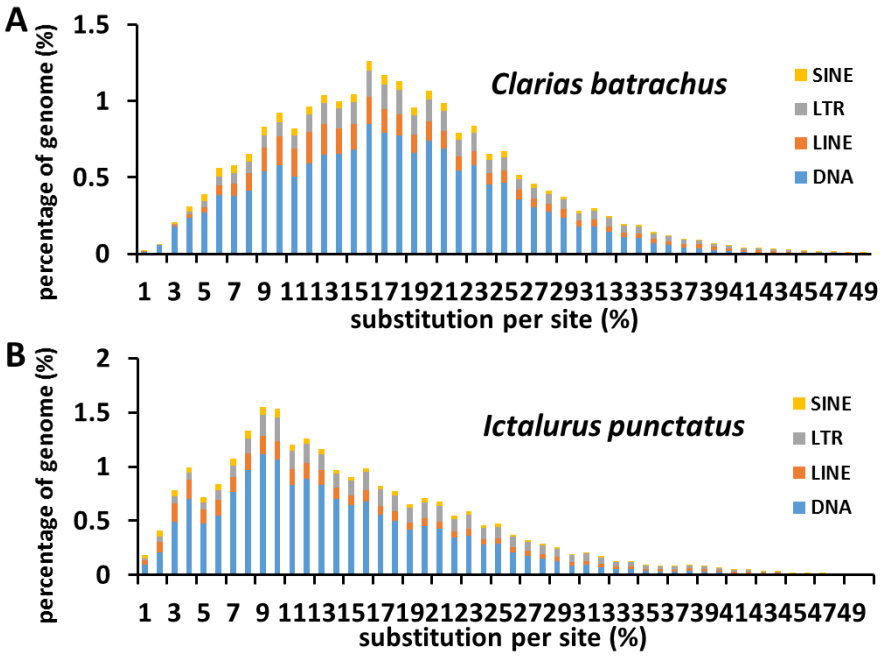


Figure 1. Comparison of repetitive elements between walking catfish genome and channel catfish genome. The distribution of repetitive elements and their contents were shown in *Clarias batrachus* (**A**) and *Ictalurus punctatus* (**B**). The average number of substitutions per site for each fragment was estimated by the Jukes-Cantor formula. SINE, short interspersed elements; LTR, long terminal repeats; LINE, long interspersed elements.

A total of 22,914 genes were annotated from the *C. batrachus* genome sequences, of which 19,834 genes (86.6%) were supported by RNA-Seq evidence from the gill and the air-breathing organ. Among the identified protein-coding genes, the majority (22,587, 98.6%) were

supported by matches from at least two publicly available databases including non-redundant protein database, SwissProt and TrEMBL subsets of the UniProt database (Boeckmann, et al. 2003), KEGG and GO terms (Figure 2).

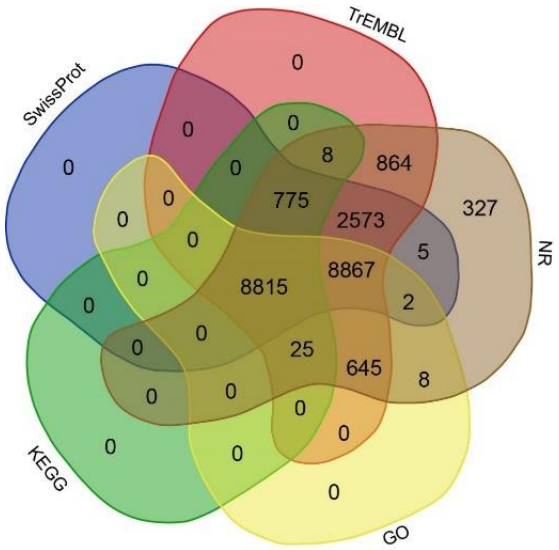


Figure 2. Venn diagram showing the number of homologues of the predicted genes from the *C. batrachus* genome in various databases. NR, non-redundant protein database.

Comparative genomic analysis

To identify the genes specific to the *C. batrachus* genome, I first compared the genes between the walking catfish and channel catfish (Figure 3). They both belong to the same order Siluriformes, and therefore shared the highest number of orthogroups compared to other fish species in this study (Table 7), but the walking catfish possesses the air-breathing organ while the channel catfish does not. A total of 1,854 genes were present in the walking catfish, but absent from channel catfish. These genes were enriched for "DNA repair", "enzyme activator activity" and "small GTPase regulator activity" (Table 8), which may be associated with its

adaptation to the terrestrial life, such as responding to increased DNA damage and accelerated metabolic processes. Small GTPases are well-known for maintaining cell adhesion, cell migration, gene transcription and cytogenesis (Narumiya 1996; Toksoz and Merdek 2001), and one of their critical modulators, namely "guanyl-nucleotide exchange factor activity", was also found to be significantly enriched (Table 8). Furthermore, small GTPases were also reported to be under selective sweeps in the alkaline-tolerant population compared with the flowing freshwater population of *Leuciscus waleckii*, indicating their roles in regulating ion transport and acid-base balance under extreme environmental conditions (Xu, et al. 2016).

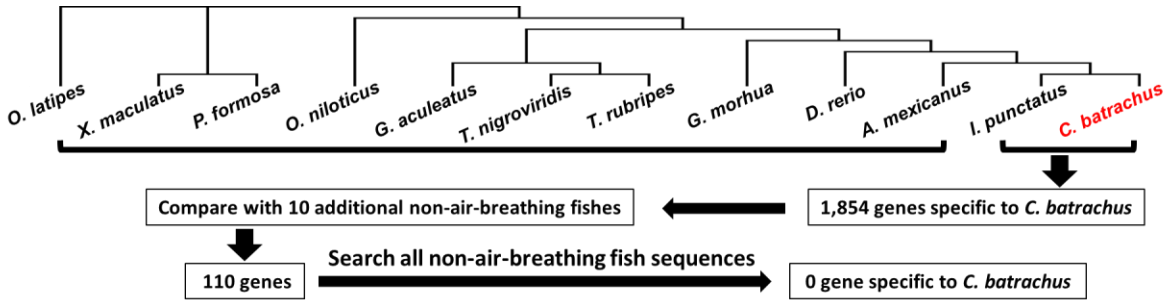


Figure 3. Schematic presentation of comparative genomic analyses of *C. batrachus* against non-air-breathing teleost fishes. In the catfish lineage, the protein sequences of *C. batrachus* genome and *I. punctatus* genome were comparatively subtracted, resulting in the identification of 1,854 genes specific to the *C. batrachus* genome; similarly, 10 additional non-air-breathing fishes were added to compare with *C. batrachus* genome, resulting in the identification of 110 genes that were still only present in the *C. batrachus* genome. The names of these 110 *C. batrachus* specific genes were searched with all existing sequences from non-air-breathing fish species in the NCBI databases, resulting in no genes specific to the *C. batrachus* genome.

Table 7. Orthogroups shared between species by OrthoFinder.

	Poecilia formosa	*Astyanax mexicanus*	*Ictalurus punctatus*	*Clarias batrachus*	*Gadus morhua*	*Takifugu rubripes*	*Oryzias latipes*	*Xiphophorus maculatus*	*Lepisosteus oculatus*	*Gasterosteus aculeatus*	*Tetraodon nigroviridis*	*Oreochromis niloticus*
Astyanax mexicanus	11,157											
Ictalurus punctatus	11,578	12,986										
Clarias batrachus	10,613	11,941	12,806									
Gadus morhua	10,969	11,684	12,184	11,205								
Takifugu rubripes	10,751	11,471	11,934	10,973	11,509							
Oryzias latipes	10,887	11,466	11,920	10,976	11,429	11,323						
Xiphophorus maculatus	11,742	12,016	12,448	11,380	11,863	11,660	11,675					
Lepisosteus oculatus	10,821	11,850	12,403	11,285	11,332	11,118	11,084	11,586				
Gasterosteus aculeatus	11,328	11,958	12,454	11,415	12,556	11,793	11,774	12,256	11,557			
Tetraodon nigroviridis	10,749	11,494	11,935	11,010	11,448	11,652	11,356	11,648	11,087	11,730		
Oreochromis niloticus	11,151	11,794	12,210	11,191	11,587	11,582	11,472	12,056	11,410	11,933	11,478	
Danio rerio	11,020	12,216	12,822	11,642	11,543	11,266	11,298	11,799	11,694	11,792	11,278	11,594

Table 8. GO slims significantly enriched in the specific genes in the *C. batrachus* genome compared with that of channel catfish (*I. punctatus*).

GO-ID	GO Term	Category	Bonferroni-corrected *P*-value
GO:0006281	DNA repair	BP	1.66E-04
GO:0006259	DNA metabolic process	BP	1.76E-03
GO:0009987	cellular process	BP	1.52E-04
GO:0008047	enzyme activator activity	MF	1.19E-02
GO:0030234	enzyme regulator activity	MF	2.63E-05
GO:0005085	guanyl-nucleotide exchange factor activity	MF	6.52E-03
GO:0005083	small GTPase regulator activity	MF	5.65E-06
GO:0030054	cell junction	CC	3.85E-02

To further narrow down the list of genes potentially present in the walking catfish but absent in non-air-breathing fishes, the status of the 1,854 genes were determined in 10 additional non-air-breathing fish species (Figure 3). Only 110 genes were then found to be present in the walking catfish, but absent in the 10 non-air-breathing fishes. However, when these genes were further searched, they had at least one match with all existing sequences from all non-air-breathing fish species in the NCBI, resulting in no genes that were specific to the walking catfish (Figure 3). Although it is possible that the genome sequence assembly is incomplete and specific genes of the walking catfish could have potentially been missed, it is unlikely that the walking catfish harbors a large number of specific genes for air-breathing in terrestrial environment.

As no genes were found to be specific to *C. batrachus*, I also determined the ratio of non-synonymous substitutions to synonymous substitutions using 705 single-copy orthologous genes in the *C. batrachus* and the other 11 non-air-breathing fishes. As shown in Figure 4, *C. batrachus* apparently evolved rapidly with the second highest dN/dS ratio, next only to *X. maculatus*, suggesting the walking catfish genome was under heavy positive selection pressure.

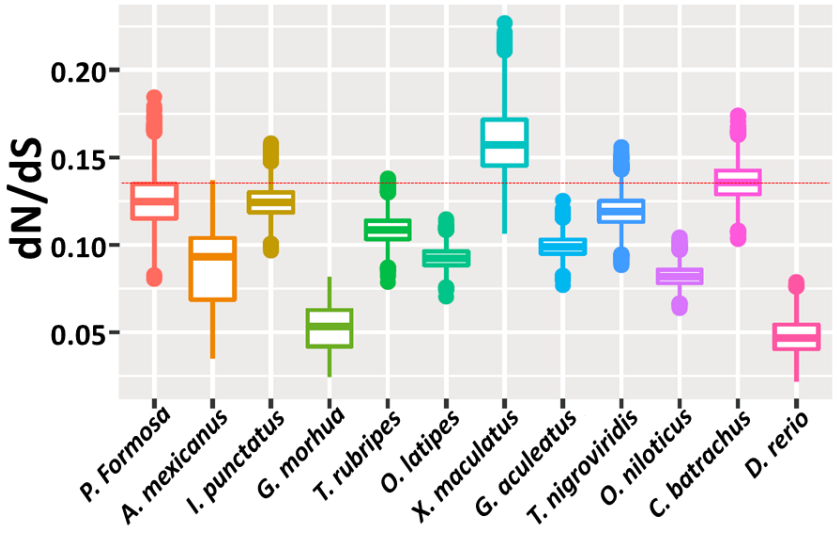


Figure 4. Comparison of the values of dN/dS ratio among various fish species against the ancestor. The average dN/dS values were estimated from 150 randomly picked single-copy genes with 10,000 bootstrap replicates. The red line represents the average dN/dS value in *C. batrachus*, noting that it is the second most rapidly evolving genome.

Of the 705 single-copy genes, 132 were positively selected genes (Appendix 1). These genes were mainly enriched in the cellular component, including "mitochondrial intermembrane space", "nucleoplasm part", "RNA polymerase II transcription factor complex" and "nuclear DNA-directed RNA polymerase complex" (Table 9), indicating the accelerated evolution of genes involved in gene expression in *C. batrachus*. The overrepresented pathway "gene expression" included a list of genes related to transcription factor (*med6*, *med14*, *gtf2e2*, *mnat1* and *nfyc*), RNA binding protein (*paip1*), mRNA splicing factor (*cstf2*, *sf3b2*, *rbm8a* and *cpsf5*), chromatin binding (*noc2l*) and translation initiation factor (*eif3m*). It has been well documented that different organisms can achieve diverse and specific responses to multiple environmental stresses by regulating gene expression to maintain homeostasis (Atkinson 2012; De Nadal, et al.

2011; Murray, et al. 2004). Additionally, GO term named "cellular nitrogen compound metabolic process" was found to be enriched in these genes under positive selection (Table 9). Ammonia is

Table 9. GO terms significantly enriched in the positively selected genes in the *C. batrachus* genome.

GO-ID	GO Term	Category	FDR
GO:0070013	intracellular organelle lumen	CC	5.30E-08
GO:0031974	membrane-enclosed lumen	CC	5.30E-08
GO:0043233	organelle lumen	CC	5.30E-08
GO:0044446	intracellular organelle part	CC	5.30E-08
GO:0044422	organelle part	CC	1.16E-07
GO:0044428	nuclear part	CC	5.82E-06
GO:0043231	intracellular membrane-bounded organelle	CC	3.44E-05
GO:0043227	membrane-bounded organelle	CC	1.23E-04
GO:0032991	macromolecular complex	CC	1.74E-04
GO:0031981	nuclear lumen	CC	1.01E-03
GO:0043229	intracellular organelle	CC	1.80E-03
GO:0044424	intracellular part	CC	2.03E-03
GO:0005739	mitochondrion	CC	2.71E-03
GO:0043226	organelle	CC	3.93E-03
GO:0044429	mitochondrial part	CC	3.93E-03
GO:0005758	mitochondrial intermembrane space	CC	6.08E-03
GO:0044451	nucleoplasm part	CC	6.42E-03
GO:0031970	organelle envelope lumen	CC	6.65E-03
GO:0090575	RNA polymerase II transcription factor complex	CC	6.99E-03
GO:0043234	protein complex	CC	8.40E-03
GO:0044798	nuclear transcription factor complex	CC	1.18E-02
GO:0005634	nucleus	CC	1.63E-02
GO:0005654	nucleoplasm	CC	1.79E-02
GO:0005622	intracellular	CC	3.13E-02
GO:0044444	cytoplasmic part	CC	3.47E-02
GO:0034641	cellular nitrogen compound metabolic process	BP	4.46E-02
GO:0055029	nuclear DNA-directed RNA polymerase complex	CC	4.58E-02
GO:0030880	RNA polymerase complex	CC	4.58E-02
GO:0000428	DNA-directed RNA polymerase complex	CC	4.58E-02

the main nitrogenous waste produced continuously in fish which is highly toxic and needs to be excreted promptly or converted to other less-toxic chemicals. *C. batrachus* usually inhabits water bodies with high levels of ammonia and sometimes dwells inside mudflats or "walks" on the land, during which excretion of ammonia directly into the aqueous environment through the gill

is impossible (Saha and Ratha 2007). To adapt to the hyper-ammonia stress, *C. batrachus* is highly tolerant to external ammonia and can convert ammonia into non-essential amino acids and less-toxic urea through ornithine-urea cycle (OUC) (Saha and Das 1999; Saha, et al. 2002; Saha, et al. 2000; Saha and Ratha 2007). Interestingly, the gene argininosuccinate synthase (*ass*) encoding one of the key enzymes in the OUC was identified to be under positive selection in comparison with non-air-breathing fish species, implying the necessity of this adaptive strategy for air-breathing walking catfish to survive in the hyper-ammonia environments.

Gene expansion

In addition to analysis of positive selection on single-copy genes, levels of gene expansion and contraction were determined. *C. batrachus* experienced expansion within 1,657 families and contraction in 1,752 families (Figure 5). Among the 12 fish species, it had the

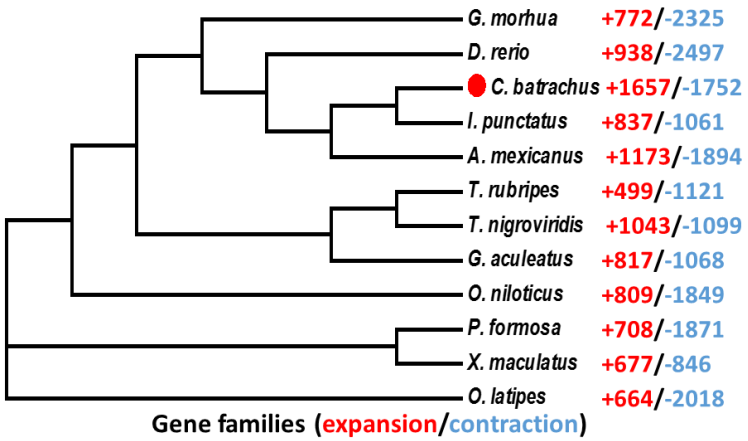


Figure 5. The number of gene families with expansion (red) / contraction (blue). *C. batrachus* is marked with red solid circle, noting that it has the largest number of expanded gene families.

largest number of expanded gene families, suggesting that its adaptation to terrestrial lifestyle may be partly mediated by gene family expansion. Among the 1,657 expanded gene families, three families were significantly expanded in *C. batrachus*: myoglobin (*mb*), olfactory receptor related to class A G protein-coupled receptor 1 (*ora1*) and sulfotransferase 6b1 (*sult6b1*).

Myoglobin gene exists in almost all the vertebrate species with one-to-two copies in the genome except the seven copies in the lungfish genome (Fraser, et al. 2006; Fuchs, et al. 2006; Hoffmann, et al. 2011; Koch, et al. 2016; Roesner, et al. 2008; Schwarze, et al. 2014; Sidell and O'Brien 2006). I found a huge expansion of myoglobin, fifteen copies of the gene, in the *C. batrachus* genome (Figure 6). Multiple sequence alignments showed some diversity among them

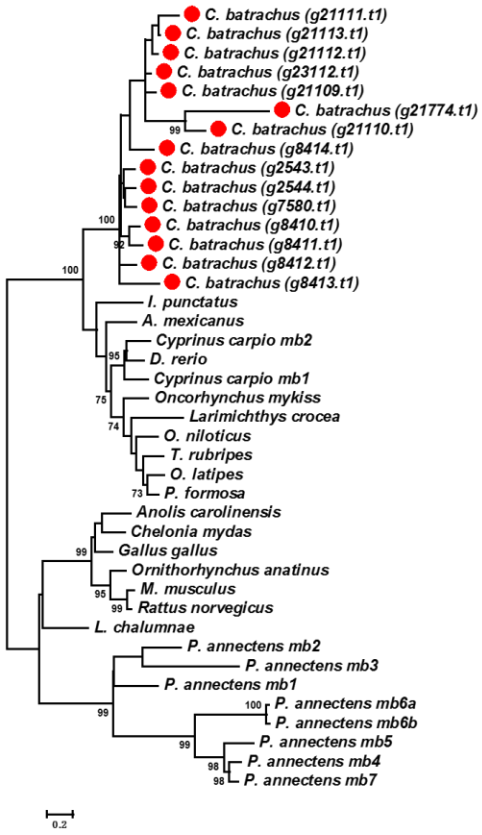


Figure 6. Phylogenetic tree of *mb* in vertebrates showing gene expansion of *mb* in the *C. batrachus* genome. The red solid circles represent the genes in the *C. batrachus* genome. Bootstrap support values (1,000 replications) are indicated on the branches.

(Figure 7). These 15 copies of the myoglobin were located on six scaffolds. I do not have information for their chromosomal locations, but their flanking genes were found on four different chromosomes in channel catfish, indicating that these fifteen genes of *C. batrachus* may be located on different chromosomes.

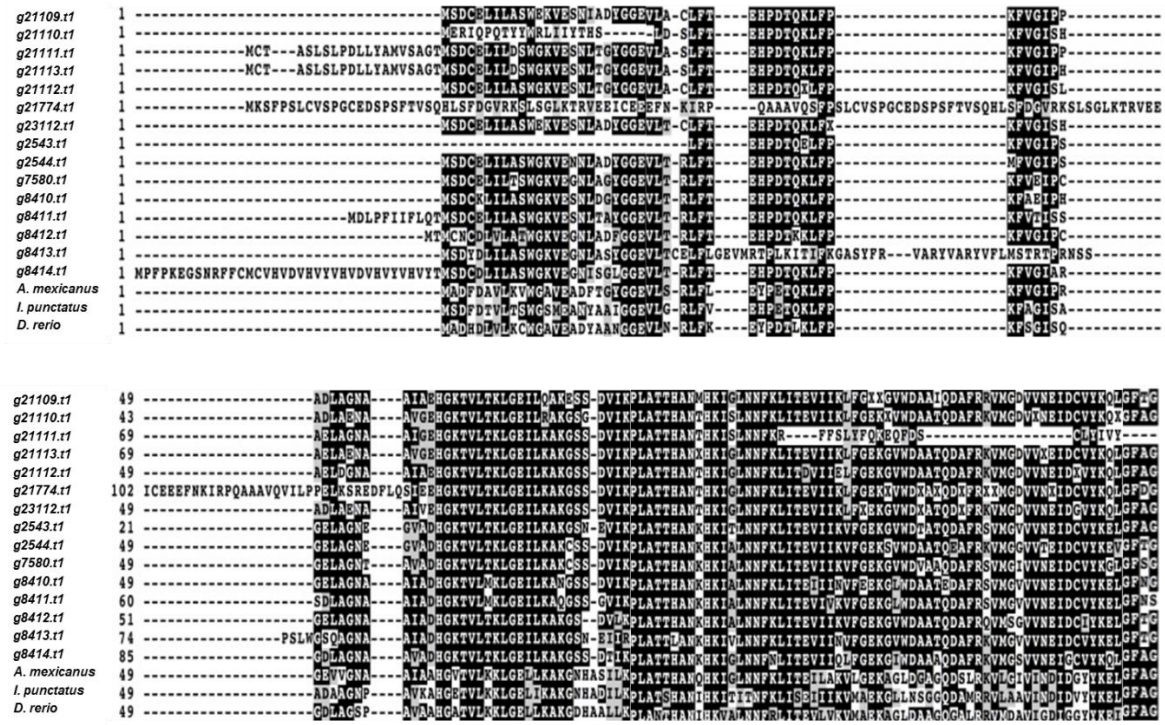


Figure 7. Multiple sequence alignment of myoglobin genes in the genomes of *C. batrachus*, *D. rerio*, *I. punctatus* and *A. mexicanus*.

The expansion of myoglobin genes in *C. batrachus* may be consistent with its frequent time spent in low-oxygen habitats and occasional terrestrial migration. Myoglobin, as a mobile oxygen carrier predominantly in skeletal and cardiac muscles, is able to store oxygen and facilitate the diffusion of oxygen (Millikan 1939; Wittenberg 1970). Millikan (Millikan 1937) reported that myoglobin maintained balance in periods of fluctuating oxygen supply and demand through rapid oxygenation and deoxygenation. Additionally, myoglobin maintains a steady level

of partial oxygenation to supply oxygen to the mitochondria during muscle contraction (Millikan 1939; Wittenberg, et al. 1975). Although myoglobin is not expanded in mammals, many studies indicated that it is regulated at the transcriptional level, with higher levels of expression in the skeletal muscle of hypoxia-tolerant animals such as deep diving and high-attitude mammals compared to surface and lowland relatives (Helbo and Fago 2012; Reynafarje 1962). In fish species, comparative studies have been conducted between *mb*-high sea raven (*Hemitripterus americanus*) and *mb*-low ocean pout (*Macrozoarces americanus*) (Driedzic, et al. 1982) and between *mb*-present icefsh (*Chionodraco rastrospinosus*) and *mb*-absent icefish (*Chaenocephalus aceratus*) (Acierno, et al. 1997). Both studies indicated that myoglobin plays a critical role in maintaining oxygen consumption in the heart and enhancing cardiac performance. In addition to those functions of oxygen storage and transport, myoglobin was also found to be involved in protecting mitochondrial respiration from nitric oxide inhibition (Brunori 2001) and in scavenging of reactive oxygen species (ROS) (Flögel, et al. 2004). Especially during the hypoxia and subsequent re-oxygenation period, the production of ROS increased significantly (Bickler and Buck 2007; Li and Jackson 2002). One pronounced example was found in common carp (*Cyprinus carpio*) that the additional myoglobin isoform *mb2* played a protective role against ROS in the brain (Fraser, et al. 2006; Helbo, et al. 2012). Similarly in lungfish, notable myoglobin expression in the brain was observed and the cell-level experiments also suggested its key role in protecting the tissues from ROS (Koch, et al. 2016).

Ora1 was also found to be significantly expanded in the *C. batrachus* genome, with fifteen copies while there is only a single copy in most teleost species (Saraiva and Korsching 2007). The 15 copies of *ora1* genes in *C. batrachus* were found on the same scaffold, suggesting tandem duplications, and these fifteen genes displayed high sequence similarities (Figure 8).

