bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Genome sequencing of the perciform fish *Larimichthys crocea* provides insights into stress adaptation

Jingqun Ao[1†], Yinnan Mu[1†], Li-Xin Xiang[2†], DingDing Fan[3†], MingJi Feng[3†], Shicui Zhang[4†], Qiong Shi[3], Lv-Yun Zhu[2], Ting Li[1], Yang Ding[1], Li Nie[2], Qiuhua Li[1], Wei-ren Dong[2], Liang Jiang[5], Bing Sun[4], XinHui Zhang[3], Mingyu Li[1], Hai-Qi Zhang[2], ShangBo Xie[3], YaBing Zhu[3], XuanTing Jiang[3], Xianhui Wang[1], Pengfei Mu[1], Wei Chen[1], Zhen Yue[3], Zhuo Wang[3], Jun Wang[3*], Jian-Zhong Shao[2*] and Xinhua Chen[1*]

[1] Key Laboratory of Marine Biogenetics and Resources, Third Institute of Oceanography, State Oceanic Administration; Fujian Collaborative Innovation Center for Exploitation and Utilization of Marine Biological Resources; Key Laboratory of Marine Genetic Resources of Fujian Province, Xiamen 361005, P.R. China

[2] College of Life Sciences, ZheJiang University, Hangzhou 310058, ZheJiang, P.R. China

[3] BGI-Tech, BGI-Shenzhen, Shenzhen 518083, Guangdong, P.R. China

[4] Ocean University of China, Qingdao 266100, Shandong, P.R. China

[5] College of Life Sciences, Shenzhen University, Shenzhen 518060, Guangdong, P.R. China

[†]These authors contributed equally to the work.

[*]To whom correspondence should be addressed: Xinhua Chen: Tel: +86-592-2195297, Fax: +86-592-2195297, Email: chenxinhua@tio.org.cn; Jun Wang: Tel: +86-755-25274247, Fax: +86-755-25274247, Email: wangj@genomics.org.cn; Jian-Zhong Shao: Tel: +86-571-88206582, Fax: +86-571-88206582, Email: shaojz@zju.edu.cn.

Keywords: Large yellow croaker, *Larimichthys crocea*; Genome sequencing; Hypoxia; Air exposure; Transcriptome; Proteome.

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Abstract

The large yellow croaker *Larimichthys crocea* (*L. crocea*) is one of the most economically important marine fish in China and East Asian countries. It also exhibits peculiar behavioral and physiological characteristics, especially sensitive to various environmental stresses, such as hypoxia and air exposure. These traits may render *L. crocea* a good model for investigating the response mechanisms to environmental stress. To understand the molecular and genetic mechanisms underlying the adaptation and response of *L. crocea* to environmental stress, we sequenced and assembled the genome of *L. crocea* using a bacterial artificial chromosome and whole-genome shotgun hierarchical strategy. The final genome assembly was 679 Mb, with a contig N50 of 63.11 kb and a scaffold N50 of 1.03 Mb, containing 25,401 protein-coding genes. Gene families underlying adaptive behaviours, such as vision-related crystallins, olfactory receptors, and auditory sense-related genes, were significantly expanded in the genome of *L. crocea* relative to those of other vertebrates. Transcriptome analyses of the hypoxia-exposed *L. crocea* brain revealed new aspects of neuro-endocrine-immune/metabolism regulatory networks that may help the fish to avoid cerebral inflammatory injury and maintain energy balance under hypoxia. Proteomics data demonstrate that skin mucus of the air-exposed *L. crocea* had a complex composition, with an unexpectedly high number of proteins (3,209), suggesting its multiple protective mechanisms involved in antioxidant functions, oxygen transport, immune defence, and osmotic and ionic regulation. Our results provide novel insights into the mechanisms of fish adaptation and response to hypoxia and air exposure.

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Introduction

Teleost fish, nearly half of all living vertebrates, display an amazing level of diversity in body forms, behaviors, physiologies, and environments that they occupy. Strategies for coping with diverse environmental stresses have evolved in different teleost species. Therefore, teleost fish are considered to be good models for investigating the adaptation and response to many natural and anthropogenic environmental stressors (Gracey et al. 2001; Cossins and Crawford 2005; van der Meer et al. 2005). Recent genome-sequencing projects in several fish have provided insights into the molecular and genetic mechanisms underlying their responses to some environmental stressors (Star et al. 2011; Schartl et al. 2013; Chen et al. 2014). However, to better clarify the conserved and differentiated features of the adaptive response to specific stresses and to trace the evolutionary process of environmental adaptation and response in teleost fish, insight from more teleost species with different evolutionary positions, such as Perciformes, is required. Perciformes are by far the largest and most diverse order of vertebrates, and thus offer a large number of models of adaptation and response to various environmental stresses.