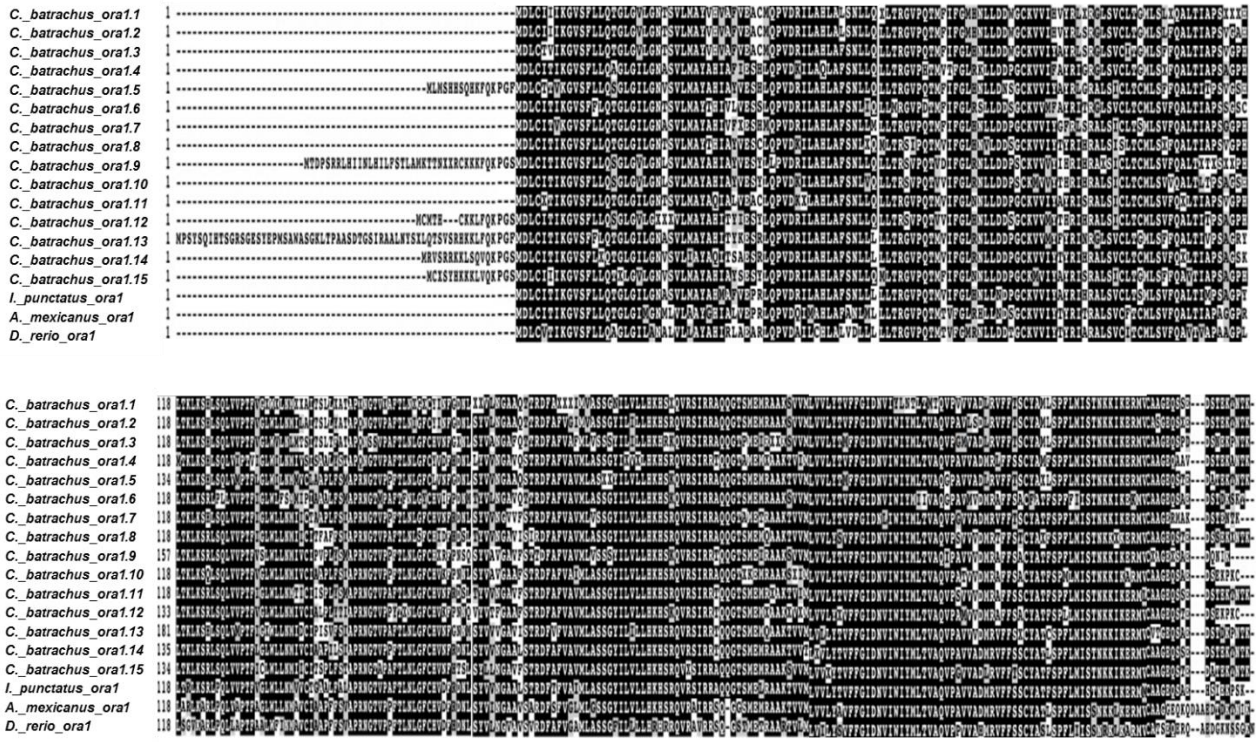


Figure 8. Multiple sequence alignment of *ora1* genes in the genomes of *C. batrachus, D. rerio, I. punctatus* and *A. mexicanus*.

Olfaction is an important sense for fish to recognize odorants due to their reduced visual ability and turbid environments. Unlike mammals possessing main olfactory epithelium (MOE) and a vomeronasal organ (VNO) to express different types of chemoreceptors, fish only have MOE (Bargmann 1997; Cao, et al. 1998; Mombaerts 2004; Yoshihara 2008). The separation of MOE and VNO in terrestrial vertebrates may have resulted evolutionarily from the segregation of distinct classes of neurons that were differentially positioned in the MOE of aquatic vertebrates (Cao, et al. 1998). Furthermore, the *ora* genes in fish species are homologs of the vomeronasal receptor 1 (*v1r*) in mammals (Pfister and Rodriguez 2005). Surprisingly, the *ora* genes are very conserved in fish species with very rare gene duplication events (Saraiva and Korsching 2007),

while mammalian genomes harbor hundreds of *v1r* genes (Young, et al. 2010). In most cases, fish species possess six *ora* genes with *ora1-ora2*, *ora3-ora4* and *ora5-ora6* forming three clades (Figure 9), suggesting a close relationship in each gene pair (Saraiva and Korsching 2007).

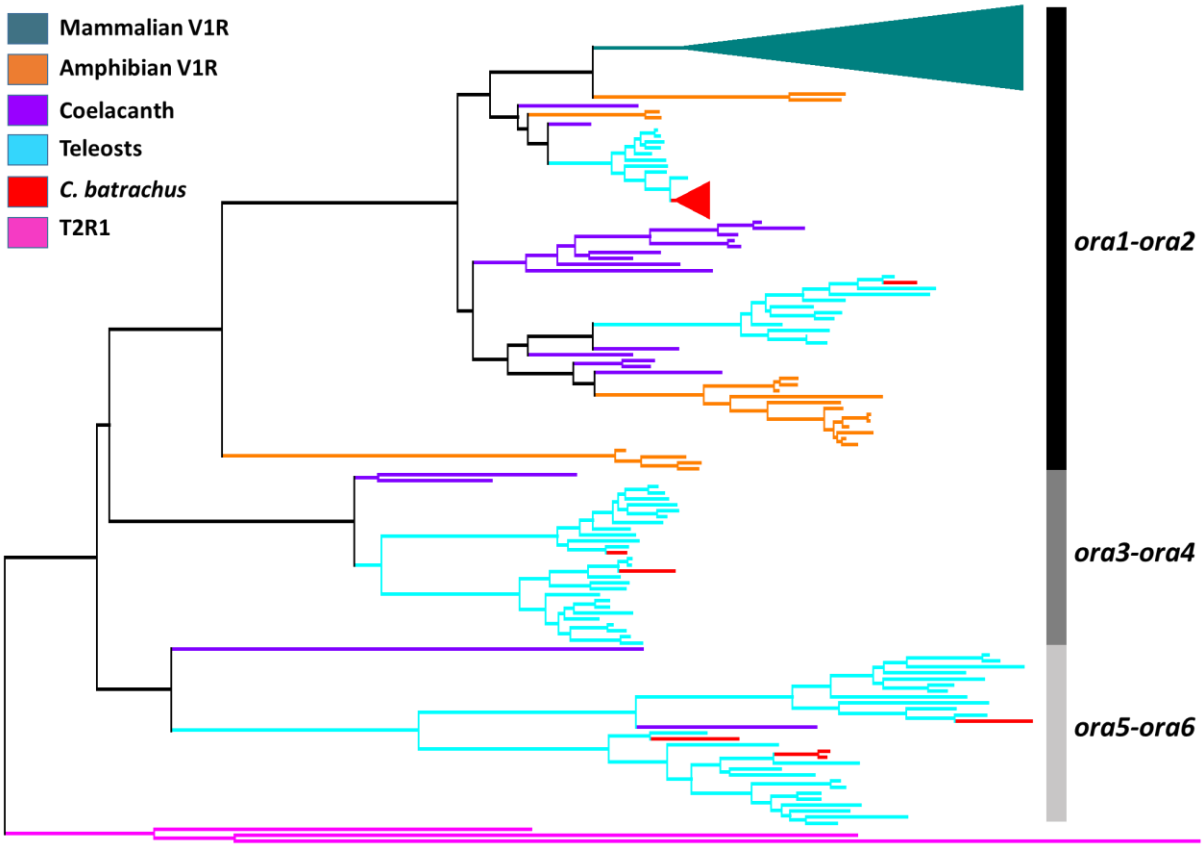


Figure 9. Phylogenetic tree of *ora* in vertebrates showing gene expansion of *ora1* in the *C. batrachus* genome with T2R1 as an outgroup. The three clades (*ora1-ora2*, *ora3-ora4* and *ora5-ora6*) formed from six members are indicated by vertical string. The dark green solid triangle represents the expansion of V1Rs in mammals. The red solid triangle represents the 15 copies in the *C. batrachus* genome. The term "teleosts" here is used to indicate the non-air-breathing fish species discussed in this study. The detailed phylogenetic tree with species names and sequence names is displayed in the appendix 2. V1R, vomeronasal type 1 receptor; T2R1, taste receptor of type 2 member 1.

In the *C. batrachus* genome, I identified all six *ora* genes but *ora1* was expanded with fifteen tandem copies. Interestingly, the expansions of *v1r* in mammals were also clustered as tandem duplications (Kurzweil, et al. 2009) and fell within *ora1-ora2* gene pair clade of teleost species (Figure 9), suggesting that the gaining of *ora3-ora6* genes in aquatic species might be due to the aquatic lifestyle (Saraiva and Korsching 2007). Coelacanth (*Latimeria chalumnae*), as an ancient lobe-finned fish which was thought to be evolutionarily close to tetrapods, not only possessed all the *ora* genes in the genome but also experienced an expansion in the *ora1-ora2* gene pair clade (Nikaido, et al. 2013; Syed and Korsching 2014) (Figure 9), which is similar to *C. batrachus*. These related observations may suggest that the expansion of *ora1* genes in *C. batrachus* could be associated with adaptation for transition from water to land, allowing recognition of airborne chemicals to help better locate food and the water sources.

In addition to *mb* and *ora1*, *sult6b1* was also found to be highly expanded in *C. batrachus* with 12 copies, as compared to 1-2 copies in non-air-breathing teleost fish (Figure 10). This gene encodes for a key enzyme in the process of detoxifying and eliminating xenobiotics. Aquatic habitats, as the ultimate sink, were loaded with contaminants affecting the health of aquatic animals (Van der Oost, et al. 2003). *C. batrachus*, as an air-breathing fish, not only suffers the same toxins in the water as the other aquatic victims, but also endures higher concentrations of toxic chemicals in the drying water bodies as well as from the land. To counteract the toxic effects of these xenobiotics, complex enzyme-based mechanisms usually need to be developed to detoxify and eliminate these chemical compounds. Sulfotransferases function by conjugation of a sulfate group with the target xenobiotics to increase their hydrophilicity for excretion (Mulder 2003). *C. batrachus* demonstrated an overall higher tolerance to three widely distributed xenobiotics than two other air-breathing fish species, *Heteropneustes fossilis* and *Channa*

punctatus (Farah, et al. 2004). Rainbow trout (*Oncorhynchus mykiss*) fry and Japanese flounder (*Paralichthys olivaceus*) also showed significantly increased expression of *sult6b1* after exposure to diesel and water accommodated fraction of crude oil, respectively, indicating that *sult6b1* does function in eliminating toxic chemicals in fish species (Mos, et al. 2008; Zhu, et al. 2016). Taken together, the expansion of *sult6b1* may play crucial roles in protecting *C. batrachus* from the deleterious effects of different xenobiotics from the aquatic and terrestrial environments.

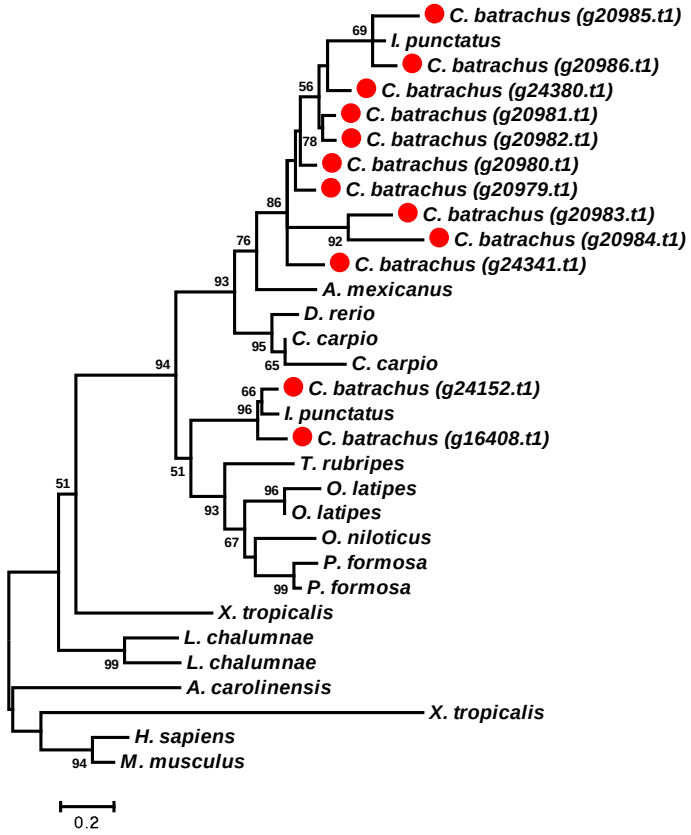


Figure 10. Phylogenetic tree of *sult6b1* in vertebrates showing gene expansion of *sult6b1* in the *C. batrachus* genome. The red solid circles represent the genes in the *C. batrachus* genome. Bootstrap support values (1,000 replications) are indicated on the branches.

Expression of significantly expanded gene families

The expression of the various copies of the *mb*, *sult6b1* and *ora1* genes in *C. batrachus* was analyzed using RNA-Seq datasets (Table 10). All 15 copies of myoglobin gene were

Table 10. Expression of myoglobin and sulfotransferase 6b1 genes (FPKM) in the air-breathing organ, gill, brain, head kidney and liver of *C. batrachus* as determined by analysis of RNA-Seq datasets. RNA-Seq datasets from air-breathing organ and gill were generated from this project; those from brain (SRR2057993), head kidney (SRR2057995), and liver (SRR2052655) were downloaded from NCBI. Each shaded area contained tandem duplicated genes. FPKM, Fragments Per Kilobase per Million mapped fragments.

Gene name	Gene ID	Air-breathing organ	Gill	Brain	Head kidney	Liver
Myoglobin	g21109.t1	0	0	0	2.91	2.84
	g21110.t1	0	0	8.48	0.85	4.14
	g21111.t1	0.25	0.18	3.14	1.57	1.53
	g21112.t1	0.59	0.05	1.45	1.55	0
	g21113.t1	0.09	0	4.48	9.88	2.92
	g21774.t1	4.52	5.07	0	1.02	0
	g23112.t1	0.46	0.76	0	0.58	0
	g2543.t1	0	0	26.84	25.61	5.11
	g2544.t1	0	0	15.71	12.23	3.31
	g7580.t1	0	0	0	2.32	0
	g8410.t1	0	0	17.4	23.41	3.31
	g8411.t1	0	0	0	0.59	0
	g8412.t1	0	0.56	16.23	13.52	3.27
	g8413.t1	0	0.16	0	0	0
	g8414.t1	6.18	139.88	838.66	643.35	145
Sulfotransferase 6b1	g16408.t1	0.2	1.16	0	0	0
	g20979.t1	2.78	1.22	3.96	1.94	1.95
	g20980.t1	23.9	28.5	13.36	18.93	2.09
	g20981.t1	3.42	3.79	6.16	3.34	1.28
	g20982.t1	2.91	2.42	1.57	1.62	1.28
	g20983.t1	0	0.58	0	0	0
	g20984.t1	0.54	0.53	0.57	1.43	1.75
	g20985.t1	9.37	16.57	5.6	4.94	0.26
	g20986.t1	24.61	14.81	89.49	44.81	0
	g24152.t1	0.13	0.53	0	0	0
	g24341.t1	0	0.06	0	0	0
	g24380.t1	0	0.31	0	0	15.45

expressed, but in different tissues and at different levels. The brain had an overall high expression, consistent with the results in common carp and lungfish (Fraser, et al. 2006; Koch, et al. 2016). The expression levels in the air-breathing organ are of particular interest: two copies of the myoglobin gene, g21774.t1 and g8414.t1, were expressed at high levels, at least 10 times higher than other copies. In all the other tissues, only one copy, g8414.t1, was expressed at very high levels. Interestingly, syntenic analysis indicated that g8414.t1 is orthologous to the single copy gene of myoglobin in other fish species (Figure 11). Its high expression in all tissues

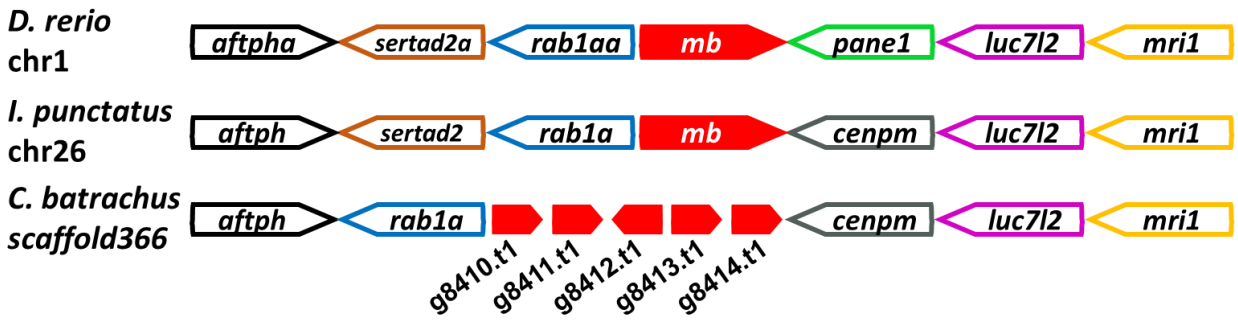


Figure 11. Syntenic analysis of *mb* gene using genomic contexts of *D. rerio*, *I. punctatus* and *C. batrachus*. The g8410.t1, g8411.t1, g8412.t1, g8413.t1 and g8414.t1 represent the myoglobin genes on the scaffold366 in the *C. batrachus* genome. *aftph*, aftiphilin; *aftpha*, aftiphilin a; *sertad2*, SERTA domain containing 2; *sertad2a*, SERTA domain containing 2a; *rab1a*, RAB1A, member RAS oncogene family; *rab1aa*, RAB1A, member RAS oncogene family a; *pane1*, proliferation associated nuclear element; *cenpm*, centromere protein M; *luc7l2*, putative RNA-binding protein Luc7-like 2; *mri1*, methylthioribose-1-phosphate isomerase 1.

suggested that its function is conserved in evolution. For the *sult6b1* gene, all 12 copies were expressed, but with some tissue specificity. For instance, g20983.t1 and g24341.t1 were only expressed in the gill; g24152.t1 and g16408.t1 were only expressed in the gill and air-breathing organ; and g24380.t1 was expressed only in the gill and liver. All 12 copies were expressed in

the gill, but the highest expressed copies were g20980.t1, g20985.t1 and g20986.t1. Notably, all the highly expressed copies were duplicates on the same scaffold (Table 10). The o*ra1* gene was known to be tissue specific and expressed only in the olfactory epithelium in other fish species (Saraiva and Korsching 2007). In regards to the five tissues examined with the RNA-Seq, no transcripts of *ora1* were detected, consistent with the tissue-specific expression of this gene only in olfactory epithelium in other fish species (Saraiva and Korsching 2007).

Comparative transcriptomic analysis between gill and the air-breathing organ

It has been suggested that air breathing was evolved as an adaptation for fish to cope with hypoxic conditions and consequently it was considered as an essential first step to terrestrial habitation in the evolution of vertebrates (Inger 1957; Janis and FARMER 1999; Martin 2014; Randall 1981; Romer 1967). To understand the transcriptomic differences between the gill and the air-breathing organ, I sequenced the transcriptomes of the gill and the air-breathing organ (Table 11) and determined the differentially expressed genes (DEG, Appendix 3). A total of 813

Table 11. Summary of transcriptome sequencing data.

Tissue	Read length (bp)	# of raw reads	Raw data (Gb)	# of trimmed reads
Gill	125	234,620,438	29.3	225,509,720
Air-breathing organ	125	231,374,100	28.9	221,641,474

differentially expressed genes were identified between the gill and the air-breathing organ, of which 335 and 478 genes were up-regulated and down-regulated in the air-breathing organ, respectively. These results were validated by determining the expression levels of a fraction of these genes using qRT-PCR (Figure 12). All the DEG were subjected to GO and pathway

enrichment analyses, and the significantly over-represented GO terms and pathways in the DEG were listed in Table 12 and Table 13, respectively. As summarized in Table 14, a total of 51 genes belonging to five functional groups were highly and differentially expressed between the gill and the air-breathing organ.

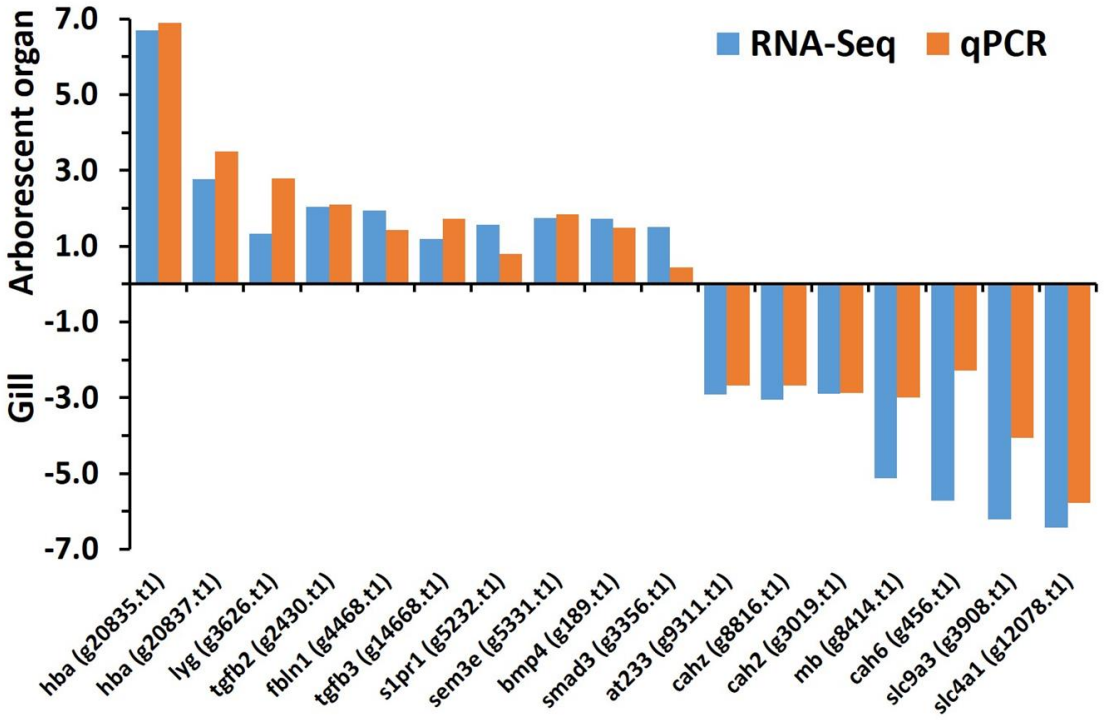


Figure 12. Comparison of relative fold changes between air-breathing organ and gill after normalization to 28S rRNA using RNA-Seq datasets and qRT-PCR. The absolute values of relative fold changes were transformed by log2(ratio). The Pearson correlation coefficient between the two results is 0.8944 before transformation.

Table 12. GO terms significantly enriched in the differentially expressed genes comparing the transcriptome of the air-breathing organ with that of the gill.

	GO-ID	GO Term	Category	FDR
Up-regulated	GO:0005833	hemoglobin complex	CC	3.34E-07
	GO:0005344	oxygen transporter activity	MF	3.70E-05
	GO:0015671	oxygen transport	BP	3.70E-05
	GO:0019825	oxygen binding	MF	3.70E-05
	GO:0001667	ameboidal-type cell migration	BP	1.33E-04
	GO:0005160	transforming growth factor beta receptor binding	MF	1.66E-04
	GO:0005615	extracellular space	CC	2.64E-04
	GO:0048703	embryonic viscerocranium morphogenesis	BP	3.73E-04
	GO:0005578	proteinaceous extracellular matrix	CC	4.78E-04
	GO:0005506	iron ion binding	MF	5.02E-04
	GO:0060395	SMAD protein signal transduction	BP	9.43E-04
	GO:0019835	cytolysis	BP	1.86E-03
	GO:0010862	positive regulation of pathway-restricted SMAD protein phosphorylation	BP	2.90E-03
	GO:0008083	growth factor activity	MF	3.45E-03
	GO:0031101	fin regeneration	BP	5.27E-03
	GO:0005540	hyaluronic acid binding	MF	8.22E-03
	GO:0034340	response to type I interferon	BP	8.22E-03
	GO:0016049	cell growth	BP	8.84E-03
	GO:0005520	insulin-like growth factor binding	MF	9.68E-03
	GO:0001756	somitogenesis	BP	1.18E-02
	GO:0035141	medial fin morphogenesis	BP	1.34E-02
	GO:0070700	BMP receptor binding	MF	1.34E-02
	GO:0035457	cellular response to interferon-alpha	BP	1.34E-02
	GO:0005125	cytokine activity	MF	1.36E-02
	GO:0030509	BMP signaling pathway	BP	1.42E-02
	GO:0020037	heme binding	MF	1.69E-02
	GO:0009617	response to bacterium	BP	2.02E-02
	GO:0042127	regulation of cell proliferation	BP	2.25E-02
	GO:0051093	negative regulation of developmental process	BP	2.46E-02
	GO:0010941	regulation of cell death	BP	2.62E-02
	GO:0090594	inflammatory response to wounding	BP	2.74E-02
	GO:0006024	glycosaminoglycan biosynthetic process	BP	2.85E-02
	GO:0048762	mesenchymal cell differentiation	BP	2.88E-02
	GO:0002040	sprouting angiogenesis	BP	2.92E-02
	GO:0070290	N-acylphosphatidylethanolamine-specific phospholipase D activity	MF	3.08E-02
	GO:0008466	glycogenin glucosyltransferase activity	MF	3.08E-02
	GO:0050840	extracellular matrix binding	MF	3.67E-02
	GO:0045765	regulation of angiogenesis	BP	3.75E-02

	GO:0051241	negative regulation of multicellular organismal process	BP	4.18E-02
	GO:0030198	extracellular matrix organization	BP	4.75E-02
Down-regulated	GO:0005882	intermediate filament	CC	1.49E-07
	GO:0005861	troponin complex	CC	1.70E-06
	GO:0031444	slow-twitch skeletal muscle fiber contraction	BP	3.69E-05
	GO:0030240	skeletal muscle thin filament assembly	BP	8.88E-05
	GO:0006096	glycolytic process	BP	1.38E-04
	GO:0071688	striated muscle myosin thick filament assembly	BP	2.72E-04
	GO:0032982	myosin filament	CC	4.63E-04
	GO:0090131	mesenchyme migration	BP	7.38E-04
	GO:0004089	carbonate dehydratase activity	MF	1.46E-03
	GO:0045214	sarcomere organization	BP	1.59E-03
	GO:0006730	one-carbon metabolic process	BP	2.71E-03
	GO:0048037	cofactor binding	MF	2.97E-03
	GO:0003272	endocardial cushion formation	BP	6.51E-03
	GO:0005198	structural molecule activity	MF	7.10E-03
	GO:0015662	ATPase activity, coupled to transmembrane movement of ions, phosphorylative mechanism	MF	8.43E-03
	GO:0030018	Z disc	CC	9.09E-03
	GO:0051373	FATZ binding	MF	1.47E-02
	GO:0014706	striated muscle tissue development	BP	1.93E-02
	GO:0030175	filopodium	CC	1.95E-02
	GO:0048747	muscle fiber development	BP	1.95E-02
	GO:0060538	skeletal muscle organ development	BP	2.06E-02
	GO:0009957	epidermal cell fate specification	BP	2.62E-02
	GO:0046539	histamine N-methyltransferase activity	MF	2.62E-02
	GO:0055067	monovalent inorganic cation homeostasis	BP	2.83E-02
	GO:0051015	actin filament binding	MF	3.14E-02
	GO:0036302	atrioventricular canal development	BP	4.03E-02
	GO:0006635	fatty acid beta-oxidation	BP	4.42E-02

Table 13. Pathways significantly enriched in the differentially expressed genes comparing the transcriptome of the air-breathing organs with that of the gill.

	Pathway name	*P*-value
Up-regulated	Formation of the cornified envelope	6.73E-07
	Molecules associated with elastic fibres	2.30E-04
	Keratinization	4.94E-04
	Elastic fibre formation	6.25E-04
	The activation of arylsulfatases	8.20E-04
	Synthesis of 15-eicosatetraenoic acid derivatives	9.56E-04
	Synthesis of 12-eicosatetraenoic acid derivatives	9.56E-04
	Syndecan interactions	1.77E-03
	Signaling by FGFR1 amplification mutants	1.88E-03
	Hydrolysis of LPC	2.16E-03
	Erythrocytes take up oxygen and release carbon dioxide	2.16E-03
	Arachidonic acid metabolism	2.62E-03
	Extracellular matrix organization	2.81E-03
	Interleukin-4 and 13 signaling	3.00E-03
	Phospholipase C-mediated cascade; FGFR2	4.59E-03
	Acyl chain remodelling of PS	5.40E-03
	FGFR2 ligand binding and activation	6.30E-03
	Signaling by FGFR2 IIIa TM	6.30E-03
	O2/CO2 exchange in erythrocytes	6.67E-03
	Erythrocytes take up carbon dioxide and release oxygen	6.67E-03
	TP53 Regulates Transcription of Death Receptors and Ligands	6.67E-03
	Constitutive Signaling by Aberrant PI3K in Cancer	9.51E-03
	PI-3K cascade: FGFR2	9.56E-03
Down-regulated	Keratinization	1.11E-16
	Formation of the cornified envelope	1.11E-16
	Striated Muscle Contraction	9.81E-13
	Muscle contraction	1.45E-06
	Reversible hydration of carbon dioxide	1.72E-05
	Apoptotic cleavage of cell adhesion proteins	1.36E-04
	Apoptotic cleavage of cellular proteins	3.04E-04
	Erythrocytes take up oxygen and release carbon dioxide	4.53E-04
	Apoptotic execution phase	5.35E-04
	Glycolysis	8.16E-04
	O2/CO2 exchange in erythrocytes	2.00E-03
	Erythrocytes take up carbon dioxide and release oxygen	2.00E-03
	Branched-chain amino acid catabolism	3.57E-03
	Pyruvate metabolism and Citric Acid (TCA) cycle	5.47E-03
	Rhesus glycoproteins mediate ammonium transport	7.75E-03
	Ion homeostasis	1.00E-02

Table 14. Five functional categories differentially expressed between the gill and the air-breathing organ in *C. batrachus*. Negative signs indicate those genes expressed differentially but at higher levels in the gill.

Category name	Gene name	Gene ID	Fold change	FDR adjusted *P*-value
Acid-base balance	Solute carrier family 4 member 1	g12078.t1	-85.83	0
	Carbonic anhydrase	g8816.t1	-8.27	0
	Carbonic anhydrase 2	g3019.t1	-7.48	1.60E-07
	Carbonic anhydrase 4	g7074.t1	-2.76	7.21E-07
	Carbonic anhydrase 9	g411.t1	-4.20	5.37E-05
	Carbonic anhydrase 6	g456.t1	-52.99	7.39E-04
Ion homeostasis	Sodium/potassium-transporting ATPase subunit alpha-1	g15272.t1	-11.77	0
	Sarcoplasmic/endoplasmic reticulum calcium ATPase 1	g16367.t1	-2159.02	0
	Sarcoplasmic/endoplasmic reticulum calcium ATPase 1	g5688.t1	-56.41	0
	Sarcoplasmic/endoplasmic reticulum calcium ATPase 2	g9634.t1	-261.57	0
	Sodium potassium-transporting ATPase subunit beta-233	g9311.t1	-7.57	0
	Adenosylhomocysteinase 3	g1970.t1	-54.65	2.72E-10
	Adenosylhomocysteinase 3	g1011.t1	-5.35	1.19E-08
	Calsequestrin-1	g3495.t1	-20.23	5.23E-04
	Calsequestrin-2	g9400.t1	-27.09	3.67E-02
	Pendrin	g2970.t1	-29.11	0
	Sodium/hydrogen exchanger 3	g3908.t1	-74.31	0
Elastic fiber formation	Transforming growth factor beta-2	g2430.t1	4.10	8.22E-03
	Fibulin-1	g4468.t1	3.82	9.66E-08
	Transforming growth factor beta-3	g14668.t1	2.27	2.68E-02
	Fibulin-2	g20129.t1	2.95	5.73E-03
	Bone morphogenetic protein 4	g189.t1	3.31	8.79E-06
	Fibronectin	g6260.t1	2.46	0
	Fibronectin	g12205.t1	4.89	9.01E-03
	Latent-transforming growth factor beta-binding protein 3	g14988.t1	2.11	8.22E-03
Oxygen binding and transport	Hemoglobin subunit alpha	g20835.t1	103.38	0
	Hemoglobin subunit beta	g20836.t1	6.36	3.55E-11
	Hemoglobin subunit alpha	g20837.t1	6.82	0
	Hemoglobin subunit beta	g20838.t1	5.80	0
	Hemoglobin subunit beta	g21168.t1	5.37	0
	Hemoglobin subunit alpha	g21169.t1	4.41	0
	Hemoglobin subunit beta	g21170.t1	3.35	0
	Hemoglobin subunit alpha	g21171.t1	3.68	0
Angiogenesis	Sphingosine 1-phosphate receptor 1	g5232.t1	2.98	7.89E-03
	Fibronectin	g6260.t1	2.46	8.79E-06
	Semaphorin-3E	g5331.t1	3.35	6.86E-05
	C-X-C chemokine receptor type 4	g23603.t1	4.53	0
	Neuropilin-1a	g3757.t1	2.62	5.00E-02
	Cadherin-5	g21848.t1	2.24	7.36E-03
	BMP-binding endothelial regulator protein	g7031.t1	5.10	2.71E-02
	Bone morphogenetic protein 4	g189.t1	3.31	5.73E-03
	Bone morphogenetic protein 5	g11279.t1	3.12	2.76E-03
	Bone morphogenetic protein 8A	g14806.t1	2.67	9.26E-03
	Bone morphogenetic protein receptor type-1A	g14843.t1	4.62	9.45E-05
	Transforming growth factor beta-2	g2430.t1	4.10	8.22E-03
	Transforming growth factor beta-3	g14668.t1	2.27	2.68E-02
	Rho-related GTP-binding protein RhoB	g5763.t1	2.49	1.23E-09
	Thrombospondin-1	g5480.t1	2.44	3.33E-02
	Mothers against decapentaplegic homolog 6	g4737.t1	2.79	6.93E-13
	Mothers against decapentaplegic homolog 6	g4739.t1	2.90	3.88E-11
	Mothers against decapentaplegic homolog 3	g3356.t1	2.83	4.28E-03

Of the five groups of DEG genes, two groups were highly expressed in the gill: six genes for acid-base balance, and 11 genes for ion homeostasis (Table 14), indicating that the gill plays critical roles in acid-base and ion regulation. Acid-base regulation in vertebrates is coupled to carbon dioxide (CO_2) excretion through the reversible hydration/dehydration reactions of CO_2 and the acid-base equivalents H^+ and HCO_3^- by carbonic anhydrase (Table 14). Acid-base

regulation is always linked to ion regulation because acid-base compensation depends on the transfer of H^+ and HCO_3^- in exchange for Na^+ and Cl^- across the gill, respectively (Gilmour and Perry 2009; Henry and Heming 1998; Randall and Val 1995).

Three groups of genes were highly expressed in the air-breathing organ: eight "Elastic fiber formation" genes, 8 hemoglobin genes, and 18 genes involved in angiogenesis (Table 14). While elastic fibers are important structural components of the arborescent organ (Ahmed, et al. 2008; Ikpegbu, et al. 2013), hemoglobin genes and genes involved in angiogenesis apparently provide functional basis for *C. batrachus* to cope with low oxygen in the terrestrial environment. The eight hemoglobin genes (four alpha subunit genes and four beta subunit genes) were found to be dramatically up-regulated in the air-breathing organ, as compared to the gill that is primarily an aquatic respiratory organ (Table 14), demonstrating that the air-breathing organ is much more committed to the respiratory processes for oxygen transport. It appears that the strategy during evolution of *C. batrachus* to adapt to the transition from aquatic to terrestrial environment was through the coupling of high expression of hemoglobin genes for oxygen transport with expansion of myoglobin genes for oxygen storage.

A large number of genes (18 genes) involved in angiogenesis were found differentially expressed in the air-breathing organ (Table 14), suggesting the structural architecture of the air-breathing organ possesses enhanced blood vessels for the air-breathing function. Angiogenesis plays a critical role in respiratory function for accessory air-breathing organs of fishes (Jiang, et al. 2016; Luo, et al. 2016; Martin 2014). Also, the air-breathing organ of *C. batrachus* is highly vascularized on the surface, and the capillaries extensively bulge out onto the surface to facilitate gas exchange between blood and atmospheric air (Chandra and Banerjee 2003; Munshi 1961).

Consequently, angiogenesis may be one additional component for the air-breathing organ to maintain high efficiency of air exchange.

CONCLUSION

The walking catfish is an aquatic species, but can move about on land. Apparently, the largest evolutionary challenge is how to utilize atmospheric oxygen without a lung. As such, it is a remarkable model to understand the transition from the aquatic to the terrestrial environment, and the adaptation to terrestrial life. Through whole genome sequencing analysis, I did not find any specific genes that were present in this air-breathing fish, but absent in non-air-breathing fishes. However, significantly higher levels of gene expansions were found in the *C. batrachus* genome, mostly in tandem fashions. Of particular interest is the expansion of the oxygen storage protein myoglobin gene, with 15 copies, while the non-air-breathing fishes harbor only one to two copies. Lungfish was found to harbor seven copies of myoglobin gene, and this expansion of myoglobin genes was believed to be crucially important for the adaptation of lungfish to survive hypoxic periods (Koch, et al. 2016). Therefore, it is likely that the expansion of myoglobin genes may be a common mechanism for the water-to-land transition. Additionally, olfactory receptor related to class A G protein-coupled receptor 1 and the sulfotransferase 6b1 genes were found to be highly expanded with the former being related to olfactory sense and the latter to provide resistance to xenobiotics.

The coupling of enhanced oxygen transport and oxygen storage may be essential for water-to-land transition. The hemoglobin genes were found to be expressed much higher in the air-breathing organ of the *C. batrachus* than in its gill. While the hemoglobin genes are also highly duplicated, they are not more expanded in the air-breathing *C. batrachus* than in the non-air-breathing fishes. Instead, the regulation appeared to be at the transcriptional level where hemoglobin RNAs were transcribed many folds higher in the air-breathing organ, ensuring the greater capacity for oxygen transport. In addition, many genes involved in angiogenesis were

found to be expressed at much higher levels in the air-breathing organ than in the gill of *C. batrachus*, providing structural basis for increased vessel systems for air exchange. Taken together, the evolution for the water-to-land transition seemed to involve mostly oxygen transport and storage genes at both the genomic and transcriptomic levels.

REFERENCES

Acierno R, Agnisola C, Tota B, Sidell B 1997. Myoglobin enhances cardiac performance in antarctic icefish species that express the protein. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology 273: R100-R106.

Ahmed A, Mohamed K, Ahmed S-A, Masoud F 2008. Anatomical, light and scanning electron microscopic studies on the air breathing dendretic organ of the sharp tooth catfish (*Clarias gariepinus*). J Vet Anat 1: 29-37.

Andrews S 2010. FastQC: A quality control tool for high throughput sequence data. Reference Source.

Atkinson B. 2012. Changes in eukaryotic gene expression in response to environmental stress: Elsevier.

Bao Z, Eddy SR 2002. Automated *de novo* identification of repeat sequence families in sequenced genomes. Genome Res 12: 1269-1276.

Bargmann CI 1997. Olfactory receptors, vomeronasal receptors, and the organization of olfactory information. Cell 90: 585-587.

Bickler PE, Buck LT 2007. Hypoxia tolerance in reptiles, amphibians, and fishes: life with variable oxygen availability. Annu. Rev. Physiol. 69: 145-170.

Biscotti MA, Gerdol M, Canapa A, Forconi M, Olmo E, Pallavicini A, Barucca M, Schartl M 2016. The Lungfish Transcriptome: A Glimpse into Molecular Evolution Events at the Transition from Water to Land. Sci Rep 6: 21571. doi: 10.1038/srep21571

Boeckmann B, Bairoch A, Apweiler R, Blatter M-C, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Research 31: 365-370.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27: 578-579. doi: 10.1093/bioinformatics/btq683

Boetzer M, Pirovano W 2012. Toward almost closed genomes with GapFiller. Genome Biology 13: R56.

Bolger AM, Lohse M, Usadel B 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics: btu170.

Borowiec ML 2016. AMAS: a fast tool for alignment manipulation and computing of summary statistics. PeerJ 4: e1660.

Brauner C, Baker D. 2009. Patterns of acid–base regulation during exposure to hypercarbia in fishes. In. Cardio-respiratory control in vertebrates: Springer. p. 43-63.

Brauner C, Matey V, Wilson J, Bernier N, Val A 2004a. Transition in organ function during the evolution of air-breathing; insights from *Arapaima gigas*, an obligate air-breathing teleost from the Amazon. Journal of Experimental Biology 207: 1433-1438.

Brauner C, Wang T, Wang Y, Richards J, Gonzalez R, Bernier N, Xi W, Patrick M, Val A 2004b. Limited extracellular but complete intracellular acid-base regulation during short-term environmental hypercapnia in the armoured catfish, *Liposarcus pardalis*. Journal of Experimental Biology 207: 3381-3390.

Brunori M 2001. Nitric oxide moves myoglobin centre stage. Trends in biochemical sciences 26: 209-210.

Cameron JN 1989. Acid-base homeostasis: past and present perspectives. Physiological zoology 62: 845-865.

Cao Y, Oh BC, Stryer L 1998. Cloning and localization of two multigene receptor families in goldfish olfactory epithelium. Proceedings of the National Academy of Sciences 95: 11987-11992.

Chandra S, Banerjee TK 2003. Histopathological analysis of the respiratory organs of the air-breathing catfish *Clarias batrachus* (Linn.) exposed to the air. Acta Zoologica Taiwanica 14: 45-64.

Chew S, Ip Y 2014. Excretory nitrogen metabolism and defence against ammonia toxicity in air-breathing fishes. Journal of Fish Biology 84: 603-638.

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M 2005. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21: 3674-3676.

Courtenay Jr W, Hensley D, Taylor J, McCann J 1986. Distribution of exotic fishes in North America.

Courtenay WR, Sahlman HF, Miley WW, Herrema DJ 1974. Exotic fishes in fresh and brackish waters of Florida. Biological Conservation 6: 292-302.

Courtenay WR, Stauffer JR 1990. The introduced fish problem and the aquarium fish industry. Journal of the World Aquaculture Society 21: 145-159.

Dahanukar N, Raut R, Bhat A 2004. Distribution, endemism and threat status of freshwater fishes in the Western Ghats of India. Journal of biogeography 31: 123-136.

Damsgaard C, Findorf I, Helbo S, Kocagoz Y, Buchanan R, Weber RE, Fago A, Bayley M, Wang T 2014. High blood oxygen affinity in the air-breathing swamp eel *Monopterus albus*. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology 178: 102-108.

Darriba D, Taboada GL, Doallo R, Posada D 2011. ProtTest 3: fast selection of best-fit models of protein evolution. Bioinformatics 27: 1164-1165.

Das B 1928. The bionomics of certain air-breathing fishes of India, together with an account of the development of their air-breathing organs. Philosophical Transactions of the Royal Society of London. Series B, Containing Papers of a Biological Character 216: 183-219.

De Nadal E, Ammerer G, Posas F 2011. Controlling gene expression in response to stress. Nature Reviews Genetics 12: 833-845.

Debnath S editor. International Conference on Asia Agriculture and Animal. IPCBEE (13), Singapore. 2011.

Delcher AL, Salzberg SL, Phillippy AM 2003. Using MUMmer to identify similar regions in large sequence sets. Current protocols in bioinformatics: 10.13. 11-10.13. 18.

Driedzic WR, Stewart JM, Scott DL 1982. The protective effect of myoglobin during hypoxic perfusion of isolated fish hearts. Journal of molecular and cellular cardiology 14: 673-677.

Edgar RC 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792-1797.

Emms D, Kelly S 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. Genome Biology 16: 157.

Evans DH, Piermarini PM, Choe KP 2005. The multifunctional fish gill: dominant site of gas exchange, osmoregulation, acid-base regulation, and excretion of nitrogenous waste. Physiological Reviews 85: 97-177.

Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, Jassal B, Jupe S, Korninger F, McKay S 2016. The reactome pathway knowledgebase. Nucleic Acids Research 44: D481-D487.

Farah MA, Ateeq B, Ali MN, Sabir R, Ahmad W 2004. Studies on lethal concentrations and toxicity stress of some xenobiotics on aquatic organisms. Chemosphere 55: 257-265.

Farrell AP. 2011. Encyclopedia of fish physiology: from genome to environment: Academic Press.

Flögel U, Gödecke A, Klotz L-O, Schrader J 2004. Role of myoglobin in the antioxidant defense of the heart. The FASEB Journal 18: 1156-1158.

Fraser J, de Mello LV, Ward D, Rees HH, Williams DR, Fang Y, Brass A, Gracey AY, Cossins AR 2006. Hypoxia-inducible myoglobin expression in nonmuscle tissues. Proc Natl Acad Sci U S A 103: 2977-2981.

Fuchs C, Burmester T, Hankeln T 2006. The amphibian globin gene repertoire as revealed by the Xenopus genome. Cytogenetic and genome research 112: 296-306.

Gilmour K, Perry S 2009. Carbonic anhydrase and acid–base regulation in fish. Journal of Experimental Biology 212: 1647-1661.

Gnerre S, MacCallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB 2011. High-quality draft assemblies of mammalian genomes from massively parallel sequence data. Proceedings of the National Academy of Sciences 108: 1513-1518. doi: 10.1073/pnas.1017351108

Gojobori T 1983. Codon substitution in evolution and the" saturation" of synonymous changes. Genetics 105: 1011-1027.

Gonçalves AF, Castro LFC, Pereira-Wilson C, Coimbra J, Wilson JM 2007. Is there a compromise between nutrient uptake and gas exchange in the gut of *Misgurnus anguillicaudatus*, an intestinal air-breathing fish? Comparative Biochemistry and Physiology Part D: Genomics and Proteomics 2: 345-355.

Goss GG, Perry SF, Fryer JN, Laurent P 1998. Gill morphology and acid-base regulation in freshwater fishes. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology 119: 107-115.

Graham J 2011. AIR-BREATHING FISHES| The Biology, Diversity, and Natural History of Air-Breathing Fishes: An Introduction.

Graham JB. 1997. Air-breathing fishes: evolution, diversity, and adaptation: Academic Press.

Graham JB, Lee HJ 2004. Breathing air in air: in what ways might extant amphibious fish biology relate to prevailing concepts about early tetrapods, the evolution of vertebrate air breathing, and the vertebrate land transition? Physiological and Biochemical Zoology 77: 720-731.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M 2013. *De novo* transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat Protoc 8: 1494-1512.

Han MV, Thomas GW, Lugo-Martinez J, Hahn MW 2013. Estimating gene gain and loss rates in the presence of error in genome assembly and annotation using CAFE 3. Molecular Biology and Evolution 30: 1987-1997.

HARRIS VA editor. Proceedings of the Zoological Society of London. 1960.

Heisler N 1984. 6 Acid-Base Regulation in Fishes. Fish physiology 10: 315-401.

Heisler N 1982. Intracellular and extracellular acid-base regulation in the tropical fresh-water teleost fish *Synbranchus marmoratus* in response to the transition from water breathing to air breathing. Journal of Experimental Biology 99: 9-28.

Helbo S, Dewilde S, Williams DR, Berghmans H, Berenbrink M, Cossins AR, Fago A 2012. Functional differentiation of myoglobin isoforms in hypoxia-tolerant carp indicates tissue-specific protective roles. American Journal of Physiology-Regulatory, Integrative and Comparative Physiology 302: R693-R701.

Helbo S, Fago A 2012. Functional properties of myoglobins from five whale species with different diving capacities. Journal of Experimental Biology 215: 3403-3410.

Henry RP, Heming TA 1998. 3 Carbonic Anhydrase and Respiratory Gas Exchange.

Hinegardner R, Rosen DE 1972. Cellular DNA content and the evolution of teleostean fishes. American Naturalist: 621-644.

Hochachka P, Moon T, Bailey J, Hulbert W 1978. The osteoglossid kidney: correlations of structure, function, and metabolism with transition to air breathing. Canadian Journal of Zoology 56: 820-832.

Hoffmann FG, Opazo JC, Storz JF 2011. Differential loss and retention of cytoglobin, myoglobin, and globin-E during the radiation of vertebrates. Genome Biology and Evolution 3: 588-600.

Huang S, Cao X, Tian X 2016. Transcriptomic Analysis of Compromise Between Air-Breathing and Nutrient Uptake of Posterior Intestine in Loach (*Misgurnus anguillicaudatus*), an Air-Breathing Fish. Marine Biotechnology 18: 521-533.

Ikpegbu E, Nlebedum U, Nnadozie O, Agbakwuru O 2013. Histological observations on the dendretic organ of the farmed adult African catfish (*Clarias gariepinus*) from Eastern Nigeria. Journal of Agricultural Sciences, Belgrade 58: 139-146.

Inger RF 1957. Ecological aspects of the origins of the tetrapods. Evolution 11: 373-376.

Islam MN, Islam MS, Alam MS 2007. Genetic structure of different populations of walking catfish (*Clarias batrachus* L.) in bangladesh. Biochemical Genetics 45: 647-662. doi: 10.1007/s10528-007-9102-1

J. AD 2011. *Clarias batrachus*. The IUCN Red List of Threatened Species 2011: e.T166613A6247551. http://dx.doi.org/10.2305/IUCN.UK.2011-1.RLTS.T166613A6247551.en.Downloaded on 17 December 2016.

Janis CM, FARMER C 1999. Proposed habitats of early tetrapods: gills, kidneys, and the water-land transition. Zoological Journal of the Linnean Society 126: 117-126.

Jiang Y, Feng S, Xu J, Zhang S, Li S, Sun X, Xu P 2016. Comparative transcriptome analysis between aquatic and aerial breathing organs of *Channa argus* to reveal the genetic basis underlying bimodal respiration. Marine Genomics 29: 89-96.

Jianxun C, Xiuhai R, Qixing Y 1991. Nuclear DNA content variation in fishes. Cytologia 56: 425-429.

Johansen K 1970. 9 Air Breathing in Fishes. Fish physiology 4: 361-411.

Johnels AG 1957. The mode of terrestrial locomotion in Clarias. Oikos 8: 122-129.

Jukes TH, Cantor CR 1969. Evolution of protein molecules. Mammalian protein metabolism 3: 132.

Kelley JL, Yee M-C, Brown AP, Richardson RR, Tatarenkov A, Lee CC, Harkins TT, Bustamante CD, Earley RL 2016. The genome of the self-fertilizing mangrove rivulus fish, *Kryptolebias marmoratus*: a model for studying phenotypic plasticity and adaptations to extreme environments. Genome Biology and Evolution 8: 2145-2154.

Kent WJ 2002. BLAT—the BLAST-like alignment tool. Genome Res 12: 656-664.

Khedkar GD, Reddy ACS, Mann P, Ravinder K, Muzumdar K 2010. *Clarias batrachus* (Linn. 1758) population is lacking genetic diversity in India. Molecular biology reports 37: 1355-1362.

Kinsella RJ, Kähäri A, Haider S, Zamora J, Proctor G, Spudich G, Almeida-King J, Staines D, Derwent P, Kerhornou A 2011. Ensembl BioMarts: a hub for data retrieval across taxonomic space. Database 2011: bar030.

Koch J, Lüdemann J, Spies R, Last M, Amemiya CT, Burmester T 2016. Unusual diversity of myoglobin genes in the lungfish. Molecular Biology and Evolution 33: 3033-3041.

Kurzweil VC, Getman M, Green ED, Lane RP 2009. Dynamic evolution of V1R putative pheromone receptors between *Mus musculus* and *Mus spretus*. Bmc Genomics 10: 1.

Lanfear R, Frandsen PB, Wright AM, Senfeld T, Calcott B 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. Molecular Biology and Evolution: msw260.

Lefevre S, Bayley M, McKenzie D, Craig J 2014a. Air-breathing fishes. Journal of Fish Biology 84: 547-553.

Lefevre S, Domenici P, McKenzie D 2014b. Swimming in air-breathing fishes. Journal of Fish Biology 84: 661-681.

Lewis S 1979. The morphology of the accessory air-breathing organs of the catfish, *Clarias batrachus*: A SEM Study. Journal of Fish Biology 14: 187-191.

Li C, Jackson RM 2002. Reactive species mechanisms of cellular hypoxia-reoxygenation injury. American Journal of Physiology-Cell Physiology 282: C227-C241.

Li W-H, Graur D. 1991. Fundamentals of molecular evolution: Sinauer Associates.

Lin H-C, Sung W-T 2003. The distribution of mitochondria-rich cells in the gills of air-breathing fishes. Physiological and Biochemical Zoology 76: 215-228.

Liu Z, Liu S, Yao J, Bao L, Zhang J, Li Y, Jiang C, Sun L, Wang R, Zhang Y, Zhou T, Zeng Q, Fu Q, Gao S, Li N, Koren S, Jiang Y, Zimin A, Xu P, Phillippy AM, Geng X, Song L, Sun F, Li C, Wang X, Chen A, Jin Y, Yuan Z, Yang Y, Tan S, Peatman E, Lu J, Qin Z, Dunham R, Li Z, Sonstegard T, Feng J, Danzmann RG, Schroeder S, Scheffler B, Duke MV, Ballard L, Kucuktas H, Kaltenboeck L, Liu H, Armbruster J, Xie Y, Kirby ML, Tian Y, Flanagan ME, Mu W, Waldbieser GC 2016. The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. Nat Commun 7: 11757. doi: 10.1038/ncomms11757

Luo W, Cao X, Xu X, Huang S, Liu C, Tomljanovic T 2016. Developmental transcriptome analysis and identification of genes involved in formation of intestinal air-breathing function of Dojo loach, *Misgurnus anguillicaudatus*. Sci Rep 6.

Martin K 2014. Theme and variations: amphibious air-breathing intertidal fishes. Journal of Fish Biology 84: 577-602.

Martin M 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet. journal 17: pp. 10-12.

Mi H, Huang X, Muruganujan A, Tang H, Mills C, Kang D, Thomas PD 2016. PANTHER version 11: expanded annotation data from Gene Ontology and Reactome pathways, and data analysis tool enhancements. Nucleic Acids Research: gkw1138.

Millikan G 1937. Experiments on muscle haemoglobin in vivo; the instantaneous measurement of muscle metabolism. Proceedings of the Royal Society of London. Series B, Biological Sciences: 218-241.

Millikan G 1939. Muscle hemoglobin. Physiological Reviews 19: 503-523.

Mohindra V, Tripathi RK, Singh A, Singh RK, Lal KK 2014. Identification of candidate reference genes for quantitative expression analysis by real-time PCR for hypoxic stress in Indian catfish, *Clarias batrachus* (Linnaeus, 1758). International Aquatic Research 6: 61.

Mombaerts P 2004. Genes and ligands for odorant, vomeronasal and taste receptors. Nature Reviews Neuroscience 5: 263-278.

Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M 2007. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Research 35: W182-W185.

Mos L, Cooper GA, Serben K, Cameron M, Koop BF 2008. Effects of diesel on survival, growth, and gene expression in rainbow trout (*Oncorhynchus mykiss*) fry. Environmental science & technology 42: 2656-2662.

Mulder GJ. 2003. Conjugation reactions in drug metabolism: an integrated approach: CRC Press.

Munshi J 1961. The accessory respiratory organs of *Clarias batrachus* (Linn.). J Morphol 109: 115-139.

Murray JI, Whitfield ML, Trinklein ND, Myers RM, Brown PO, Botstein D 2004. Diverse and specific gene expression responses to stresses in cultured human cells. Molecular biology of the cell 15: 2361-2374.

Narumiya S 1996. The small GTPase Rho: cellular functions and signal transduction. Journal of Biochemistry 120: 215-228.

Nelson J 2014. Breaking wind to survive: fishes that breathe air with their gut. Journal of Fish Biology 84: 554-576.

Nico L editor. American Fisheries Society Symposium. 2005.

Nikaido M, Noguchi H, Nishihara H, Toyoda A, Suzuki Y, Kajitani R, Suzuki H, Okuno M, Aibara M, Ngatunga BP 2013. Coelacanth genomes reveal signatures for evolutionary transition from water to land. Genome Res 23: 1740-1748.

Olson K, Ghosh T, Roy P, Munshi J 1995. Microcirculation of gills and accessory respiratory organs of the walking catfish *Clarias batrachus*. The Anatomical Record 242: 383-399.

Pace C, Gibb A 2014. Sustained periodic terrestrial locomotion in air-breathing fishes. Journal of Fish Biology 84: 639-660.

Pace CM, Gibb AC 2011. Locomotor behavior across an environmental transition in the ropefish, *Erpetoichthys calabaricus*. Journal of Experimental Biology 214: 530-537.

Parra G, Bradnam K, Korf I 2007. CEGMA: a pipeline to accurately annotate core genes in eukaryotic genomes. Bioinformatics 23: 1061-1067. doi: 10.1093/bioinformatics/btm071

Perry S, Euverman R, Wang T, Loong A, Chew S, Ip Y, Gilmour K 2008. Control of breathing in African lungfish (*Protopterus dolloi*): a comparison of aquatic and cocooned (terrestrialized) animals. Respiratory physiology & neurobiology 160: 8-17.

Perry SF 1997. The chloride cell: structure and function in the gills of freshwater fishes. Annual Review of Physiology 59: 325-347.

Pfaffl MW, Horgan GW, Dempfle L 2002. Relative expression software tool (REST©) for group-wise comparison and statistical analysis of relative expression results in real-time PCR. Nucleic Acids Research 30: e36-e36.

Pfister P, Rodriguez I 2005. Olfactory expression of a single and highly variable V1r pheromone receptor-like gene in fish species. Proc Natl Acad Sci U S A 102: 5489-5494.

Price AL, Jones NC, Pevzner PA 2005. *De novo* identification of repeat families in large genomes. Bioinformatics 21: i351-i358.

Randall D, Ip Y, Chew S, Wilson J 2004. Air breathing and ammonia excretion in the giant mudskipper, *Periophthalmodon schlosseri*. Physiological and Biochemical Zoology 77: 783-788.

Randall D, Val A. 1995. The role of carbonic anhydrase in aquatic gas exchange. In. Mechanisms of Systemic Regulation: Springer. p. 25-39.

Randall DJ. 1981. The evolution of air breathing in vertebrates: Cambridge University Press.

Reynafarje B 1962. Myoglobin content and enzymatic activity of muscle and altitude adaptation. Journal of Applied Physiology 17: 301-305.

Roesner A, Mitz SA, Hankeln T, Burmester T 2008. Globins and hypoxia adaptation in the goldfish, Carassius auratus. FEBS journal 275: 3633-3643.

Romer AS 1967. Major steps in vertebrate evolution. Science 158: 1629-1637.

Saha N, Das L 1999. Stimulation of ureogenesis in the perfused liver of an Indian air-breathing catfish, *Clarias batrachus*, infused with different concentrations of ammonium chloride. Fish Physiology and Biochemistry 21: 303-311.

Saha N, Dutta S, Bhattacharjee A 2002. Role of amino acid metabolism in an air-breathing catfish, *Clarias batrachus* in response to exposure to a high concentration of exogenous ammonia. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology 133: 235-250.

Saha N, Dutta S, Haussinger D 2000. Changes in free amino acid synthesis in the perfused liver of an air-breathing walking catfish, *Clarias batrachus* infused with ammonium chloride: a strategy to adapt under hyperammonia stress. The Journal of experimental zoology 286: 13-23.

Saha N, Ratha B 2007. Functional ureogenesis and adaptation to ammonia metabolism in Indian freshwater air-breathing catfishes. Fish Physiology and Biochemistry 33: 283-295.

Saha N, Ratha B 1998. Ureogenesis in Indian air-breathing teleosts: adaptation to environmental constraints. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology 120: 195-208.

Saraiva LR, Korsching SI 2007. A novel olfactory receptor gene family in teleost fish. Genome Res 17: 1448-1457.

Sayer M, Davenport J 1991. Amphibious fish: why do they leave water? Reviews in Fish Biology and Fisheries 1: 159-181.

Schwarze K, Campbell KL, Hankeln T, Storz JF, Hoffmann FG, Burmester T 2014. The globin gene repertoire of lampreys: convergent evolution of hemoglobin and myoglobin in jawed and jawless vertebrates. Molecular Biology and Evolution 31: 2708-2721.

Shartau R, Brauner C 2014. Acid–base and ion balance in fishes with bimodal respiration. Journal of Fish Biology 84: 682-704.

Sidell BD, O'Brien KM 2006. When bad things happen to good fish: the loss of hemoglobin and myoglobin expression in Antarctic icefishes. Journal of Experimental Biology 209: 1791-1802.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. doi: 10.1093/bioinformatics/btv351

Simpson JT, Wong K, Jackman SD, Schein JE, Jones SJ, Birol I 2009. ABySS: a parallel assembler for short read sequence data. Genome Res 19: 1117-1123. doi: 10.1101/gr.089532.108

Srivastava S, Kushwaha B, Prakash J, Kumar R, Nagpure N, Agarwal S, Pandey M, Das P, Joshi C, Jena J 2016. Development and characterization of genic SSR markers from low depth genome sequence of *Clarias batrachus* (magur). Journal of genetics 95: 603-609.

Stamatakis A 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312-1313.

Stanke M, Steinkamp R, Waack S, Morgenstern B 2004. AUGUSTUS: a web server for gene finding in eukaryotes. Nucleic Acids Research 32: W309-W312. doi: 10.1093/nar/gkh379

Suyama M, Torrents D, Bork P 2006. PAL2NAL: robust conversion of protein sequence alignments into the corresponding codon alignments. Nucleic Acids Research 34: W609-W612.

Syed AS, Korsching SI 2014. Positive Darwinian selection in the singularly large taste receptor gene family of an 'ancient' fish, *Latimeria chalumnae*. Bmc Genomics 15: 1.

Talavera G, Castresana J 2007. Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst Biol 56: 564-577.

Tamura K, Stecher G, Peterson D, Filipski A, Kumar S 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution: mst197.

Tarallo A, Angelini C, Sanges R, Yagi M, Agnisola C, D'Onofrio G 2016. On the genome base composition of teleosts: the effect of environment and lifestyle. Bmc Genomics 17: 1.

Thompson JD, Gibson T, Higgins DG 2002. Multiple sequence alignment using ClustalW and ClustalX. Current protocols in bioinformatics: 2.3. 1-2.3. 22.

Toksoz D, Merdek K 2001. The Rho small GTPase: functions in health and disease. Histology and histopathology 17: 915-927.

Trapnell C, Pachter L, Salzberg SL 2009. TopHat: discovering splice junctions with RNA-Seq. Bioinformatics 25: 1105-1111.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7: 562-578.

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, Van Baren MJ, Salzberg SL, Wold BJ, Pachter L 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28: 511-515.

Turko AJ, Robertson CE, Bianchini K, Freeman M, Wright PA 2014. The amphibious fish *Kryptolebias marmoratus* uses different strategies to maintain oxygen delivery during aquatic hypoxia and air exposure. Journal of Experimental Biology 217: 3988-3995.

Van der Oost R, Beyer J, Vermeulen NP 2003. Fish bioaccumulation and biomarkers in environmental risk assessment: a review. Environmental toxicology and pharmacology 13: 57-149.

Vinogradov AE 1998. Genome size and GC-percent in vertebrates as determined by flow cytometry: The triangular relationship. Cytometry 31: 100-109.

Wells MW, Turko AJ, Wright PA 2015. Fish embryos on land: terrestrial embryo deposition lowers oxygen uptake without altering growth or survival in the amphibious fish *Kryptolebias marmoratus*. Journal of Experimental Biology 218: 3249-3256.

Wittenberg B, Wittenberg J, Caldwell P 1975. Role of myoglobin in the oxygen supply to red skeletal muscle. Journal of Biological Chemistry 250: 9038-9043.

Wittenberg JB 1970. Myoglobin-facilitated oxygen diffusion: role of myoglobin in oxygen entry into muscle. Physiological Reviews 50: 559-636.

Wright PA 2012. Environmental physiology of the mangrove rivulus, *Kryptolebias marmoratus*, a cutaneously breathing fish that survives for weeks out of water. Integrative and Comparative Biology 52: 792-800.

Xu J, Li J-T, Jiang Y, Peng W, Yao Z, Chen B, Jiang L, Feng J, Ji P, Liu G 2016. Genomic basis of adaptive evolution: the survival of Amur ide (*Leuciscus waleckii*) in an extremely alkaline environment. Molecular Biology and Evolution: msw230.

Yang Z 2007. PAML 4: phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution 24: 1586-1591.

Yoshihara Y. 2008. Molecular genetic dissection of the zebrafish olfactory system. In. Chemosensory Systems in Mammals, Fishes, and Insects: Springer. p. 1-19.

You X, Bian C, Zan Q, Xu X, Liu X, Chen J, Wang J, Qiu Y, Li W, Zhang X, Sun Y, Chen S, Hong W, Li Y, Cheng S, Fan G, Shi C, Liang J, Tom Tang Y, Yang C, Ruan Z, Bai J, Peng C, Mu Q, Lu J, Fan M, Yang S, Huang Z, Jiang X, Fang X, Zhang G, Zhang Y, Polgar G, Yu H, Li J, Liu Z, Zhang G, Ravi V, Coon SL, Wang J, Yang H, Venkatesh B, Wang J, Shi Q 2014. Mudskipper genomes provide insights into the terrestrial adaptation of amphibious fishes. Nat Commun 5: 5594. doi: 10.1038/ncomms6594

Young JM, Massa HF, Hsu L, Trask BJ 2010. Extreme variability among mammalian V1R gene families. Genome Res 20: 10-18.

Zhu L, Qu K, Xia B, Sun X, Chen B 2016. Transcriptomic response to water accommodated fraction of crude oil exposure in the gill of Japanese flounder, *Paralichthys olivaceus*. Marine pollution bulletin 106: 283-291.

Zvaritch E, Depreux F, Kraeva N, Loy RE, Goonasekera SA, Boncompagni S, Kraev A, Gramolini AO, Dirksen RT, Franzini-Armstrong C 2007. An Ryr1I4895T mutation abolishes Ca^{2+} release channel function and delays development in homozygous offspring of a mutant mouse line. Proceedings of the National Academy of Sciences 104: 18537-18542.

APPENDICES

Appendix 1. Positively selected genes in the *C. batrachus* genome compared with those of 11 non-air-breathing teleost fishes.

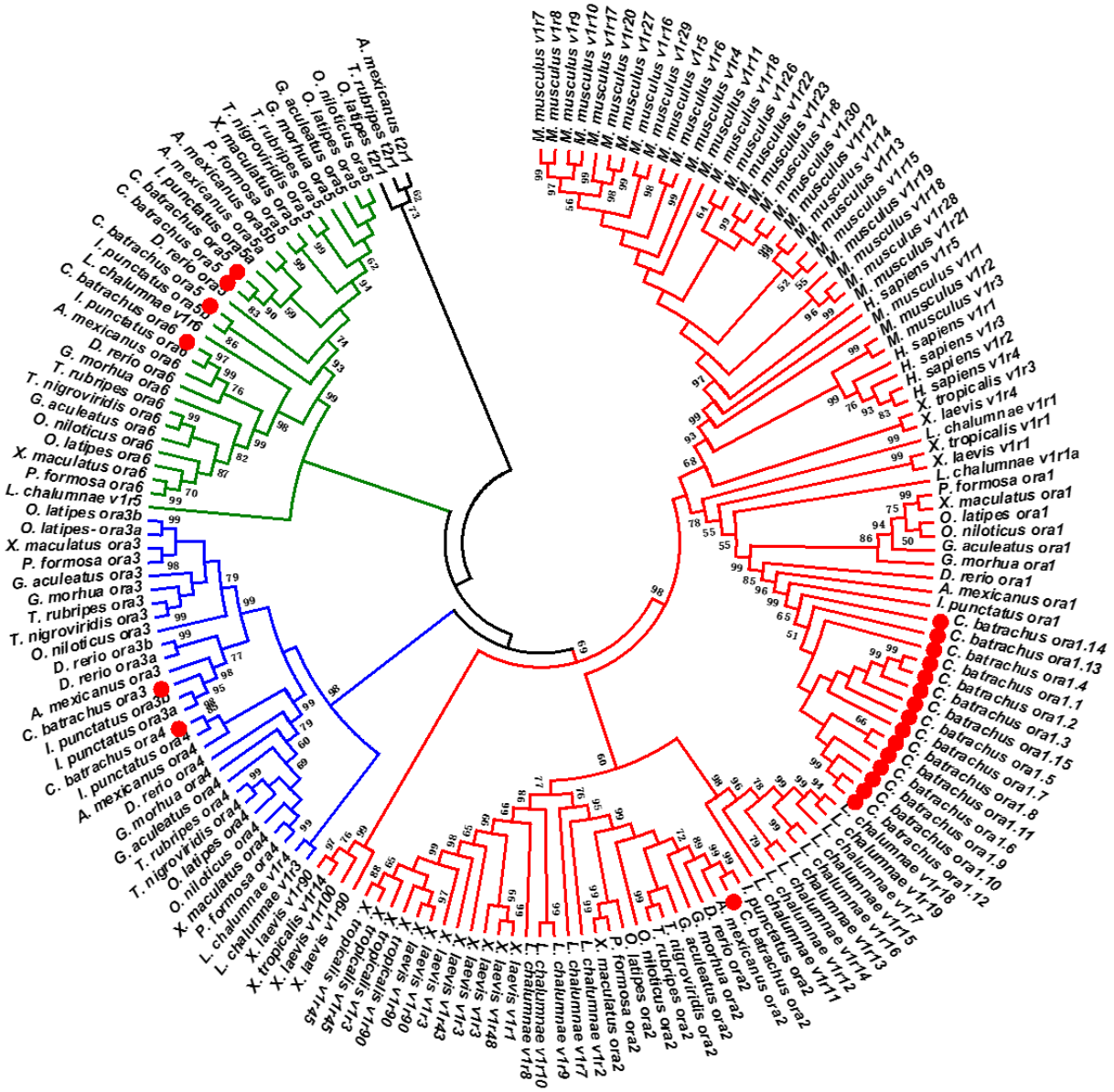
Gene ID	Gene name	Zebrafish protein ID	FDR-adjusted *P*-value
g705.t1	isovaleryl- mitochondrial	ENSDARP00000062893	0.0021173
g5574.t1	cell division cycle 16 homolog	ENSDARP00000072280	2.02E-05
g23227.t1	origin recognition complex subunit 5	ENSDARP00000033056	0
g2369.t1	nibrin-like	ENSDARP00000058973	0
g9971.t1	COP9 signalosome complex subunit 3	ENSDARP00000021983	0.009628888
g6795.t1	DNA repair XRCC3	ENSDARP00000004619	0.016387317
g20799.t1	dnaJ homolog subfamily C member 12	ENSDARP00000105475	0.03822202
g16100.t1	ribonuclease H2 subunit B	ENSDARP00000006173	0.038761503
g20577.t1	peptidyl-prolyl cis-trans isomerase FKBP3	ENSDARP00000132658	6.83E-08
g6229.t1	peroxisomal membrane 2	ENSDARP00000026766	8.81E-05
g9181.t1	A-kinase anchor mitochondrial isoform X1	ENSDARP00000114357	0.000448024
g7941.t1	ADP-ribosylation factor 2 isoform X2	ENSDARP00000071659	0
g660.t1	COP9 signalosome complex subunit 5 isoform X2	ENSDARP00000074782	0.002205753
g1151.t1	ATP-dependent RNA helicase DDX54	ENSDARP00000138094	5.57E-05
g20503.t1	transcription initiation factor IIE subunit beta	ENSDARP00000020885	5.49E-05
g4567.t1	Golgi-specific brefeldin A-resistance guanine nucleotide exchange factor 1 isoform X1	ENSDARP00000046791	7.06E-07
g4465.t1	tyrosine--tRNA mitochondrial	ENSDARP00000070911	0.006199169
g2182.t1	adenylyltransferase and sulfurtransferase MOCS3	ENSDARP00000013020	0.02048943
g19244.t1	CDK-activating kinase assembly factor MAT1	ENSDARP00000027502	0
g13777.t1	calsyntenin-3-like	ENSDARP00000128441	0.007272252
g11132.t1	eukaryotic translation initiation factor 3 subunit M	ENSDARP00000024129	0.000282053
g20754.t1	26S proteasome non-ATPase regulatory subunit 8	ENSDARP00000025160	0
g5913.t1	GTP-binding nuclear Ran	ENSDARP00000032532	0.007866615
g3738.t1	cleavage and polyadenylation specificity factor subunit 5	ENSDARP00000017229	0
g5700.t1	post-GPI attachment to s factor 3	ENSDARP00000085952	0
g945.t1	N-alpha-acetyltransferase 10 isoform X2	ENSDARP00000095601	0.015568892
g18774.t1	trafficking particle complex subunit 3	ENSDARP00000066711	6.07E-07
g18362.t1	ribulose-phosphate 3-epimerase	ENSDARP00000023863	5.74E-08
g8932.t1	peptidyl-prolyl cis-trans isomerase-like 1	ENSDARP00000127620	0.023580507
g8994.t1	citrate mitochondrial	ENSDARP00000135698	0
g1135.t1	stress-induced-phospho 1	ENSDARP00000100503	3.40E-05
g21120.t1	splicing factor 3B subunit 2	ENSDARP00000007002	5.86E-06
g21032.t1	oxygen-dependent coproporphyrinogen-III mitochondrial	ENSDARP00000108678	0
g7235.t1	argininosuccinate synthase	ENSDARP00000133962	1.60E-06
g6049.t1	probable 18S rRNA (guanine-N(7))-methyltransferase	ENSDARP00000087506	3.40E-05
g22990.t1	succinate dehydrogenase [ubiquinone] flavo mitochondrial	ENSDARP00000018027	0.019288562

g24348.t1	probable arginine--tRNA mitochondrial	ENSDARP00000119548	0.006978767
g13578.t1	adenylate kinase mitochondrial isoform X1	ENSDARP00000131617	0
g8534.t1	ER membrane complex subunit 2	ENSDARP00000006378	1.69E-05
g5479.t1	acid phosphatase type 7 isoform X1	ENSDARP00000141174	0.000114807
g4594.t1	SID1 transmembrane family member 2	ENSDARP00000067582	0
g3097.t1	tRNA modification GTPase mitochondrial	ENSDARP00000004190	1.69E-05
g22295.t1	BSD domain-containing 1 isoform X2	ENSDARP00000003318	0
g2967.t1	MCM10 homolog isoform X1	ENSDARP00000067338	6.71E-06
g16705.t1	V-type proton ATPase subunit H isoform X1	ENSDARP00000024886	7.06E-07
g22360.t1	mediator of RNA polymerase II transcription subunit 6 isoform X2	ENSDARP00000044114	1.68E-06
g18046.t1	signal peptidase complex catalytic subunit SEC11A	ENSDARP00000009538	0.00324268
g4072.t1	calcium uptake mitochondrial isoform X3	ENSDARP00000086953	6.43E-05
g7366.t1	ubiquitin carboxyl-terminal hydrolase 47 isoform X2	ENSDARP00000134662	2.97E-06
g2683.t1	SCY1 2 isoform X2	ENSDARP00000039876	0.017492378
g8958.t1	NADPH:adrenodoxin mitochondrial	ENSDARP00000106579	0.049959127
g17141.t1	acyl- synthetase family member 4	ENSDARP00000118390	1.67E-05
g15564.t1	DNA polymerase alpha subunit B	ENSDARP00000002618	0.00071127
g18180.t1	dynamin-like 120 kDa mitochondrial isoform X3	ENSDARP00000133518	0
g4014.t1	beta-catenin 1	ENSDARP00000026984	3.37E-06
g875.t1	urocanate hydratase	ENSDARP00000094017	0.00092137
g8908.t1	KIAA0100 homolog	ENSDARP00000127487	0.023908842
g11413.t1	ribosomal RNA small subunit methyltransferase NEP1	ENSDARP00000118075	1.26E-05
g1845.t1	mediator of RNA polymerase II transcription subunit 14 isoform X1	ENSDARP00000010465	0
g18658.t1	probable ATP-dependent RNA helicase DDX52	ENSDARP00000133983	0
g7981.t1	pseudouridine-metabolizing bifunctional -like isoform X2	ENSDARP00000095090	0.00037304
g1651.t1	mitochondrial ribosome-associated GTPase 1	ENSDARP00000112811	3.58E-05
g10048.t1	ubiquinone biosynthesis COQ4 mitochondrial isoform X2	ENSDARP00000139950	0
g2049.t1	glycerophosphocholine phosphodiesterase GPCPD1-like	ENSDARP00000107604	0
g5456.t1	vacuolar -sorting-associated 36	ENSDARP00000074049	0.025268507
g13838.t1	THO complex subunit 5 homolog	ENSDARP00000055844	0
g464.t1	nucleolar complex 2 homolog	ENSDARP00000105346	3.03E-08
g24300.t1	malectin	ENSDARP00000112772	7.35E-05
g22699.t1	RNA-binding 8A	ENSDARP00000026575	1.47E-05
g12919.t1	cleavage stimulation factor subunit 2 isoform X1	ENSDARP00000110939	0
g13738.t1	tetratricopeptide repeat 17	ENSDARP00000065863	1.16E-05
g20041.t1	serine--tRNA mitochondrial	ENSDARP00000140088	0.0018394
g7870.t1	geranylgeranyl transferase type-2 subunit alpha	ENSDARP00000089797	1.20E-07
g11100.t1	coiled-coil domain-containing 94	ENSDARP00000035484	8.23E-06
g10307.t1	folliculin	ENSDARP00000133243	0
g9797.t1	pyridine nucleotide-disulfide oxidoreductase domain-containing 2	ENSDARP00000007840	0
g6652.t1	BUD13 homolog	ENSDARP00000065543	0
g18083.t1	exosome complex component RRP46	ENSDARP00000059351	6.06E-06
g4886.t1	G patch domain and KOW motifs-containing	ENSDARP00000047911	0.038862051

g9202.t1	Williams-Beuren syndrome chromosomal region 16	ENSDARP00000077387	9.86E-05
g4251.t1	proto-oncogene tyrosine- kinase receptor Ret	ENSDARP00000116296	0
g24106.t1	E3 ubiquitin- ligase AMFR-like	ENSDARP00000002529	0.00066772
g21220.t1	39S ribosomal mitochondrial	ENSDARP00000008022	0.000958468
g20130.t1	dystrophin-related 2 isoform X1	ENSDARP00000134164	0.00924134
g7322.t1	IQ and ubiquitin-like domain-containing isoform X1	ENSDARP00000134606	0.016092715
g22102.t1	syntaxin-8	ENSDARP00000139377	0.010538275
g6632.t1	E3 ubiquitin- ligase FANCL	ENSDARP00000003498	0
g5657.t1	ribosome-recycling mitochondrial	ENSDARP00000039535	0.002837616
g370.t1	RB1-inducible coiled-coil 1 isoform X1	ENSDARP00000097746	0.000905394
g1325.t1	telomere-associated RIF1	ENSDARP00000103401	0.000311061
g13865.t1	UPF0489 C5orf22 homolog	ENSDARP00000004506	1.56E-07
g5178.t1	zinc finger MYND domain-containing 10	ENSDARP00000006141	1.83E-07
g5395.t1	apoptosis inhibitor 5	ENSDARP00000126600	4.37E-05
g10005.t1	LIM homeobox Lhx1-like	ENSDARP00000132843	1.28E-06
g5828.t1	armadillo repeat-containing 6	ENSDARP00000131155	0
g8527.t1	E3 ubiquitin- ligase E3D	ENSDARP00000030264	0
g1695.t1	erythroid differentiation-related factor 1 isoform X2	ENSDARP00000134137	0
g3381.t1	methenyltetrahydrofolate synthase domain-containing isoform X1	ENSDARP00000090099	0
g9703.t1	polyadenylate-binding -interacting 1	ENSDARP00000062956	5.60E-07
g10814.t1	E3 ubiquitin- ligase RFWD2-like	ENSDARP00000101197	0.000603049
g13755.t1	RNA demethylase ALKBH5	ENSDARP00000086116	3.96E-05
g10306.t1	leukocyte receptor cluster member 9-like	ENSDARP00000117502	5.26E-05
g11610.t1	Fanconi anemia group J homolog isoform X1	ENSDARP00000102913	0
g21916.t1	aldehyde dehydrogenase family 16 member A1-like isoform X1	ENSDARP00000111827	0
g11957.t1	zinc finger FYVE domain-containing 21 isoform X1	ENSDARP00000042587	1.62E-07
g2309.t1	meiosis 1 arrest isoform X1	ENSDARP00000138928	0.008427642
g17737.t1	early endosome antigen 1 isoform X2	ENSDARP00000085782	0.000905394
g3542.t1	lisH domain and HEAT repeat-containing KIAA1468 homolog isoform X2	ENSDARP00000137498	1.62E-07
g11991.t1	glutaminase liver mitochondrial-like isoform X1	ENSDARP00000016280	0.014266232
g8399.t1	protrudin isoform X1	ENSDARP00000113701	6.70E-06
g15712.t1	synaptosomal-associated 47	ENSDARP00000087173	0.000323233
g14558.t1	ankyrin repeat domain-containing 40	ENSDARP00000100846	0.000113513
g16943.t1	HAUS augmin-like complex subunit 1	ENSDARP00000067732	0
g11655.t1	N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase	ENSDARP00000108126	6.06E-06
g13580.t1	NADH dehydrogenase [ubiquinone] iron-sulfur 5	ENSDARP00000011338	0.000557609
g15546.t1	TNFAIP3-interacting 2-like	ENSDARP00000102784	0.04096195
g11662.t1	U11 U12 small nuclear ribonucleo 48 kDa isoform X1	ENSDARP00000081780	4.42E-08
g5122.t1	GTPase IMAP family member 8-like	ENSDARP00000101279	0.000230225
g2767.t1	integrin beta 1	ENSDARP00000060053	0
g24774.t1	Hermansky-Pudlak syndrome 4 isoform X3	ENSDARP00000012323	9.63E-05
g17670.t1	microcephalin isoform X2	ENSDARP00000124269	6.83E-08
g16861.t1	transcobalamin-2-like	ENSDARP00000053000	0.01051738

g1059.t1	spindle and centriole-associated 1-like	ENSDARP00000003902	0.015568892
g9821.t1	chromodomain-helicase-DNA-binding 1-like	ENSDARP00000129508	3.03E-08
g18865.t1	anti-mullerian hormone	ENSDARP00000015395	2.18E-05
g17452.t1	transmembrane 117-like	ENSDARP00000105155	0.000355812
g15370.t1	SAC3 domain-containing 1-like	ENSDARP00000096054	0.002404338
g8796.t1	enhancer of mRNA-decapping 4	ENSDARP00000130250	1.60E-08
g11950.t1	B-cell lymphoma leukemia 11A-like	ENSDARP00000089751	0
g19306.t1	farnesyltransferase subunit beta-like	ENSDARP00000034563	0
g9345.t1	Peptide-N4-N-acetyl-beta-D-glucosaminylasparagine amidase F precursor	ENSDARP00000033960	0.042038328
g18374.t1	nuclear transcription factor Y subunit gamma isoform X1	ENSDARP00000133372	2.26E-06

Appendix 2. Detailed phylogenetic tree with species names and sequence names for Figure 9. Three clades are indicated with red (*ora1-ora2* clade), blue (*ora3-ora4* clade) and green (*ora5-ora6* clade). The red solid circles represent the *ora* genes in the *C. batrachus* genome. Bootstrap support (1,000 replications) is indicated on the branches.



Appendix 3. Differentially expressed genes in the transcriptome of air-breathing organ compared with that of the gill.

Gene ID	Gene name	Zebrafish protein ID	Fold change	FDR-adjusted *P*-value
g10034.t1	G2 mitotic-specific cyclin-B1	ENSDARP00000063356.4	-2.452966521	0.009492372
g10071.t1	electrogenic sodium bicarbonate cotransporter 1 isoform X3	ENSDARP00000048398.7	-2.10870214	0.000830777
g10078.t1	cytohesin-interacting -like	ENSDARP00000082775.5	-2.354619422	0.000314532
g10082.t1	DNA fragmentation factor subunit alpha-like	ENSDARP00000124281.1	-2.631548805	0.005779516
g1011.t1	adenosylhomocysteinase 2 isoform X1	ENSDARP00000073256.4	-5.346346494	1.18766E-08
g10124.t1	alpha-2A adrenergic receptor	ENSDARP00000059868.4	15.2465161	0.01891849
g10125.t1	GTPase IMAP family member 8-like isoform X1	ENSDARP00000127893.1	2.887503457	0.004979643
g10147.t1	tensin-2-like isoform X4	ENSDARP00000123088.2	2.421606382	0.038786442
g10242.t1	PREDICTED: uncharacterized protein LOC108255376		5.019039461	3.98811E-05
g10352.t1	cadherin 26	ENSDARP00000102644.2	-58.76752395	0.011899843
g10354.t1	cadherin 26	ENSDARP00000102644.2	-132.284319	3.02142E-05
g10355.t1	cadherin 26	ENSDARP00000109048.2	-21.7221756	7.67444E-07
g10356.t1	cadherin 26	ENSDARP00000109048.2	-2.656930749	5.03406E-05
g10404.t1	alpha-actinin-2 isoform X1	ENSDARP00000115698.1	-38.11914845	1.81475E-10
g10510.t1	Friend leukemia integration 1 transcription factor isoform X5	ENSDARP00000058612.5	-2.203900478	0.007197169
g1054.t1	brother of CDO-like	ENSDARP00000098383.3	2.750902986	0.00574796
g10555.t1	PREDICTED: uncharacterized protein LOC108260052	ENSDARP00000119741.1	-14.47051894	6.40007E-10
g10573.t1	amiloride-sensitive amine oxidase [copper-containing]	ENSDARP00000081856.4	-2.661709274	8.30846E-11
g10629.t1	Vesicle-associated membrane protein 727	ENSDARP00000122979.1	-11.04760582	4.31547E-12
g10658.t1	interferon-induced 44-like	ENSDARP00000022947.7	-6.660609822	4.38615E-13
g10684.t1	elongation factor 1-alpha-like	ENSDARP00000111742.1	-9.928691366	0
g10686.t1	kelch 31	ENSDARP00000070382.3	-147.684012	0.000137773
g10781.t1	middle subunit-like	ENSDARP00000044237.6	-3.486656774	0
g10795.t1	cytochrome c oxidase subunit mitochondrial-like	ENSDARP00000113555.1	-3.572537467	1.79511E-07
g10797.t1	glycine-rich cell wall structural 1-like isoform X4	ENSDARP00000127968.1	16.71857041	0
g108.t1	methylosome subunit pICln	ENSDARP00000133397.1	2.017584085	0.002385954
g10802.t1	perforin-1-like	ENSDARP00000080171.5	-47.05993129	0.035001379
g1081.t1	cystatin-B-like	ENSDARP00000032643.6	-1.7977E+308	7.43381E-07
g10827.t1	disintegrin and metallo ase domain-containing 33-like	ENSDARP00000077690.4	2.268117	0.024649469
g1084.t1	cystatin-B-like	ENSDARP00000121287.1	51.0589336	0
g10880.t1	solute carrier organic anion transporter family member 2B1-like	ENSDARP00000120042.1	-2.396161813	0.000368153
g1094.t1	villin	ENSDARP00000048939.5	-4.435200723	0.008930064
g10958.t1	cytosolic sulfotransferase 3-like	ENSDARP00000137427.1	3.701702769	0
g10968.t1	probable G- coupled receptor 160	ENSDARP00000133078.1	-6.427697932	0.016910434
g11027.t1	fibroblast growth factor receptor 2 isoform X2	ENSDARP00000124472.1	2.21327492	0.012122938
g11046.t1	ceramide kinase-like	ENSDARP00000036893.6	-7.072949462	0.044935843
g11057.t1	nitric oxide inducible	ENSDARP00000132752.1	-3.739999581	0.001580848
g11123.t1	metallo ase inhibitor 2-like	ENSDARP00000141124.1	4.546163385	0

g11262.t1	myosin-binding fast-type isoform X1	ENSDARP00000107798.2	-156.7899174	1.90878E-05
g11279.t1	bone morphogenetic 5	ENSDARP00000134612.1	3.129416277	0.002757629
g11285.t1	kelch 31	ENSDARP00000093566.2	-20.04240976	0.020947111
g11323.t1	glycogen muscle form	ENSDARP00000072337.4	-120.5193362	0.023045393
g11327.t1	troponin fast skeletal muscle-like	ENSDARP00000130768.1	-19.18471066	9.49749E-05
g11328.t1	troponin fast skeletal muscle-like	ENSDARP00000027342.5	-150.3622195	1.26052E-13
g11332.t1	troponin fast skeletal muscle isoforms-like isoform X1	ENSDARP00000089786.3	-13.04871009	0.000136703
g11344.t1	tomoregulin-1-like	ENSDARP00000073739.5	3.315197935	0.041929563
g11345.t1	muscle-related coiled-coil -like	ENSDARP00000073748.4	-1.7977E+308	0.003670976
g11351.t1	toll-like receptor 5	ENSDARP00000124387.1	11.74546425	0.006178306
g11353.t1	myosin regulatory light chain skeletal muscle isoform type 2	ENSDARP00000069948.4	-3.498290907	0
g11364.t1	ABI gene family member 3-like isoform X2	ENSDARP00000078719.5	-2.07186309	0.012075457
g11371.t1	RING finger 208	ENSDARP00000100778.2	7.121354104	0.003945306
g11415.t1	glyceraldehyde-3-phosphate dehydrogenase	ENSDARP00000102247.2	-14.69067436	0
g11459.t1	globoside alpha-1,3-N-acetylgalactosaminyltransferase 1-like	ENSDARP00000139562.1	-4.858509444	0.000250064
g1146.t1	histone-lysine N-methyltransferase SMYD1-like	ENSDARP00000027046.7	-49.35553769	0.007374681
g11478.t1	Immunoglobulin light partial	ENSDARP00000119905.2	-5.131355492	0.010965491
g11522.t1	prolactin receptor-like	ENSDARP00000039548.4	-22.0304163	4.12998E-07
g11539.t1	fibroblast growth factor receptor 1-A isoform X1	ENSDARP00000013742.8	2.194583908	0.004676425
g11547.t1	tyrosine- kinase ZAP-70	ENSDARP00000006727.7	-4.591212808	0.003667288
g11548.t1	tyrosine- kinase ZAP-70	ENSDARP00000114598.1	-3.299934206	0.001263437
g11591.t1	angiopoietin-related 1-like	ENSDARP00000006394.5	-4.395246408	0.010796353
g11596.t1	proactivator polypeptide-like	ENSDARP00000088025.4	-2.095695475	1.81837E-06
g11613.t1	T-box transcription factor TBX2	ENSDARP00000107769.2	2.32451443	0.00782993
g11643.t1	C-X-C motif chemokine 10-like	ENSDARP00000102486.2	-3.826010674	0.00441575
g11671.t1	retinol dehydrogenase 3-like	ENSDARP00000027139.6	-2.611703504	0
g11691.t1	omega-amidase NIT2	ENSDARP00000121828.1	-4.026140463	0.013303025
g11759.t1	myosin-binding slow-type	ENSDARP00000009148.9	-282.2448074	0
g11795.t1	cingulin 1 isoform X1	ENSDARP00000004196.8	-5.706531014	0.01045829
g11811.t1	claudin-4-like	ENSDARP00000092862.3	-11.47803202	0.000533499
g11812.t1	claudin-4-like	ENSDARP00000025285.5	-28.63466936	0
g11814.t1	claudin-4-like	ENSDARP00000063319.4	2.252573162	0
g11815.t1	claudin-4-like	ENSDARP00000076279.1	-3.055182053	0.001344123
g11844.t1	beta-2-glyco 1-like	ENSDARP00000071352.6	9.002704745	0.022144795
g11858.t1	arylacetamide deacetylase-like	ENSDARP00000087565.3	24.61223313	9.26782E-05
g11924.t1	matrix metallo ase-23	ENSDARP00000063241.5	2.984607961	0.018159385
g11997.t1	aggrecan core -like	ENSDARP00000124852.1	10.96915544	0
g12019.t1	connective tissue growth factor-like	ENSDARP00000063027.5	2.529016554	0
g12070.t1	antigen KI-67-like isoform X2	ENSDARP00000126723.1	-2.644073954	0.046081303
g12078.t1	band 3 anion exchange -like	ENSDARP00000134130.1	-85.83415821	0
g12105.t1	dispanin subfamily A member 2b-like	ENSDARP00000121961.1	-2.348581937	0.047305119
g12106.t1	histamine N-methyltransferase-like	ENSDARP00000004085.5	-2.05016421	0.035278621
g12108.t1	histamine N-methyltransferase-like	ENSDARP00000004085.5	-13.45144805	3.55489E-13

g12115.t1	acetyl- mitochondrial	ENSDARP00000067447.4	-2.843298326	2.84907E-09
g12116.t1	collagenase 3-like	ENSDARP00000067446.5	-56.91190877	0
g12123.t1	anterior gradient 2 homolog	ENSDARP00000135194.1	-9.906707851	0
g12202.t1	interferon-induced GTP-binding Mx1-like isoform X1	ENSDARP00000094364.3	7.720895838	0.001053751
g12205.t1	fibronectin-like isoform X2	ENSDARP00000131015.1	4.896387879	0
g12228.t1	carbohydrate sulfotransferase 6-like	ENSDARP00000124000.1	2.660297459	0.001397525
g12265.t1	arylsulfatase D-like	ENSDARP00000030967.7	2.200528097	0.02339616
g12385.t1	calponin-3-like isoform X2	ENSDARP00000058483.4	2.282661884	1.16268E-05
g12421.t1	probable glutamate receptor	ENSDARP00000123565.1	2.649248862	0.004098562
g12504.t1	tubulin alpha-1A chain	ENSDARP00000019888.7	2.35572667	0.001744271
g12506.t1	junctophilin-2	ENSDARP00000043428.6	-34.01671308	0.009888673
g1259.t1	homeobox Nkx-	ENSDARP00000054821.4	39.11361875	0
g126.t1	oligodendrocyte-myelin glyco -like	ENSDARP00000118578.1	-36.04102055	0.03013822
g12624.t1	ammonium transporter Rh type B	ENSDARP00000058869.5	-5.625601416	0
g12684.t1	guanylate-binding 1-like	ENSDARP00000135031.1	-2.223134188	2.87593E-07
g12708.t1	keratinocyte differentiation factor 1-like	ENSDARP00000138353.1	-2.285493601	0.031359623
g12842.t1	propionyl- carboxylase beta mitochondrial	ENSDARP00000133677.1	-2.197045042	2.78013E-06
g12853.t1	purine nucleoside phosphorylase-like	ENSDARP00000093456.4	2.797751043	1.15352E-09
g12918.t1	cytochrome b-245 heavy chain-like	ENSDARP00000080539.5	2.103482162	9.75956E-05
g12974.t1	EMILIN-1-like isoform X2	ENSDARP00000095986.2	4.954050049	2.77347E-05
g13010.t1	creatine kinase M-type	ENSDARP00000037871.5	-80.90808585	0
g13013.t1	immunoglobulin light chain	ENSDARP00000119905.2	-4.904250045	0.003667288
g13133.t1	suppressor of cytokine signaling 1-like	ENSDARP00000122638.1	-2.110682555	0.016912018
g1318.t1	formin 2 isoform X4	ENSDARP00000138630.1	2.763531384	0.012105528
g13232.t1	desmin-like	ENSDARP00000065355.4	-12.21262607	0.000189416
g1324.t1	nebulin isoform X1	ENSDARP00000117320.1	-42.18176768	0.009950007
g13251.t1	epidermal growth factor receptor kinase substrate 8 2 isoform X2	ENSDARP00000095696.3	7.595334028	3.07944E-11
g1329.t1	ETS translocation variant 5 isoform X1	ENSDARP00000108566.2	-3.629321393	0.006965118
g13332.t1	lysyl oxidase homolog 4-like	ENSDARP00000117086.2	5.451784545	3.48461E-05
g13335.t1	secreted frizzled-related 5	ENSDARP00000056995.4	-6.611346444	0.022750741
g13368.t1	cytoskeleton-associated 4	ENSDARP00000049169.4	3.084307976	0.001472665
g13377.t1	monocarboxylate transporter 1-like	ENSDARP00000124803.1	-66.14215952	1.59522E-07
g13383.t1	lipopolysaccharide-induced tumor necrosis factor-alpha factor homolog	ENSDARP00000117757.1	-6.687201264	0.000964257
g13386.t1	---NA---	ENSDARP00000136239.1	-18.44883683	0
g13389.t1	somatostatin receptor type 5-like isoform X1	ENSDARP00000132886.1	-7.317245413	0.000599893
g13409.t1	serine threonine- kinase Sgk2	ENSDARP00000086932.4	-10.55978946	0.000219053
g13412.t1	PRA1 family 3	ENSDARP00000015357.5	2.206666202	0.000453245
g13445.t1	type I cytoskeletal 18-like	ENSDARP00000064982.5	-20.02916588	1.19272E-05
g13452.t1	tyrosine- kinase ITK TSK	ENSDARP00000022015.5	-2.612078892	0.026766696
g13465.t1	syndecan-1	ENSDARP00000131513.1	2.289294515	1.50993E-05
g13467.t1	ornithine decarboxylase	ENSDARP00000124918.1	3.55227108	0
g13488.t1	troponin fast skeletal muscle isoforms-like isoform X3	ENSDARP00000121196.1	-115.3983681	0
g13575.t1	collagen alpha-2(VIII) chain	ENSDARP00000114013.1	13.1249268	0

g13589.t1	thioredoxin-related transmembrane 2-B	ENSDARP00000022891.5	-4.11491352	2.75947E-06
g13602.t1	ERBB receptor feedback inhibitor 1-like	ENSDARP00000105673.2	4.659183802	6.73054E-05
g13645.t1	myosin heavy fast skeletal muscle-like	ENSDARP00000128891.1	-36.45791345	0
g13669.t1	forkhead box E1	ENSDARP00000098221.1	-83.64615835	4.66783E-14
g13680.t1	dentin sialophospho - partial		-138.3102859	2.08037E-06
g13693.t1	pyruvate kinase PKM-like	ENSDARP00000132920.1	-80.98389259	0
g13803.t1	filaggrin-like isoform X9		-276.3336209	0
g13889.t1	forkhead box I1-like	ENSDARP00000002195.2	-24.10386724	0
g13928.t1	ephrin type-B receptor 3 isoform X7	ENSDARP00000140419.1	2.11523789	0.049078415
g13956.t1	gelsolin-like isoform X1	ENSDARP00000026895.6	4.051488377	0
g13957.t1	GDNF family receptor alpha-1-like isoform X2	ENSDARP00000137090.1	5.681931466	0.03968896
g13963.t1	glutathione peroxidase 1-like	ENSDARP00000020656.8	3.153605342	0.000270384
g13973.t1	myosin light chain 4-like	ENSDARP00000048394.2	-1.7977E+308	0
g13975.t1	transmembrane protease serine 9-like	ENSDARP00000021264.6	3.237174982	0
g13976.t1	transmembrane protease serine 9-like	ENSDARP00000118878.1	47.25312496	0
g1398.t1	dual specificity tyrosine-phosphorylation-regulated kinase 4-like isoform X1	ENSDARP00000139074.1	-2.051463471	0.010228179
g13996.t1	tenascin isoform X1	ENSDARP00000114191.1	-12.29511227	0.026359646
g14015.t1	FAM43B-like isoform X2	ENSDARP00000135701.1	2.339412523	0.025366151
g14062.t1	tripartite motif-containing 35-like	ENSDARP00000062165.4	-2.063065839	7.88692E-07
g14068.t1	frizzled-9-like	ENSDARP00000006458.6	2.52405581	0.033212016
g14129.t1	xylosyltransferase 1	ENSDARP00000081552.4	6.915384088	0.011522455
g14160.t1	interferon-induced GTP-binding Mx1	ENSDARP00000094364.3	2.292904937	8.47626E-06
g14161.t1	lanC 1	ENSDARP00000018694.5	-2.205983414	0.001409192
g14248.t1	UDP-glucose 6-dehydrogenase	ENSDARP00000108630.2	4.535278754	7.82996E-05
g1425.t1	achaete-scute homolog 1b-like	ENSDARP00000022655.5	-6.062755375	3.21017E-05
g14277.t1	immunoglobulin heavy chain variable partial	ENSDARP00000125907.1	-5.021639009	0.020502987
g14283.t1	atypical chemokine receptor 3-like	ENSDARP00000063664.6	3.326268151	2.36648E-14
g14301.t1	Friend leukemia integration 1 transcription factor isoform X2	ENSDARP00000071331.4	2.357444506	0.001524077
g14302.t1	ATP-sensitive inward rectifier potassium channel 1-like	ENSDARP00000089221.2	-12.53131025	0
g14303.t1	ATP-sensitive inward rectifier potassium channel 1-like	ENSDARP00000058988.5	-8.272098939	3.24982E-08
g14304.t1	ATP-sensitive inward rectifier potassium channel 1-like	ENSDARP00000058986.6	-12.29789145	0.008336564
g14307.t1	EMILIN-2-like	ENSDARP00000081460.5	2.280638104	0.000486487
g14327.t1	tumor necrosis factor receptor superfamily member 11B-like	ENSDARP00000133879.1	2.150988874	0.00031788
g14332.t1	CD209 antigen C isoform X1	ENSDARP00000073318.4	3.907358021	0
g14364.t1	guanine nucleotide-binding G(i) subunit alpha-1	ENSDARP00000037326.5	-3.053589651	0.03149761
g1439.t1	aggrecan core -like	ENSDARP00000137330.1	6.000439737	4.2839E-05
g14395.t1	lysophosphatidic acid receptor 6-like	ENSDARP00000067835.5	7.097957915	0.002599845
g14414.t1	PDZ and LIM domain 2-like	ENSDARP00000109858.2	3.06672781	4.90049E-05
g14421.t1	guanine nucleotide-binding subunit alpha-14-like	ENSDARP00000028232.6	-48.20773449	2.44068E-05
g14429.t1	angiotensin-converting enzyme	ENSDARP00000099084.3	-10.95463376	1.30812E-05
g14453.t1	fatty acid-binding heart	ENSDARP00000039143.6	-2.69642657	0
g14481.t1	clumping factor A-like	ENSDARP00000140656.1	-288.0986037	1.12433E-11
g14483.t1	heme oxygenase-like	ENSDARP00000038993.6	5.157678589	3.55592E-11

g14484.t1	heme oxygenase-like	ENSDARP00000038993.6	2.949474841	5.88755E-05
g14486.t1	GTPase IMAP family member 8-like isoform X2	ENSDARP00000025704.8	-2.300107593	0.017875971
g14491.t1	muscle-related coiled-coil	ENSDARP00000132655.1	-47.63383289	0.000953699
g14534.t1	complement C4-like	ENSDARP00000115743.1	-2.195401363	0.006091002
g14552.t1	Elastin	ENSDARP00000140438.1	12.65545696	0
g14556.t1	Tob1-like	ENSDARP00000049239.5	-2.122844273	0.00038572
g14564.t1	hyaluronidase-2-like	ENSDARP00000078315.4	3.17496132	0.049814662
g14629.t1	ADP ATP translocase 1	ENSDARP00000137238.1	-87.32869363	0
g14631.t1	long-chain-fatty-acid-- ligase 1 isoform X1	ENSDARP00000105676.1	-2.713228164	0.001923707
g14634.t1	PREDICTED: uncharacterized protein LOC108260474	ENSDARP00000130493.1	8.909788646	0
g14668.t1	transforming growth factor beta-3	ENSDARP00000025519.8	2.277748289	0.026766696
g14683.t1	influenza virus NS1A-binding	ENSDARP00000061658.4	-2.077818104	0.012406857
g14691.t1	transmembrane protease serine 13-like isoform X2	ENSDARP00000079274.5	-23.43198244	1.7931E-13
g14717.t1	von Willebrand factor A domain-containing 7-like	ENSDARP00000092438.3	-23.19770682	1.39217E-05
g14755.t1	complement C1q 4	ENSDARP00000107225.1	-4.605383218	1.74751E-06
g14757.t1	peroxisome proliferator-activated receptor gamma coactivator 1-alpha-like isoform X2	ENSDARP00000088481.3	-23.32500079	0.003373551
g14762.t1	disintegrin and metallo ase domain-containing 33-like	ENSDARP00000077690.4	2.906340852	0.005532973
g14794.t1	glycogenin-1-like isoform X2	ENSDARP00000097305.4	4.124554807	0
g14795.t1	glycogenin-1	ENSDARP00000097305.4	3.912725995	0
g14802.t1	NDRG1-like isoform X1	ENSDARP00000093670.4	-4.973347289	0
g14806.t1	bone morphogenetic 8A	ENSDARP00000094030.4	2.67149517	0.009262591
g14843.t1	bone morphogenetic receptor type-1A-like	ENSDARP00000141520.1	4.621841557	9.44966E-05
g14874.t1	sodium potassium-transporting ATPase subunit beta-3-like	ENSDARP00000062873.2	2.220229994	0.042272953
g14899.t1	alpha cardiac	ENSDARP00000133299.1	-57.05007027	0
g14910.t1	pituitary homeobox 1 isoform X1	ENSDARP00000062802.4	-9.182425617	0.004532333
g14961.t1	monocarboxylate transporter 7-like	ENSDARP00000126253.1	4.754548883	6.60025E-07
g14965.t1	hepatocyte growth factor isoform X2	ENSDARP00000132446.1	4.317119258	0
g14979.t1	musculoskeletal embryonic nuclear 1	ENSDARP00000109538.1	-28.69508005	0.006903476
g14981.t1	inter-alpha-trypsin inhibitor heavy chain H3-like	ENSDARP00000113767.1	4.1214489	1.52933E-05
g14982.t1	ribonuclease-like 3	ENSDARP00000123932.1	4.670758107	0.022607901
g14988.t1	latent-transforming growth factor beta-binding 3 isoform X1	ENSDARP00000048106.7	2.115568301	0.009012587
g15010.t1	troponin skeletal muscle-like	ENSDARP00000116801.1	-28.53110817	2.19733E-11
g15015.t1	troponin slow skeletal and cardiac muscles-like	ENSDARP00000054663.4	-1.7977E+308	0
g15052.t1	sulfotransferase family cytosolic 2B member 1-like	ENSDARP00000136304.1	-6.802421918	1.98792E-05
g15053.t1	neural cell adhesion molecule 1-like isoform X1	ENSDARP00000128855.1	5.265747159	0
g15132.t1	RAS guanyl-releasing 1-like isoform X2	ENSDARP00000082937.4	-2.623136106	0.012075457
g15154.t1	fumarylacetoacetase	ENSDARP00000136083.2	9.975577677	2.73736E-06
g15155.t1	cortexin-2-like	ENSDARP00000041584.5	1.7977E+308	0.028337546
g15156.t1	aryl hydrocarbon receptor nuclear translocator 2 isoform X1	ENSDARP00000136894.1	7.14144174	0.00279429
g15161.t1	beta-enolase	ENSDARP00000120742.2	-29.34607291	0
g15221.t1	ciliary neurotrophic factor receptor subunit alpha-like	ENSDARP00000114970.1	8.522897199	7.2766E-06
g15242.t1	E3 ubiquitin- ligase TRIM39-like	ENSDARP00000137535.1	-27.74920519	0
g1526.t1	atypical kinase mitochondrial	ENSDARP00000024635.5	-2.169531911	0.027555554

g15272.t1	sodium potassium-transporting ATPase subunit alpha-1-like	ENSDARP00000075490.4	-11.76547115	0
g15276.t1	type II cytoskeletal 8-like	ENSDARP00000045668.7	2.721463235	0
g15279.t1	sodium- and chloride-dependent taurine transporter-like	ENSDARP00000123063.1	-13.70190073	0.022577313
g15305.t1	troponin cardiac muscle isoforms-like	ENSDARP00000130469.1	-81.26446671	3.06314E-12
g15356.t1	tyrosine- kinase receptor Tie-2-like	ENSDARP00000055680.5	3.026964002	0.024649469
g15394.t1	procollagen-lysine,2-oxoglutarate 5-dioxygenase 2 isoform X2	ENSDARP00000097266.1	4.769859785	0.000116648
g15478.t1	hyaluronan and proteoglycan link 1-like isoform X2	ENSDARP00000136494.1	8.042033442	0
g15545.t1	suppressor SRP40-like		-1.7977E+308	0.005626205
g15551.t1	heat shock beta-1-like	ENSDARP00000060161.4	-9.182425617	4.66783E-14
g15591.t1	ATP-sensitive inward rectifier potassium channel 1-like	ENSDARP00000058986.6	-23.83898958	4.02678E-05
g15647.t1	interferon alpha-inducible 27 2A	ENSDARP00000113098.1	4.945508584	0
g15648.t1	Interferon alpha-inducible protein 27-like protein 1	ENSDARP00000123470.1	-2.526320616	9.66901E-07
g15659.t1	2-oxoisovalerate dehydrogenase subunit mitochondrial	ENSDARP00000076508.5	-2.150333067	8.08589E-06
g15669.t1	nucleolar and spindle-associated 1 isoform X1	ENSDARP00000015118.8	-2.180826084	0.01240092
g15678.t1	galectin-related B-like	ENSDARP00000080154.4	2.827050799	0.000203642
g15764.t1	ankyrin repeat domain-containing 1-like	ENSDARP00000098001.3	-80.34622415	3.77626E-05
g15861.t1	butyrophilin subfamily 3 member A3-like	ENSDARP00000127325.1	-6.601121472	1.73128E-11
g15868.t1	myosin-binding fast-type-like isoform X1	ENSDARP00000115895.1	-42.57306422	0.005656161
g15874.t1	hypothetical protein EH28_05532		2.000271791	0.000682461
g1601.t1	urea transporter 2-like	ENSDARP00000135209.1	-12.13595396	0
g1603.t1	urea transporter 2-like	ENSDARP00000135209.1	-20.32286846	0.000149522
g16061.t1	arachidonate 5-lipoxygenase-like	ENSDARP00000091900.3	1.7977E+308	6.84133E-06
g16125.t1	Complement factor H	ENSDARP00000071352.6	62.94633075	0.022718053
g16126.t1	---NA---		-22.35195709	0
g1613.t1	creatine kinase S- mitochondrial isoform X1	ENSDARP00000048209.3	-595.7098619	5.99645E-10
g16132.t1	branched-chain-amino-acid cytosolic-like isoform X2	ENSDARP00000122776.1	-2.294347848	6.77709E-10
g16165.t1	Krueppel-like factor 2	ENSDARP00000059187.4	2.749360853	7.20236E-10
g16198.t1	serine threonine- kinase D3-like	ENSDARP00000101844.2	-2.952780002	0.030607228
g16201.t1	alpha-actinin-3	ENSDARP00000000803.9	-3.70677541	3.02747E-07
g16222.t1	prostaglandin-H2 D-isomerase-like	ENSDARP00000122807.1	3.741161934	4.92978E-05
g16298.t1	rap guanine nucleotide exchange factor 5-like	ENSDARP00000100132.2	-5.130332494	0.009691021
g16301.t1	macrophage-expressed gene 1 -like	ENSDARP00000072103.4	-4.027379656	0.015150955
g16338.t1	c-binding -like	ENSDARP00000138855.1	-1.7977E+308	0.023293293
g1634.t1	gamma-synuclein isoform X1	ENSDARP00000073336.3	-21.23435924	2.10583E-07
g16348.t1	transmembrane 158		7.546187749	2.87105E-11
g16349.t1	CUB domain-containing 1-like	ENSDARP00000124759.1	4.289768067	0.002414122
g16350.t1	tetranectin-like	ENSDARP00000098343.2	15.80859964	0
g16367.t1	sarcoplasmic endoplasmic reticulum calcium ATPase 1-like isoform X1	ENSDARP00000043931.5	-2159.017823	0
g1642.t1	myozenin-1-like	ENSDARP00000060575.5	-22.95606404	0.004040236
g16423.t1	lipopolysaccharide-binding -like	ENSDARP00000139184.1	-2.130196168	5.76492E-06
g16428.t1	single-stranded DNA-binding 2	ENSDARP00000136811.1	3.514960362	0.013673174
g16513.t1	semaphorin-3ab-like isoform X2	ENSDARP00000062514.5	2.268571158	0.002165171
g16556.t1	PREDICTED: uncharacterized protein LOC108260524		-2.627435766	0.000299604

g16563.t1	eyes absent homolog 2	ENSDARP00000006502.6	-6.10148018	0.002767825
g16590.t1	mucin-19- partial	ENSDARP00000135368.1	-5.120823153	2.13302E-11
g16618.t1	proto-oncogene vav isoform X2	ENSDARP00000099666.3	-2.385224672	0.016304326
g16620.t1	microfibril-associated glyco 4-like	ENSDARP00000106408.3	2.593774657	0
g16657.t1	type I cytoskeletal 13-like	ENSDARP00000095806.5	-75.67272398	0
g16658.t1	type I cytoskeletal 13-like	ENSDARP00000133447.1	-3.073467718	0
g16711.t1	zinc fingers and homeoboxes isoform 2-like	ENSDARP00000025864.7	-3.080945437	0.009006542
g16743.t1	ephrin type-B receptor 5-like	ENSDARP00000113830.2	-10.47663229	6.91153E-08
g16788.t1	apolipo L6-like	ENSDARP00000134943.1	53.14499898	1.42775E-08
g16894.t1	serine threonine- kinase PLK1	ENSDARP00000075786.4	-2.501462413	0.008957697
g16947.t1	GTPase IMAP family member 7-like	ENSDARP00000126490.2	-1.7977E+308	1.15313E-09
g1697.t1	antimicrobial peptide NK-lysin-like	ENSDARP00000109920.1	-3.510271929	0.009691021
g16978.t1	colorectal cancer-associated 2 isoform X2	ENSDARP00000128784.1	-11.13794218	2.03981E-10
g17016.t1	transmembrane 45b	ENSDARP00000100942.2	-7.184397821	0.003619341
g17020.t1	thread biopolymer filament subunit alpha-like	ENSDARP00000140491.1	-1.7977E+308	2.21666E-06
g17053.t1	hematopoietic progenitor cell antigen CD34-like isoform X1	ENSDARP00000115242.1	3.576999131	6.61097E-06
g17056.t1	sec1 family domain-containing 1	ENSDARP00000121390.2	2.626518932	1.8181E-11
g17093.t1	MAL2	ENSDARP00000099973.1	3.199268952	5.76492E-06
g1711.t1	tyrosine- kinase Fes Fps isoform X1	ENSDARP00000124270.1	-4.5518518	0
g17117.t1	band 3 anion transport	ENSDARP00000005654.5	10.57203479	0.002489119
g17215.t1	sarcalumenin-like isoform X2	ENSDARP00000138572.1	-1.7977E+308	0
g17228.t1	ammonium transporter Rh type C	ENSDARP00000053520.5	-48.0777945	0
g1724.t1	pro convertase subtilisin kexin type 6 isoform X1	ENSDARP00000079749.4	7.537148158	1.30812E-05
g17270.t1	C-C chemokine receptor type 9-like	ENSDARP00000071978.3	-2.458566118	0.000557379
g17271.t1	thread biopolymer filament subunit gamma-like isoform X2	ENSDARP00000076687.4	-358.1145991	6.68931E-12
g17293.t1	rho GTPase-activating 20-like	ENSDARP00000106388.2	-2.425934656	0.023736746
g17330.t1	WNT1-inducible-signaling pathway 2	ENSDARP00000053992.6	2.332534734	1.49309E-05
g17354.t1	plasminogen activator inhibitor 1	ENSDARP00000073803.4	2.009448666	6.18055E-10
g17361.t1	T cell receptor delta partial	ENSDARP00000138884.1	-6.442508296	0.000217815
g17416.t1	telethonin-like	ENSDARP00000027408.6	-229.5606404	0
g17467.t1	low-density lipo receptor-like	ENSDARP00000115492.1	3.451622571	0.045059383
g17528.t1	kelch 35	ENSDARP00000087460.4	-9.314864448	0.03190225
g17539.t1	LIM domain transcription factor LMO4-like	ENSDARP00000138986.1	3.702725338	2.12241E-05
g17546.t1	ependymin-like	ENSDARP00000101021.2	-2.954036148	0.000445371
g17606.t1	manganese superoxide dismutase	ENSDARP00000062555.3	-2.15778774	3.2707E-09
g17620.t1	solute carrier family facilitated glucose transporter member 4	ENSDARP00000116768.2	-39.02530887	0.040076786
g17633.t1	C-C motif chemokine 14-like		-5.994083389	0.009008707
g17634.t1	C-C motif chemokine 4 homolog	ENSDARP00000138949.1	-30.70373566	4.8051E-09
g17635.t1	C-C motif chemokine 4 homolog	ENSDARP00000141835.1	-16.49088682	0
g17657.t1	phosphatidylinositol 4,5-bisphosphate 3-kinase catalytic subunit gamma isoform	ENSDARP00000002818.6	-2.199009105	0.021957951
g1766.t1	SH3 and PX domain-containing 2A isoform X3	ENSDARP00000078892.4	2.584839816	0.01409936
g1773.t1	carbonyl reductase [NADPH] 1	ENSDARP00000068939.5	-2.249947934	0.004636176
g17738.t1	pleckstrin homology domain-containing family G member 7 isoform X1	ENSDARP00000106435.2	-10.54738078	0.008016386

g17764.t1	PREDICTED: uncharacterized protein LOC108256173	ENSDARP00000106769.1	-66.18998465	6.30327E-06
g17783.t1	alpha-tectorin-like isoform X20	ENSDARP00000142113.1	-2.733724814	0
g17834.t1	BTB POZ domain-containing kctd15 isoform X1	ENSDARP00000122355.1	4.873439968	0.02185417
g17848.t1	inactive rhomboid 1-like isoform X1	ENSDARP00000040908.6	-4.37865666	0.034605497
g17874.t1	PREDICTED: uncharacterized protein LOC108256181	ENSDARP00000096709.4	-2.088118902	0.000963575
g17880.t1	CBL-interacting kinase 24-like	ENSDARP00000013115.8	-311.6285694	4.5821E-08
g17881.t1	CBL-interacting kinase 24-like	ENSDARP00000013115.8	-4.092898247	3.61822E-07
g17923.t1	mucin-2-like isoform X1	ENSDARP00000117940.2	4.414378734	9.86545E-08
g17959.t1	DENN domain-containing 2D-like isoform X1	ENSDARP00000096652.4	-2.14395365	0.009153627
g17990.t1	membrane-spanning 4-domains subfamily A member 12-like	ENSDARP00000123613.2	-2.802624736	5.26366E-05
g18008.t1	LIM homeobox Lhx8 isoform X1	ENSDARP00000028001.3	3.368754033	0.029967682
g18017.t1	serum deprivation-response	ENSDARP00000030479.6	-2.637777977	7.33007E-13
g18084.t1	2-oxoisovalerate dehydrogenase subunit mitochondrial	ENSDARP00000059346.4	-2.582557205	0.047008301
g18114.t1	primary amine liver isozyme-like	ENSDARP00000121793.1	6.065752814	0
g1814.t1	sciellin isoform X1	ENSDARP00000048547.5	-3.849262672	8.31305E-05
g18162.t1	PREDICTED: uncharacterized protein LOC108270429 isoform X1	ENSDARP00000129998.1	32.73333661	0.000323515
g18172.t1	collagen alpha-2(V) chain-like	ENSDARP00000031878.7	2.36919002	0.002559088
g18258.t1	myozenin-3 isoform X1	ENSDARP00000097622.3	-229.5606404	3.85804E-11
g1829.t1	coiled-coil domain-containing 80	ENSDARP00000000005.6	3.364854537	0.003020431
g18293.t1	hyaluronan synthase 1	ENSDARP00000128010.1	62.14770372	7.31982E-05
g18354.t1	SH2 domain-containing 1A-like	ENSDARP00000102265.3	-3.962128361	1.57906E-10
g18359.t1	purine nucleoside phosphorylase-like	ENSDARP00000093456.4	11.00917266	0.000141212
g18392.t1	LRR and PYD domains-containing 4E-like isoform X1	ENSDARP00000126848.1	-4.139047911	0.00112291
g18417.t1	proheparin-binding EGF-like growth factor	ENSDARP00000099541.3	2.064508795	0.003516958
g18426.t1	ornithine mitochondrial	ENSDARP00000125055.2	-3.122319965	2.6793E-05
g18561.t1	mucin-2-like isoform X1	ENSDARP00000108535.1	-3.120241714	0.041437924
g18592.t1	insulin-like growth factor II	ENSDARP00000047568.4	3.557520423	3.71953E-05
g18609.t1	phosphate-regulating neutral endopeptidase	ENSDARP00000084443.3	4.974846661	3.99444E-07
g18611.t1	jagunal homolog 1-A	ENSDARP00000006367.6	172.3496929	0
g18614.t1	alpha-actinin-2	ENSDARP00000115698.1	-20.44524454	2.2746E-05
g18617.t1	stathmin-like	ENSDARP00000023151.9	-2.098508885	0.021474379
g18629.t1	rap1 GTPase-activating 1-like isoform X6	ENSDARP00000109551.2	3.099489251	0.010007607
g18636.t1	n-acylneuraminate cytidylyltransferase	ENSDARP00000074418.4	-8.075615386	0.006406883
g18642.t1	sialic acid synthase-like	ENSDARP00000067085.5	-17.38921851	3.32581E-05
g18706.t1	mucolipin-3 isoform X2	ENSDARP00000098991.2	-4.531843677	0.039159566
g18775.t1	angiopoietin-2-like	ENSDARP00000070502.4	4.356147457	1.85752E-05
g18776.t1	angiopoietin-2-like	ENSDARP00000070502.4	4.327860785	6.91153E-08
g18786.t1	PREDICTED: uncharacterized protein LOC108438151		-9.252699282	2.70171E-08
g18819.t1	traf2 and NCK-interacting kinase-like		-5.565106434	0
g18820.t1	traf2 and NCK-interacting kinase-like	ENSDARP00000007356.7	-4.840586718	0
g18890.t1	glucagon receptor-like isoform X1	ENSDARP00000139469.1	-3.272190344	0.005857055
g189.t1	bone morphogenetic 4	ENSDARP00000069635.3	3.314921967	0.005731973
g18914.t1	tripartite motif-containing 35-like	ENSDARP00000062165.4	-20.87760959	6.26037E-10

g18938.t1	myosin-7-like	ENSDARP00000137611.1	-3115.13789	0
g18939.t1	myosin-7-like	ENSDARP00000137611.1	-1.7977E+308	0
g18940.t1	myosin-7-like	ENSDARP00000141470.1	-1.7977E+308	9.44611E-05
g18967.t1	B-cell lymphoma leukemia 11B-like isoform X1	ENSDARP00000088314.3	-3.535233862	0.02500716
g18991.t1	FAM234B	ENSDARP00000084079.4	-2.620090854	0.043588219
g19041.t1	alpha-tectorin-like	ENSDARP00000129792.1	-22.68940269	0
g19079.t1	sarcalumenin isoform X2	ENSDARP00000138572.1	-215.787002	0.003564587
g19159.t1	C-factor-like	ENSDARP00000067677.6	-3.295194778	3.09206E-12
g19175.t1	runt-related transcription factor 3-like	ENSDARP00000117722.1	-2.924339368	0.000209812
g19177.t1	myelin-associated glyco -like isoform X2	ENSDARP00000127576.1	5.570588566	2.25005E-09
g19198.t1	sodium- and chloride-dependent GABA transporter 2-like	ENSDARP00000082132.5	-48.39903502	1.60985E-11
g19251.t1	E3 ubiquitin- ligase RNF213-like	ENSDARP00000105582.1	-5.088594196	5.30315E-06
g19285.t1	transcription factor jun-B-like	ENSDARP00000109347.1	4.236332006	0
g19292.t1	cytosolic acyl coenzyme A thioester hydrolase isoform X2	ENSDARP00000118830.1	2.463177067	0.00085243
g19301.t1	butyrophilin subfamily 1 member A1-like isoform X2	ENSDARP00000141552.1	-6.020110672	0.000153093
g19355.t1	CYR61-like	ENSDARP00000092583.2	7.776263229	3.38943E-11
g19380.t1	zinc finger 385D-like isoform X1	ENSDARP00000110557.1	-6.244049419	0.032835171
g19382.t1	proteasome subunit beta type-7	ENSDARP00000112303.2	2.690324402	5.40864E-05
g19390.t1	cytosolic phospholipase A2 zeta-like	ENSDARP00000129605.1	48.73760273	3.59841E-06
g19397.t1	drebrin isoform X2	ENSDARP00000026512.7	-2.155227595	0.008385765
g19434.t1	insulin-like growth factor-binding 5	ENSDARP00000057368.5	-9.074819067	5.24481E-05
g19499.t1	type I cytoskeletal 13-like	ENSDARP00000000221.8	-111.9629851	0
g19500.t1	type I cytoskeletal 13-like	ENSDARP00000000221.8	-219.2304116	1.29778E-06
g19501.t1	type I cytoskeletal 13-like	ENSDARP00000096584.4	-758.9848674	0
g19512.t1	filamin-C isoform X1	ENSDARP00000138996.1	-18.59830273	0.009497236
g19532.t1	Proline-rich partial	ENSDARP00000090128.3	-3.131721668	0.000907439
g19568.t1	intelectin-like	ENSDARP00000052386.5	-26.4203428	0
g19582.t1	cytochrome P450 1A1-like	ENSDARP00000139599.1	2.489948633	0.003623094
g19674.t1	E3 ubiquitin- ligase LNX isoform X1	ENSDARP00000131036.1	-2.728913637	0.029774104
g19693.t1	tripartite motif-containing 66-like isoform X1	ENSDARP00000070982.3	-2.572185489	0.041926466
g1970.t1	adenosylhomocysteinase 3	ENSDARP00000073256.4	-54.65309093	2.7178E-10
g19817.t1	deoxynucleoside triphosphate triphosphohydrolase SAMHD1-like isoform X1	ENSDARP00000096087.3	-6.518181688	6.41427E-13
g19900.t1	nexilin isoform X2	ENSDARP00000124483.1	-1.7977E+308	0.007704866
g19938.t1	nesprin-2 isoform X8	ENSDARP00000131353.1	3.979551483	1.36284E-13
g1994.t1	cytochrome b-c1 complex subunit mitochondrial-like	ENSDARP00000011707.8	-3.303987183	1.04763E-11
g19964.t1	serine protease HTRA1	ENSDARP00000048135.5	2.270056125	0.007115891
g19970.t1	PREDICTED: uncharacterized protein LOC108179621		9.279607141	0
g19971.t1	PREDICTED: uncharacterized protein LOC108263565		6.093711892	0
g19997.t1	isocitrate dehydrogenase [NADP] mitochondrial	ENSDARP00000023774.7	-2.667440087	0
g20.t1	follistatin-related 3-like	ENSDARP00000038299.7	4.113852865	2.50674E-10
g20073.t1	eva-1 homolog C-like	ENSDARP00000088660.4	4.173797098	0.040992406
g20099.t1	equilibrative nucleoside transporter 1-like	ENSDARP00000137952.1	2.460662217	6.88072E-05
g20106.t1	focal adhesion kinase 1	ENSDARP00000137309.1	2.103886064	0.009139249

g20182.t1	S100-A1-like	ENSDARP00000033294.6	2.492378039	0.000753391
g2020.t1	elongation factor 2	ENSDARP00000051080.4	-2.366500131	0.003535851
g20204.t1	pentraxin fusion -like	ENSDARP00000104927.2	-7.519282201	0.000474162
g20207.t1	myosin light chain 3	ENSDARP00000061954.5	-1119.682024	0
g2022.t1	atrial natriuretic peptide receptor 3 isoform X1	ENSDARP00000139195.1	3.413270804	1.0465E-05
g20223.t1	ribonuclease-like 3	ENSDARP00000123932.1	-27.10090894	0
g20262.t1	forkhead box C1	ENSDARP00000105372.1	2.280170935	0.006091002
g20286.t1	immediate early response gene 2 -like	ENSDARP00000135138.1	2.299753825	3.15956E-05
g20295.t1	NXPE family member 3	ENSDARP00000119028.1	-6.568966018	0.00987731
g20356.t1	caspase 8	ENSDARP00000111270.2	-4.387804646	7.71135E-06
g20465.t1	carnitine O-palmitoyltransferase liver isoform-like isoform X1	ENSDARP00000077866.5	-6.033864539	0.000746683
g20558.t1	GTPase IMAP family member 7-like	ENSDARP00000092414.3	2.207344285	0.014198121
g20585.t1	epidermal differentiation-specific -like	ENSDARP00000046171.7	-20.33089038	0
g20638.t1	ferric-chelate reductase 1	ENSDARP00000082157.4	2.025794274	0.011066208
g20674.t1	PREDICTED: uncharacterized protein LOC108440808		-20.66045764	0.004543458
g20729.t1	BOLA class I histocompatibility alpha chain BL3-7-like isoform X1	ENSDARP00000117273.2	-2.053963625	0.002548596
g2076.t1	tubulin beta-4B chain	ENSDARP00000061148.3	-228.4128372	0.001428493
g20794.t1	immunoglobulin light partial	ENSDARP00000129656.1	-4.707938558	5.0389E-11
g20835.t1	hemoglobin subunit alpha-like	ENSDARP00000100779.1	103.3858996	0
g20836.t1	hemoglobin subunit beta-A	ENSDARP00000109518.1	6.368988006	3.54836E-11
g20837.t1	hemoglobin subunit alpha-like	ENSDARP00000100779.1	6.828925137	0
g20838.t1	hemoglobin subunit beta-like	ENSDARP00000109518.1	5.804913716	0
g2086.t1	prostaglandin E synthase	ENSDARP00000012269.6	2.502649421	0.003278971
g20865.t1	endothelial lipase-like	ENSDARP00000089620.3	2.502768053	0.032835171
g20869.t1	Krueppel-like factor 15	ENSDARP00000107271.2	6.717637921	0.006839599
g20887.t1	myosin heavy fast skeletal muscle	ENSDARP00000088845.4	-155.5828727	0
g20888.t1	myosin heavy fast skeletal muscle-like	ENSDARP00000088845.4	-16.70463589	5.63852E-10
g2089.t1	ubiquitin carboxyl-terminal hydrolase 2-like isoform X2	ENSDARP00000100330.3	-2.935429135	7.00802E-05
g20908.t1	ependymin-like	ENSDARP00000101021.2	-2.128005144	0
g20915.t1	apolipo L3-like	ENSDARP00000134943.1	194.2841766	0
g20939.t1	cytochrome P450 2B4-like	ENSDARP00000136547.1	-4.686532105	8.62588E-05
g20966.t1	immunoglobulin light partial	ENSDARP00000081168.5	-6.772038892	0.020544308
g21003.t1	macrophage mannose receptor 1-like	ENSDARP00000067412.3	7.032968503	0
g21068.t1	EH domain-containing 1-like	ENSDARP00000020181.7	-2.794418076	1.68847E-13
g21075.t1	tyrosine- phosphatase non-receptor type 18-like isoform X1	ENSDARP00000120732.1	-3.569196259	0.049748861
g211.t1	DNA fragmentation factor subunit alpha	ENSDARP00000072629.4	-2.030090483	0.003175267
g21168.t1	hemoglobin subunit beta-A-like	ENSDARP00000109518.1	5.372581864	0
g21169.t1	hemoglobin subunit alpha-like	ENSDARP00000100779.1	4.412941315	0
g21170.t1	hemoglobin subunit beta-A-like	ENSDARP00000109518.1	3.353922993	0
g21171.t1	hemoglobin subunit alpha-like	ENSDARP00000100779.1	3.681484619	0
g21173.t1	multidrug resistance 1-like	ENSDARP00000121883.1	-12.78094376	1.53293E-06
g2122.t1	calcitonin gene-related peptide type 1 receptor	ENSDARP00000118958.1	2.332512661	0.000568467
g21266.t1	interferon-induced GTP-binding Mx1	ENSDARP00000094364.3	9.583524405	0.00070394

g21267.t1	globoside alpha-1,3-N-acetylgalactosaminyltransferase 1-like isoform X2	ENSDARP00000116300.1	-3.85710546	0
g21283.t1	Golgi-associated plant pathogenesis-related 1-like	ENSDARP00000139543.1	-142.3275971	5.8875E-06
g21311.t1	proto-oncogene c-Fos-like	ENSDARP00000043297.6	4.292166541	0
g21373.t1	extracellular matrix 2-like	ENSDARP00000119295.1	2.828933407	0.032409888
g21374.t1	interferon-induced 44-like	ENSDARP00000121495.1	-4.591212808	0.018264959
g21381.t1	LBH-like	ENSDARP00000108481.1	-3.015796649	0.001873991
g21468.t1	interferon regulatory factor 1-like	ENSDARP00000140878.1	-2.024116524	0
g21550.t1	GTPase IMAP family member 8-like	ENSDARP00000137196.1	-87.23304336	0.018745809
g21556.t1	serine protease 27-like	ENSDARP00000019814.6	-146.9188099	3.3888E-05
g21592.t1	myristoylated alanine-rich C-kinase substrate	ENSDARP00000106995.2	-2.283251149	4.85025E-11
g21638.t1	B-cadherin-like isoform X1	ENSDARP00000137597.1	-2.690017469	0
g21654.t1	ski	ENSDARP00000028344.5	2.52396484	7.8876E-05
g21674.t1	interferon regulatory factor 4-like	ENSDARP00000005897.7	-2.124330648	0.009139249
g21724.t1	MGARP isoform X2	ENSDARP00000133493.1	-7.396953969	0.005265435
g21741.t1	epidermal differentiation-specific -like	ENSDARP00000096371.4	-12.19540902	9.39369E-05
g21742.t1	epidermal differentiation-specific -like	ENSDARP00000096371.4	-10.28400856	0
g21743.t1	phospholipase B-like 1	ENSDARP00000140223.1	-2.287384318	2.66614E-05
g21748.t1	claudin-10-like	ENSDARP00000132047.1	-11.47803202	0.03869397
g21773.t1	arachidonate 12- 12S-type-like	ENSDARP00000091900.3	1.7977E+308	4.7216E-12
g21786.t1	cordon-bleu-like isoform X3	ENSDARP00000126238.1	3.441894287	0.000220399
g21809.t1	myosin regulatory light chain cardiac muscle isoform-like	ENSDARP00000134838.1	-1.7977E+308	0
g21833.t1	ATP synthase F(0) complex subunit mitochondrial-like	ENSDARP00000113590.1	-2.474835567	0
g21848.t1	cadherin-5-like	ENSDARP00000104309.1	2.240856342	0.007355853
g2188.t1	mitochondrial amidoxime-reducing component 1-like	ENSDARP00000094269.4	-2.859942979	1.65693E-05
g21891.t1	dickkopf-related 4-like	ENSDARP00000017197.5	71.78931009	0
g21903.t1	intestinal mucin	ENSDARP00000130459.1	-32.3680503	1.1544E-12
g21909.t1	angiopoietin-related 5-like	ENSDARP00000073617.3	-5.325806858	0.005409428
g21914.t1	middle subunit-like	ENSDARP00000055286.5	2.466721777	0
g21970.t1	chorion-specific transcription factor GCMb	ENSDARP00000113822.1	-3.633068075	1.8181E-11
g21992.t1	lysoplasmalogenase TMEM86A	ENSDARP00000064645.5	-1.7977E+308	0.005142765
g22054.t1	ferric-chelate reductase 1	ENSDARP00000130544.1	-21.56230301	0.010040497
g22068.t1	GLIPR1 1	ENSDARP00000041428.4	6.526300917	5.36783E-07
g22091.t1	DENN domain-containing 2D-like	ENSDARP00000027732.8	2.344604636	0.000152693
g22167.t1	glutathione-specific gamma-glutamylcyclotransferase 1	ENSDARP00000094107.4	2.402977582	0
g22204.t1	uncharacterized oxidoreductase -like	ENSDARP00000067677.6	-6.542478252	0.001767565
g22223.t1	interferon-induced 44-like	ENSDARP00000022947.7	-3.905719229	0.030437723
g22324.t1	sacsin-like	ENSDARP00000140359.1	-32.06675196	9.66901E-07
g22325.t1	sacsin-like	ENSDARP00000140359.1	-15.63646657	0
g22336.t1	T-cell receptor beta partial	ENSDARP00000140922.1	-3.239594084	3.84547E-05
g22399.t1	E3 ubiquitin- ligase TRIM39-like isoform X1	ENSDARP00000137909.1	-3.179546801	0.041437924
g22441.t1	provirus ancestral Env partial	ENSDARP00000133661.1	-5.921621065	0.006504388
g22500.t1	cingulin-like isoform X2	ENSDARP00000130472.1	-2.358788232	0.030081727
g22544.t1	signal CUB and EGF-like domain-containing 2	ENSDARP00000054365.5	5.702593035	0.011169927

g22553.t1	collagenase 3-like	ENSDARP00000138624.1	-24.28321149	0
g22610.t1	interferon-induced 44-like	ENSDARP00000022947.7	-3.581145991	0.001432853
g22662.t1	interferon alpha beta receptor 2-like isoform X1	ENSDARP00000135777.1	2.470863967	0.025888891
g22669.t1	jun dimerization 2-like	ENSDARP00000058706.3	7.405450677	3.78157E-06
g22710.t1	trifunctional enzyme subunit mitochondrial	ENSDARP00000067447.4	-2.109977854	0.000229554
g22711.t1	trifunctional enzyme subunit mitochondrial-like	ENSDARP00000140496.1	-9.655050465	0.00033987
g22837.t1	fibrillin-1	ENSDARP00000134527.1	2.173104739	0.001305065
g22850.t1	leukocyte elastase inhibitor-like	ENSDARP00000072235.5	-101.0066818	0.016910434
g2286.t1	E3 ubiquitin- ligase RNF182-like	ENSDARP00000111797.1	2.562439681	0.002084785
g2288.t1	very long-chain specific acyl- mitochondrial	ENSDARP00000112163.1	-2.503040718	0.001693615
g2289.t1	very long-chain specific acyl- mitochondrial	ENSDARP00000112163.1	-2.60107016	1.43531E-05
g22895.t1	sortilin-like isoform X2	ENSDARP00000073048.3	-17.85471648	0.004532333
g22907.t1	von Willebrand factor A domain-containing 1-like	ENSDARP00000099327.3	-6.006836758	0.007844014
g22919.t1	long-chain-fatty-acid-- ligase 4-like	ENSDARP00000015290.7	-2.357449249	0.000646668
g22921.t1	metalloprotease TIKI2-like	ENSDARP00000107884.2	3.133615751	0.011355999
g22946.t1	serine protease 27-like	ENSDARP00000019814.6	-640.4741868	1.26052E-13
g22978.t1	prostaglandin reductase 1-like	ENSDARP00000038835.6	-2.340914425	0.012989961
g22999.t1	endogenous retrovirus group K member 11 Pol - partial	ENSDARP00000109033.2	-5.817965437	0
g23052.t1	mannose-specific lectin-like		-53.5163243	0
g231.t1	transcriptional activator Myb isoform X2	ENSDARP00000070211.5	-2.002462421	0.041984704
g23208.t1	zinc finger GLIS1		16.84377017	0.007208029
g23265.t1	pentraxin fusion -like	ENSDARP00000101055.2	-3.631031284	9.89112E-05
g23278.t1	C-C motif chemokine 4 homolog	ENSDARP00000138949.1	4.672958181	0.023228606
g23294.t1	nesprin-2 isoform X5	ENSDARP00000119338.1	4.686128142	3.00271E-12
g23404.t1	histamine N-methyltransferase-like	ENSDARP00000004085.5	-3.765354407	0.001430554
g23406.t1	immunoglobulin light chain precursor	ENSDARP00000120591.1	-6.10148018	0
g23431.t1	Ig kappa chain V-IV region STH	ENSDARP00000135064.1	-3.790771102	5.22648E-08
g23433.t1	myosin heavy fast skeletal muscle-like	ENSDARP00000088845.4	-6.341612692	0.000699801
g23434.t1	myosin heavy fast skeletal muscle- partial	ENSDARP00000088845.4	-6.175181227	0.014229474
g23440.t1	interferon-induced 44-like	ENSDARP00000022947.7	-2.20285463	0.002306965
g23475.t1	transmembrane protease serine 9-like	ENSDARP00000138127.1	3.799776008	0
g23507.t1	interferon-induced GTP-binding Mx1-like isoform X2	ENSDARP00000131851.1	10.08136983	2.64558E-08
g23519.t1	prohibitin-2-like isoform X2	ENSDARP00000022440.6	-2.009621768	0.041926466
g23555.t1	bone morphogenetic 2-like	ENSDARP00000133018.1	6.378113512	0
g23565.t1	peptidyl-prolyl cis-trans isomerase FKBP1B	ENSDARP00000021881.6	3.982763389	0.018745809
g2357.t1	basal cell adhesion molecule-like isoform X2	ENSDARP00000105899.2	3.04930322	0.012989961
g23584.t1	alpha-(1,3)-fucosyltransferase 9-like	ENSDARP00000139142.1	-2.048347211	0.005819689
g2359.t1	apolipo C-I	ENSDARP00000118222.1	2.677215625	6.33262E-13
g23603.t1	C-X-C chemokine receptor type 4-like	ENSDARP00000074800.4	4.531489241	0
g23605.t1	ferric-chelate reductase 1		-20.84344076	0
g23634.t1	transcription factor SOX-10	ENSDARP00000100237.2	3.18215027	0.0106182
g23649.t1	coiled-coil domain-containing 86-like	ENSDARP00000125424.1	-2.410297816	0
g23663.t1	transcription factor jun-B-like	ENSDARP00000131822.1	2.371484324	0

g23686.t1	Ig kappa chain V-IV region STH	ENSDARP00000135064.1	-3.222678221	0.046299476
g23694.t1	macrophage mannose receptor 1-like	ENSDARP00000056912.5	-4.02755945	0.000166465
g23700.t1	monooxygenase p33MONOX	ENSDARP00000099389.2	6.147008078	0.002182941
g23786.t1	butyrophilin subfamily 1 member A1-like isoform X28	ENSDARP00000133849.1	-2.923965821	0.001959458
g23791.t1	pyruvate kinase PKM-like isoform X2	ENSDARP00000104356.2	-2.115053092	0.02038094
g23808.t1	repressor of RNA polymerase III transcription MAF1 homolog	ENSDARP00000089819.3	-2.063312899	0.010299442
g23816.t1	cell surface glyco 1-like		-33.23638837	1.05174E-07
g23872.t1	Ig kappa chain V-IV region STH	ENSDARP00000081172.5	-2.446137972	5.67953E-05
g23874.t1	PREDICTED: uncharacterized protein LOC107075374	ENSDARP00000137908.1	-57.65503777	0
g23906.t1	myosin-7-like	ENSDARP00000131677.1	-1060.570159	1.18877E-14
g23907.t1	Slow myosin heavy chain 1	ENSDARP00000090306.4	-1.7977E+308	0.000435434
g23930.t1	Immunoglobulin light partial	ENSDARP00000119905.2	-3.820544944	0.00012298
g2395.t1	ephrin type-A receptor 2-like	ENSDARP00000011069.6	2.47586686	0.007673976
g24086.t1	nesprin-2 isoform X7	ENSDARP00000126725.1	4.898245807	0.004174068
g24095.t1	prostaglandin G H synthase 2-like	ENSDARP00000022450.7	2.083723991	0.033664208
g24100.t1	suppressor of cytokine signaling 3-like	ENSDARP00000038723.4	2.151852258	9.66999E-09
g24182.t1	myosin heavy fast skeletal muscle-like	ENSDARP00000088845.4	-18.36485123	0.021399833
g24191.t1	intestinal mucin	ENSDARP00000130459.1	-28.12117845	0.009482339
g24224.t1	myozenin-2 isoform X1	ENSDARP00000120440.1	-1.7977E+308	5.97628E-05
g24275.t1	MHC class II beta chain	ENSDARP00000137499.1	-2.074875019	0.004088079
g2430.t1	transforming growth factor beta-2	ENSDARP00000123937.1	4.107224745	0.008218164
g24326.t1	hepacam family member 2	ENSDARP00000081878.5	-11.86063309	0.001794306
g24388.t1	C19orf12 homolog	ENSDARP00000141027.1	-2.028675427	4.40557E-05
g24522.t1	collagen alpha-1(X) chain	ENSDARP00000130066.1	-29.73281993	0
g24618.t1	myosin heavy fast skeletal muscle	ENSDARP00000088845.4	-4.463679119	0.018676061
g24629.t1	actin filament-associated 1-like 2 isoform X2	ENSDARP00000134907.1	2.573723819	0.000526842
g24693.t1	TRPM8 channel-associated factor homolog	ENSDARP00000120589.2	2.906702877	0
g24709.t1	nesprin-2 isoform X7	ENSDARP00000113565.1	5.082172033	0.00074136
g2474.t1	aconitate mitochondrial-like	ENSDARP00000123883.2	-2.238012973	0.004090618
g2487.t1	lysozyme milk isozyme-like		-42.38673253	0
g2488.t1	filaggrin-like isoform X8		361.3974818	0
g2498.t1	alpha cardiac muscle 1	ENSDARP00000133299.1	-83.02443162	0.000184808
g25.t1	granzyme B-like	ENSDARP00000107249.1	-4.256436874	0.046667052
g2514.t1	tubulin alpha-1B chain	ENSDARP00000072006.4	2.606648805	0
g2561.t1	transcription factor SOX-2	ENSDARP00000095266.3	4.775628323	1.10477E-07
g2583.t1	lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase mitochondrial	ENSDARP00000137558.1	-2.32981246	0.027951289
g2586.t1	vascular cell adhesion 1-like	ENSDARP00000137248.1	-3.138142797	0.002352967
g2593.t1	podoplanin isoform X1		6.118867591	0
g2619.t1	histone-lysine N-methyltransferase SMYD1-like isoform X2	ENSDARP00000067092.6	-72.31160173	0.039132397
g2676.t1	transmembrane and coiled-coil domains 3	ENSDARP00000107843.2	3.982763389	0.016839379
g268.t1	immunoglobulin heavy partial	ENSDARP00000140922.1	-3.692243308	0
g2686.t1	troponin slow skeletal muscle-like	ENSDARP00000141905.1	-123.5036245	0
g2774.t1	protocadherin-20	ENSDARP00000052911.5	-8.003175752	6.03863E-06

g2788.t1	polycystic kidney disease 1-like 2	ENSDARP00000142251.1	-1.7977E+308	0.005396085
g2801.t1	NAD(P)H dehydrogenase [quinone] 1-like	ENSDARP00000021150.6	-4.040640541	1.11144E-05
g2802.t1	NAD(P)H dehydrogenase [quinone] 1-like	ENSDARP00000021150.6	-9.391117108	7.07318E-06
g2868.t1	55 kDa erythrocyte membrane	ENSDARP00000131512.1	2.004673684	0.049250924
g2870.t1	V-set and immunoglobulin domain-containing 1-like	ENSDARP00000089974.3	130.6844237	0.007197169
g2924.t1	transcription factor SOX-11-like	ENSDARP00000122216.1	2.265196678	0.025788057
g2926.t1	DNA-binding inhibitor ID-2	ENSDARP00000072092.3	2.624060254	0.028018258
g2934.t1	alpha cardiac	ENSDARP00000133299.1	-514.2158345	0
g2936.t1	gremlin-1-like	ENSDARP00000043381.4	3.828824344	0.009150685
g2945.t1	thrombospondin-1-like	ENSDARP00000012255.8	2.515587566	0.007670428
g2970.t1	pendrin-like	ENSDARP00000131698.1	-29.11246304	0
g299.t1	insulin-like growth factor-binding 7	ENSDARP00000139497.1	3.616312631	0
g3019.t1	carbonic anhydrase-like	ENSDARP00000034327.6	-7.475436239	1.59522E-07
g307.t1	probable basic-leucine zipper transcription factor H	ENSDARP00000137886.1	-1.7977E+308	1.19773E-07
g3082.t1	growth arrest and DNA damage-inducible GADD45 beta-like	ENSDARP00000028333.6	2.216285548	1.02761E-08
g3109.t1	FYN-binding -like isoform X1	ENSDARP00000096399.4	-3.013505134	0.000824473
g3160.t1	ankyrin repeat and MYND domain-containing 1	ENSDARP00000034733.6	2.035870465	0.005148245
g3202.t1	guanylate-binding 1-like	ENSDARP00000114509.1	-2.262421618	0.002576291
g321.t1	golgin subfamily A member 6 22 isoform X1	ENSDARP00000131828.1	-3.75046524	0.012723495
g3277.t1	secreted frizzled-related 3	ENSDARP00000026312.6	2.120194846	0.000140976
g329.t1	receptor activity-modifying 3-like	ENSDARP00000110442.2	2.294499291	1.59522E-07
g3331.t1	tetraspanin 4 variant partial	ENSDARP00000123336.1	6.81608955	0.006307248
g3336.t1	myelin basic -like isoform X2	ENSDARP00000126314.1	-66.57258572	9.54275E-05
g3356.t1	mothers against decapentaplegic homolog 3	ENSDARP00000045373.5	2.837319574	0.004278738
g3380.t1	forkhead box F1	ENSDARP00000022888.8	2.369452592	0.007115891
g3398.t1	chloride channel 2-like	ENSDARP00000123262.2	-26.31892606	0
g3473.t1	probable N-acetyltransferase CML1	ENSDARP00000098162.2	-5.947707502	0.001200296
g3493.t1	Golgi resident GCP60-like	ENSDARP00000111270.2	2.078567822	0
g3495.t1	calsequestrin-1-like isoform X2	ENSDARP00000136986.1	-20.23003144	0.0005225
g3504.t1	lysophospholipid acyltransferase LPCAT4	ENSDARP00000050686.5	2.720275061	0.000115242
g3527.t1	cis-aconitate decarboxylase-like	ENSDARP00000085591.4	94.21724643	3.40874E-10
g3531.t1	cis-aconitate decarboxylase-like	ENSDARP00000085591.4	36.17895099	1.51725E-09
g3532.t1	tRNA threonylcarbamoyladenosine biosynthesis	ENSDARP00000138061.1	21.73233565	6.71095E-06
g3533.t1	probable alpha-ketoglutarate-dependent hypophosphite dioxygenase	ENSDARP00000099599.2	28.0843389	0
g3534.t1	threonylcarbamoyl-AMP synthase	ENSDARP00000097499.3	36.10762226	3.13247E-08
g3535.t1	ES1 mitochondrial-like	ENSDARP00000075482.3	8.000000316	0
g3562.t1	chondroadherin-like	ENSDARP00000066263.3	7.318327728	4.25273E-05
g3600.t1	cytokine 1	ENSDARP00000107063.1	3.506038681	3.89767E-10
g3601.t1	homeobox MSX-1	ENSDARP00000012308.3	3.401425139	9.65435E-11
g3626.t1	lysozyme g-like	ENSDARP00000073908.4	2.509519731	0.015069766
g3636.t1	early growth response 1	ENSDARP00000054459.5	2.736782123	1.0523E-07
g3652.t1	phosphoglycerate mutase 2	ENSDARP00000074719.3	-75.59103945	5.80814E-14
g3711.t1	interferon-induced with tetratricopeptide repeats 5-like isoform X1	ENSDARP00000007361.7	2.881241975	5.80903E-11

g3757.t1	neuropilin-1 isoform X2	ENSDARP00000140900.1	2.62315836	0.049960781
g3762.t1	5 -AMP-activated kinase subunit beta-2-like	ENSDARP00000064865.4	-2.415638112	0.021957951
g3771.t1	bone morphogenetic 2-like	ENSDARP00000037833.4	-49.01119673	0.000192486
g3841.t1	estradiol 17-beta-dehydrogenase 2-like	ENSDARP00000140406.1	-2.16250272	3.58953E-06
g3857.t1	L-lactate dehydrogenase B-B chain	ENSDARP00000013329.6	-8.212117755	2.19733E-11
g3889.t1	proline-rich 15-like	ENSDARP00000138276.1	2.638580745	0.034559149
g3908.t1	sodium hydrogen exchanger 3-like	ENSDARP00000109287.1	-74.30778121	0
g3932.t1	microtubule-associated tau-like isoform X1	ENSDARP00000128528.1	-6.096528483	0.042006174
g3970.t1	serine threonine- kinase Sgk1-like	ENSDARP00000009591.8	8.548370201	0
g398.t1	ADAMTS 4	ENSDARP00000087059.5	2.073364208	0.000739411
g3988.t1	regulator of G- signaling 5-like	ENSDARP00000004766.6	4.5489195	0
g3993.t1	Protein-glutamine gamma-glutamyltransferase K	ENSDARP00000135992.1	2.187065776	0.000711348
g4058.t1	ornithine decarboxylase antizyme 2-like	ENSDARP00000139868.1	-2.138595589	0.000486527
g411.t1	carbonic anhydrase 14-like	ENSDARP00000141999.1	-4.200723059	5.37162E-05
g414.t1	tropomyosin beta chain isoform X1	ENSDARP00000136122.1	-3.71988775	6.18255E-08
g4303.t1	EGF-containing fibulin-like extracellular matrix 2	ENSDARP00000120392.1	4.128870198	0.001923707
g4306.t1	fos-related antigen 1-like isoform X1	ENSDARP00000011823.6	4.892288683	5.67953E-05
g4331.t1	four and a half LIM domains 1-like isoform X1	ENSDARP00000096545.3	-3.909066421	0
g4352.t1	period circadian homolog 3 isoform X1	ENSDARP00000004561.7	-2.320202187	0.000394507
g4360.t1	F-box only 2-like	ENSDARP00000134933.1	-11.08071553	3.45655E-08
g4384.t1	cathepsin L1-like	ENSDARP00000106602.2	3.084958519	0
g4390.t1	doublesex- and mab-3-related transcription factor 2	ENSDARP00000002460.4	-6.675381781	1.33618E-06
g4467.t1	fibulin-1 isoform X1	ENSDARP00000136228.1	5.227376948	4.86782E-06
g4468.t1	fibulin-1-like isoform X2	ENSDARP00000136402.1	3.821026805	9.66057E-08
g4475.t1	growth differentiation factor 10-like	ENSDARP00000098571.2	7.796001828	2.08985E-05
g4482.t1	transcription factor Sox-9-A-like	ENSDARP00000006945.4	4.700490542	1.91061E-08
g4496.t1	syndecan-2-A precursor	ENSDARP00000115190.2	3.217020525	0
g4500.t1	exostosin-1c	ENSDARP00000135521.1	3.753327051	0.004684567
g456.t1	carbonic anhydrase 6	ENSDARP00000073464.4	-52.99024783	0.000739411
g4737.t1	mothers against decapentaplegic homolog 6	ENSDARP00000091342.4	2.790729761	6.92825E-13
g4739.t1	mothers against decapentaplegic homolog 6	ENSDARP00000091342.4	2.906858854	3.88025E-11
g4771.t1	rho guanine nucleotide exchange factor 4 isoform X1	ENSDARP00000057215.4	2.78218097	0.015723425
g4822.t1	glycogen muscle form-like	ENSDARP00000111947.1	-7.14032176	3.36568E-06
g4828.t1	cornifelin homolog	ENSDARP00000052855.3	-3.725945779	3.71993E-05
g4831.t1	cornifelin homolog B-like	ENSDARP00000123156.1	-8.67229086	6.74126E-08
g4833.t1	acetylcholine receptor subunit beta-like	ENSDARP00000034131.8	12.53602435	0.028019032
g4940.t1	forkhead box I1-like	ENSDARP00000072803.4	-19.93258244	0
g4957.t1	CASP isoform X7	ENSDARP00000075613.4	2.29986653	0.011312306
g4988.t1	NADPH oxidase activator 1-like	ENSDARP00000089913.4	4.470960646	1.29027E-08
g5049.t1	vitrin isoform X3	ENSDARP00000118990.1	-4.120201933	8.10683E-05
g5119.t1	aminopeptidase N-like	ENSDARP00000053443.5	2.096163264	0.027994582
g5120.t1	aminopeptidase N-like	ENSDARP00000053443.5	2.42699644	0.02216043
g5172.t1	granzyme K-like	ENSDARP00000106760.2	-6.045096864	0.045281425

g5175.t1	NAD(P) mitochondrial	ENSDARP00000032730.6	-2.939797434	2.84907E-09
g5221.t1	cytochrome c oxidase subunit mitochondrial-like	ENSDARP00000122388.1	-1.7977E+308	0.000353196
g5227.t1	tumor necrosis factor alpha-induced 2-like	ENSDARP00000129449.1	-3.57307775	2.64558E-08
g5229.t1	creatine kinase M-type-like	ENSDARP00000059365.5	-65.07078714	0
g5232.t1	sphingosine 1-phosphate receptor 1	ENSDARP00000062632.3	2.987072542	0.007890647
g524.t1	5-hydroxytryptamine receptor 3A-like	ENSDARP00000135876.1	10.55155717	2.66492E-11
g5242.t1	ciliary neurotrophic factor receptor subunit alpha-like isoform X1	ENSDARP00000032045.5	4.750275084	0.011003745
g5243.t1	ciliary neurotrophic factor receptor subunit alpha-like	ENSDARP00000114970.1	4.072050884	1.4352E-05
g5281.t1	ras-related rab-11b	ENSDARP00000111631.1	2.260785389	0.008930064
g5331.t1	semaphorin-3E isoform X1	ENSDARP00000053124.6	3.355017424	6.86303E-05
g5365.t1	basement membrane-specific heparan sulfate proteoglycan core -like	ENSDARP00000090925.3	144.6240956	0.004635956
g5373.t1	cdc42-interacting 4 homolog	ENSDARP00000040854.5	-4.282663561	0.001151516
g5391.t1	MDS1 and EVI1 complex locus EVI1-like isoform X2	ENSDARP00000080486.5	2.202126076	0.022328315
g5396.t1	PREDICTED: uncharacterized protein C11orf96 homolog	ENSDARP00000129219.1	3.771506614	0
g5411.t1	brambleberry-like	ENSDARP00000126692.1	4.27835911	0.02608713
g5417.t1	ras-related Rab-19-like	ENSDARP00000131109.1	-8.034622415	0.023927033
g5437.t1	nidogen and EGF-like domain-containing 1	ENSDARP00000065254.4	7.78139217	0.01074806
g5471.t1	cysteine-rich motor neuron 1 -like isoform X2	ENSDARP00000050533.4	6.38901627	0.004704129
g5473.t1	alpha cardiac	ENSDARP00000130083.1	-72.82319402	0
g5480.t1	thrombospondin-1-like	ENSDARP00000120705.2	2.445063411	0.033342398
g5617.t1	WNT1-inducible-signaling pathway 1-like	ENSDARP00000118864.1	-6.87599088	2.01734E-06
g5619.t1	PHD finger 20 1 isoform X1	ENSDARP00000018938.6	-8.929320295	0.002909506
g5629.t1	S100 calcium binding V2-like	ENSDARP00000094117.4	5.800554245	0
g5643.t1	coxsackievirus and adenovirus receptor homolog	ENSDARP00000134709.1	-2.140688181	8.66171E-06
g5655.t1	annexin A3-like	ENSDARP00000064973.2	2.109055917	4.47245E-12
g5665.t1	type I cytoskeletal 18-like	ENSDARP00000064976.5	-7.133151407	0
g5674.t1	butyrophilin 2 isoform X2	ENSDARP00000131129.1	-4.081078052	0.045319444
g5684.t1	transforming growth factor beta-1-induced transcript 1 isoform X1	ENSDARP00000104565.2	2.039900528	0.022373951
g5688.t1	sarcoplasmic endoplasmic reticulum calcium ATPase 1 isoform X1	ENSDARP00000096674.2	-56.4129763	0
g5723.t1	tumor necrosis factor receptor superfamily member 11B-like	ENSDARP00000036002.6	-3.495582479	0.011066208
g5763.t1	rho-related GTP-binding	ENSDARP00000018377.4	2.491490052	1.22507E-09
g5792.t1	interleukin-1 beta	ENSDARP00000140751.1	2.836002207	2.64675E-06
g5820.t1	olfactomedin-4-like	ENSDARP00000140836.2	-5.796406171	0.022669629
g5821.t1	olfactomedin-4-like	ENSDARP00000136182.1	-14.60840439	0.000429934
g5853.t1	solute carrier organic anion transporter family member 5A1	ENSDARP00000126538.1	11.60256418	6.10411E-05
g5865.t1	cysteine dioxygenase type 1	ENSDARP00000141126.1	2.625788884	0.00017857
g5905.t1	large neutral amino acids transporter small subunit 3	ENSDARP00000096925.4	-2.533365639	0.020862747
g5912.t1	carabin isoform X3	ENSDARP00000054436.3	-2.822658895	0.03289057
g595.t1	C-C motif chemokine 28 precursor	ENSDARP00000104032.2	-19.51265444	3.13241E-05
g5992.t1	bcl-2 13	ENSDARP00000084589.3	-3.677655158	0.001087698
g6008.t1	purine nucleoside phosphorylase-like	ENSDARP00000060000.3	-181.3529059	0.004111652
g6011.t1	fibulin-2-like isoform X2	ENSDARP00000100955.3	3.302567607	0.046299476
g6035.t1	uveal autoantigen with coiled-coil domains and ankyrin repeats-like	ENSDARP00000118060.1	-5.972827774	0.036714906

g6044.t1	myosin regulatory light chain cardiac muscle isoform-like	ENSDARP00000134838.1	-1.7977E+308	4.19434E-09
g6241.t1	kelch 41a	ENSDARP00000090560.4	-1.7977E+308	1.29637E-06
g6253.t1	desmin-like isoform X2	ENSDARP00000075991.4	-258.3991959	0
g6257.t1	tubulin alpha chain	ENSDARP00000093490.4	3.763105607	0
g6260.t1	fibronectin-like	ENSDARP00000112231.2	2.464131368	8.79395E-06
g6290.t1	carbohydrate sulfotransferase 15	ENSDARP00000129538.1	-3.780998783	0.000534975
g6306.t1	plasmalemma vesicle-associated -like	ENSDARP00000075680.5	2.329591901	0.004166879
g6374.t1	activin receptor type-1-like	ENSDARP00000123002.1	2.497256883	1.45867E-07
g6429.t1	FAM167A-like	ENSDARP00000132065.1	4.727410024	0
g6478.t1	endothelin-1 receptor-like	ENSDARP00000124118.1	4.364289789	0.000642631
g6575.t1	alpha-2,8-sialyltransferase 8F-like isoform X1	ENSDARP00000104838.2	-18.61080906	1.27276E-05
g6682.t1	organic solute transporter subunit alpha-like	ENSDARP00000066630.4	2.821013139	6.35528E-07
g6686.t1	APC membrane recruitment 2	ENSDARP00000102405.1	4.114492562	0.002705499
g6692.t1	HERV-H LTR-associating 2	ENSDARP00000109982.2	-2.958115141	0.007663098
g6701.t1	gap junction alpha-1	ENSDARP00000061260.3	2.966536418	1.33182E-05
g6702.t1	gap junction -like	ENSDARP00000061246.4	2.340361575	0.011327128
g6736.t1	annexin A6-like isoform X2	ENSDARP00000124518.2	-2.042141783	0.03386901
g6787.t1	nesprin-2-like	ENSDARP00000119338.1	-20.72422448	0.007655583
g6802.t1	homeobox SIX4-like	ENSDARP00000048109.3	-4.37222404	0.001174197
g6838.t1	hyaluronan and proteoglycan link 1-like	ENSDARP00000062630.5	12.92792149	0
g6849.t1	tubulin polymerization-promoting family member 3-like	ENSDARP00000117760.1	-9.239815777	2.61942E-08
g6859.t1	epithelial splicing regulatory 2 isoform X1	ENSDARP00000015568.9	-2.136734732	0.016390476
g6868.t1	myosin light chain skeletal muscle isoform	ENSDARP00000004932.5	-55.85975584	0
g6921.t1	barttin isoform X1	ENSDARP00000126598.1	3.676588454	0.002018699
g7002.t1	sperm acrosome membrane-associated 4-like	ENSDARP00000138050.1	5.077819842	0
g7003.t1	sperm acrosome membrane-associated 4-like	ENSDARP00000138050.1	-8.489949305	0
g7031.t1	BMP-binding endothelial regulator	ENSDARP00000134708.1	5.102915593	0.02707706
g7073.t1	transient receptor potential cation channel subfamily M member 4-like	ENSDARP00000086924.5	-2.584928699	0.006146676
g7074.t1	carbonic anhydrase 4-like	ENSDARP00000138709.1	-2.76493038	7.20991E-07
g7077.t1	serine protease 33-like	ENSDARP00000059297.3	-2.71238797	1.05174E-07
g7082.t1	fructose-bisphosphate aldolase A	ENSDARP00000042199.3	-25.62608753	0
g7091.t1	collagen alpha-1(I) chain-like	ENSDARP00000003038.8	-2.128458116	0
g7100.t1	serine threonine- kinase WNK4	ENSDARP00000129294.1	5.700674631	0.00032925
g7114.t1	jagunal homolog 1	ENSDARP00000129340.2	9.91307063	0
g7115.t1	target of Nesh-SH3-like isoform X5	ENSDARP00000083979.4	2.382743514	1.09415E-08
g7122.t1	catechol O-methyltransferase-like	ENSDARP00000038939.5	-7.116379853	0.03013822
g7128.t1	solute carrier family 43 member 3	ENSDARP00000141278.1	-3.38362819	0.023183745
g7259.t1	tumor necrosis factor receptor superfamily member 10B-like	ENSDARP00000107373.2	2.653899066	0.045342568
g7288.t1	prostate-associated microsemino	ENSDARP00000117235.1	587.9831034	0
g7438.t1	death-associated kinase 2-like	ENSDARP00000044425.7	4.384869308	0.000388931
g7494.t1	A disintegrin and metallo ase with thrombospondin motifs 15-like	ENSDARP00000120003.1	4.143290252	0.004179176
g7504.t1	perforin-1-like	ENSDARP00000133079.1	-19.36917904	1.4356E-12
g7506.t1	atonal homolog 1-like	ENSDARP00000105581.2	-1.7977E+308	0.01949323

g7517.t1	group 10 secretory phospholipase A2-like	ENSDARP00000132585.1	-4.227983947	2.57073E-08
g7548.t1	glutathione peroxidase 1-like	ENSDARP00000011398.6	2.499030991	0
g7605.t1	peroxisomal trans-2-enoyl- reductase	ENSDARP00000097753.3	-18.38692437	0
g7622.t1	poly(U)-specific endoribonuclease-C-like	ENSDARP00000099395.2	-2.321457828	0.001295326
g7646.t1	complement component C7-like	ENSDARP00000114284.1	4.501352372	0.015484746
g768.t1	transmembrane 238-like	ENSDARP00000127408.1	-5.48793406	1.14697E-08
g7686.t1	periostin isoform X2	ENSDARP00000141296.1	-2.140080998	0.004185429
g7729.t1	claudin-3-like	ENSDARP00000091985.3	-3.683060824	3.55182E-09
g7731.t1	bicaudal D-related 2-like	ENSDARP00000112331.2	-2.734257309	0.003420483
g7754.t1	desmoplakin-like isoform X1	ENSDARP00000103914.3	-3.465117468	0.003603872
g7770.t1	erythroferrone-like	ENSDARP00000072327.3	4.179376256	6.46949E-05
g7796.t1	apoptosis-inducing factor 3-like	ENSDARP00000112087.1	4.040207092	0.008628422
g7809.t1	kinase C and casein kinase substrate in neurons 3 isoform X1	ENSDARP00000138125.2	-3.745179462	4.23869E-05
g7867.t1	succinyl- ligase [ADP GDP-forming] subunit mitochondrial	ENSDARP00000069096.4	-2.123616965	6.79402E-06
g7925.t1	homeobox DLX-1	ENSDARP00000011033.5	2.069662819	0.013728264
g7926.t1	homeobox DLX-2	ENSDARP00000103004.2	2.992684858	2.03481E-10
g7964.t1	ferric-chelate reductase 1	ENSDARP00000121549.1	-18.42224139	1.66597E-09
g8043.t1	G- coupled receptor 1-like	ENSDARP00000139581.1	5.463198465	7.67329E-09
g8080.t1	pleckstrin homology domain-containing family G member 4B isoform X2	ENSDARP00000106098.2	2.940781652	0.027294242
g8121.t1	type I cytoskeletal 19-like	ENSDARP00000121430.1	-1.7977E+308	0
g8168.t1	sialic acid-binding Ig-like lectin 14	ENSDARP00000098106.3	4.673889977	3.96462E-13
g8171.t1	heat shock cognate 71 kDa -like	ENSDARP00000039324.7	-2.672547692	0.032291365
g8237.t1	sodium-coupled neutral amino acid transporter 3-like	ENSDARP00000004787.6	2.123422914	3.2464E-11
g8249.t1	matrix metallo ase-9-like	ENSDARP00000062844.4	3.478717044	3.55489E-13
g8274.t1	cationic amino acid transporter 2 isoform X1	ENSDARP00000053920.4	4.656078921	0.024439406
g8329.t1	glucose-6-phosphate isomerase	ENSDARP00000133574.1	-88.81127276	6.53621E-10
g8338.t1	growth arrest-specific 1-like	ENSDARP00000088466.1	3.739911963	0.044053176
g8342.t1	leptin	ENSDARP00000133784.1	5.428429908	9.90064E-11
g8345.t1	proline-rich transmembrane 4-like	ENSDARP00000108434.1	19.74786847	0.029536623
g8369.t1	heat shock beta-11-like	ENSDARP00000026207.4	-127.4061554	0.000380896
g8372.t1	ictacalcin-like	ENSDARP00000120686.1	-325.115257	0
g8403.t1	regulator of G- signaling 5-like	ENSDARP00000030228.5	-6.828456338	5.32021E-11
g8414.t1	myoglobin	ENSDARP00000039806.3	-34.96973756	0
g8450.t1	potassium voltage-gated channel subfamily KQT member 5-like	ENSDARP00000118905.1	-28.81590144	5.89124E-06
g8452.t1	cyclic GMP-AMP synthase	ENSDARP00000036480.6	-2.223868704	0.002234558
g8537.t1	L-lactate dehydrogenase A chain	ENSDARP00000140963.1	-2.424179064	7.00802E-05
g8622.t1	major facilitator superfamily domain-containing 4A-like	ENSDARP00000032062.8	-5.243373719	0.038274744
g8635.t1	homeobox DLX-5	ENSDARP00000100492.2	2.244952341	3.13885E-06
g8679.t1	hyaluronan synthase 2	ENSDARP00000053753.5	15.08714485	5.68842E-07
g8716.t1	myomesin-1 isoform X1	ENSDARP00000138459.1	-9.543532242	0.026624332
g879.t1	inositol hexakisphosphate kinase 2-like	ENSDARP00000024046.5	-2.313085641	0.00285021
g880.t1	inositol hexakisphosphate kinase 2-like	ENSDARP00000024046.5	-2.320158344	0.002237452
g8816.t1	carbonic anhydrase	ENSDARP00000022592.5	-8.265985887	0

g8902.t1	thioredoxin domain-containing 17	ENSDARP00000075336.3	-2.15158547	2.69053E-12
g8933.t1	enolase isoform X1	ENSDARP00000101018.2	-2.437715372	0.013333806
g8943.t1	estrogen-related receptor gamma-like isoform X2	ENSDARP00000106954.2	-46.86863075	9.92396E-05
g9006.t1	adenylate kinase isoenzyme 1	ENSDARP00000013777.6	-3.63823967	1.46985E-13
g9007.t1	transforming growth factor beta receptor type 3-like	ENSDARP00000129928.1	3.102228376	0.047008301
g901.t1	adenylosuccinate synthetase isozyme 1	ENSDARP00000141430.1	-16.29880547	0.001156835
g9045.t1	mitochondrial basic amino acids transporter isoform X2	ENSDARP00000131547.1	-2.981921721	0.011846437
g9137.t1	C8orf4 homolog	ENSDARP00000098558.2	2.249592567	6.82861E-05
g9176.t1	metallo ase inhibitor 3	ENSDARP00000139308.1	2.395881101	0.000684502
g9231.t1	intermediate filament ON3-like	ENSDARP00000016453.6	-2.765652418	0
g9259.t1	transforming growth factor-beta-induced ig-h3	ENSDARP00000096711.3	-2.643981178	0.010708751
g9260.t1	PCTP	ENSDARP00000096710.4	-8.676041851	1.82045E-05
g9282.t1	fibroblast growth factor-binding 2-like	ENSDARP00000071237.4	2.719001929	3.10155E-05
g9287.t1	ribonuclease ZC3H12A-like	ENSDARP00000136372.1	-2.950912488	0.005589917
g9290.t1	eva-1 homolog B-like	ENSDARP00000130607.1	2.108708323	6.97986E-06
g9311.t1	sodium potassium-transporting ATPase subunit beta-233-like	ENSDARP00000003281.6	-7.571629638	0
g9312.t1	dermatopontin	ENSDARP00000073086.2	2.002299357	0.00015251
g9324.t1	tetraspanin-1-like	ENSDARP00000122014.1	-11.86063309	6.82179E-10
g9333.t1	PREDICTED: uncharacterized protein LOC108279429 isoform X2		-45.22344616	0.000153766
g9349.t1	collagen triple helix repeat-containing 1-like	ENSDARP00000117902.2	4.367920829	3.52325E-05
g9400.t1	calsequestrin-2-like	ENSDARP00000020399.5	-27.08815557	0.036749172
g9408.t1	-lysine 6-oxidase-like	ENSDARP00000123413.1	12.07638543	0
g9438.t1	uroplakin-3a-like isoform X1	ENSDARP00000113175.1	-3.925486951	0.000608366
g9449.t1	high affinity immunoglobulin gamma Fc receptor I-like	ENSDARP00000088608.4	2.352791838	2.49842E-10
g9455.t1	myelin basic -like isoform X1	ENSDARP00000126314.1	-15.49534323	0.007213833
g9472.t1	claudin-8-like	ENSDARP00000070266.3	-29.20521481	9.36889E-05
g9473.t1	claudin-8-like	ENSDARP00000059483.5	-5.863102843	0.003943452
g9496.t1	desmoglein-2-like	ENSDARP00000103538.2	-5.401426833	1.25173E-06
g9499.t1	cadherin-2	ENSDARP00000003949.9	2.812781137	0.000334801
g9522.t1	phospholipid phosphatase 3 isoform X1	ENSDARP00000078333.4	2.026987095	0.000357785
g961.t1	type II cytoskeletal 8-like	ENSDARP00000016453.6	-25.64716818	0
g962.t1	type II cytoskeletal 8-like	ENSDARP00000075658.4	2.220388995	0
g9634.t1	sarcoplasmic endoplasmic reticulum calcium ATPase 2 isoform X2	ENSDARP00000140204.1	-261.5715964	0
g9658.t1	caspase-6-like	ENSDARP00000093958.3	-3.45992476	4.55474E-06
g9669.t1	T-box transcription factor TBX1 isoform X1	ENSDARP00000118808.1	2.212666664	0.00015121
g9679.t1	docking 4	ENSDARP00000102258.2	7.135784406	0.022445671
g9716.t1	olfactomedin 2A	ENSDARP00000078874.5	2.224055285	0.000566749
g9756.t1	carboxypeptidase E	ENSDARP00000072733.5	5.177354203	0
g9812.t1	troponin slow skeletal muscle-like	ENSDARP00000132136.1	-1.7977E+308	0.00027935
g9819.t1	adapter CIKS-like isoform X2	ENSDARP00000128747.1	-3.130372369	3.12939E-09
g994.t1	dentin sialophospho -like isoform X1	ENSDARP00000103899.2	-2.768468991	0.010234971
g9974.t1	arginase-1	ENSDARP00000097045.3	-30.37263858	0
g9996.t1	interferon lambda receptor 1-like	ENSDARP00000068753.2	-5.010252073	0.013212725

g9999.t1	alpha-crystallin B chain-like	ENSDARP00000062522.3	-73.45940493	0.006767895