The large yellow croaker, *Larimichthys crocea* (*L. crocea*), is a temperate-water migratory fish that belongs to the order Perciformes and the family Sciaenidae. It is mainly distributed in the southern Yellow Sea, the East China Sea, and the northern South China Sea. *L. crocea* is one of the most economically important marine fish in China and East Asian countries due to its rich nutrients and trace elements, especially selenium. In China, the annual yield from *L. crocea* aquaculture exceeds that of any other net-cage-farmed marine fish species (Liu et al. 2013; Liu et al. 2014). Recently, the basic studies on genetic improvement for growth and

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

disease resistance traits of *L. crocea* are increasingly performed for farming purpose (Ning et al. 2007; Mu et al. 2010; Liu et al. 2013; Ye et al. 2014). *L. crocea* also exhibits peculiar behavioral and physiological characteristics, such as loud sound production, high sensitivity to sound, and well-developed photosensitive and olfactory systems (Su 2004; Zhou et al. 2011). Most importantly, *L. crocea* is especially sensitive to various environmental stresses, such as hypoxia and air exposure. For example, the response of its brain to hypoxia is quick and robust, and a large amount of mucus is secreted from its skin when it is exposed to air (Su 2004; Gu and Xu 2011). These traits may render *L. crocea* a good model for investigating the response mechanisms to environmental stress. Several studies have reported transcriptomic and proteomic responses of *L. crocea* to pathogenic infections or immune stimuli (Mu et al. 2010; Yu et al. 2010; Mu et al. 2014). The effect of hypoxia on the blood physiology of *L. crocea* has been evaluated (Gu and Xu 2011). However, little is known about the molecular response mechanisms of *L. crocea* against environmental stress.

To understand the molecular and genetic mechanisms underlying the responses of *L. crocea* to environmental stress, we sequenced its whole genome. Furthermore, we sequenced the transcriptome of the hypoxia-exposed *L. crocea* brain and profiled the proteome of its skin mucus under exposure to air. Our results revealed the molecular and genetic basis of fish adaptation and response to hypoxia and air exposure.

Results

Genome features

We applied a bacterial artificial chromosome (BAC) and whole-genome shotgun (WGS)

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

hierarchical assembly strategy for the *L. crocea* genome to overcome the high levels of genome heterozygosity (**Table 1**; **Supplemental Fig. S1-S2**). The 42,528 BACs were sequenced by the HiSeq 2000 platform and each BAC was assembled by SOAPdenovo (Luo et al. 2012) (**Supplemental Table S1**). The total length of all combined BACs was 3,006 megabases (Mb), which corresponded to approximately 4.3-fold genome coverage (**Supplemental Tables S2-S3**). All BAC assemblies were then merged into super-contigs and oriented to super-scaffolds with large mate-paired libraries (2-40 kb). Gap filling was made with reads from short insert-sized libraries (170-500 bp) (**Supplemental Tables S3-S4**). In total, we sequenced 563-fold coverage bases of the estimated 691 Mb genome size. The final assembly was 679 Mb, with a contig N50 of 63.11 kb and a scaffold N50 of 1.03 Mb (**Table 1**). The 672 longest scaffolds (11.2% of all scaffolds) covered more than 90% of the assembly (**Supplemental Table S5**). To assess the completeness of the *L. crocea* assembly, 52-fold coverage paired-end high-quality reads were aligned against the assembly (**Supplemental Fig. S3**). More than 95.63% of the generated reads could be mapped to the assembly. Furthermore, the integrity of the assembly was validated by the successful mapping of 98.80% of the transcripts from the mixed-tissue transcriptomes (**Supplemental Table S6**). These results indicate that the genome assembly of *L. crocea* has high coverage and is of high quality (**Supplemental Table S7**).

The repetitive elements comprise 18.1% of the *L.crocea* genome (**Supplemental Table S8**), which is a relatively low percentage when compared with other fish species, such as *Danio rerio* (52.2%), *Gadus morhua* (25.4%), and *Gasterosteus aculeatus* (25.2%). This suggests that *L. crocea* may have a more compact genome (**Supplemental Tables S9-S10**).

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

We identified 25,401 protein-coding genes based on *ab initio* gene prediction and evidence-based searches from the reference proteomes of six other teleost fish and humans **(Supplemental Fig. S4; Table S11)**, in which 24,941 genes (98.20% of the whole gene set) were supported by homology or RNAseq evidence (**Supplemental Fig. S5**). Over 97.35% of the inferred proteins matched entries in the InterPro, SWISS-PROT, KEGG or TrEMBL database (**Supplemental Table S12**).

Phylogenetic relationships and genomic comparison

L. crocea is the first species of *Sciaenidae* of the order *Perciformes* with a complete genome available, therefore we estimated its phylogenetic relationships to seven other sequenced teleost species based on 2,257 one-to-one high-quality orthologues, using the maximum likelihood method. According to the phylogeny and the fossil record of teleosts, we dated the divergence of *L. crocea* from the other teleost species to approximately 64.7 million years ago (**Fig. 1A**). We also detected 19,283 orthologous gene families (**Supplemental Table S3**), of which 14,698 families were found in *L. crocea*. The gene components of *L. crocea* were similar to those of *D. rerio* (**Fig. 1B**). The gene contents in four representative teleost species and *L. crocea* genomes were also analysed, and 11,205 (76.23%) gene families were found to be shared by five teleosts (**Fig. 1C**). We confirmed that the one-to-one orthologous genes of *G. aculeatus* and *L. crocea* have higher sequence identities from the distribution of the percent identity of proteins (**Fig. 1D**), which indicates that Sciaenidae has a closer affinity to Gasterosteiformes and coincides with our genome-level phylogeny position.

Furthermore, 121 significantly expanded and 27 contracted gene families ($P < 0.01$) were identified by comparing the family size of *L. crocea* with that of the other vertebrates used in

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

the phylogenetic analysis (**Supplemental Tables S14-S15**). Based on the ratios of the number of nonsynonymous substitutions per nonsynonymous site (K_a) to the number of synonymous substitutions per synonymous site (K_s; K_a/K_s ratios) in a branch-site model of PAML (Yang 1997), 92 genes were found to be positively selected in *L. crocea* compared with their orthologues in the other six teleost species ($P < 0.001$, **Supplemental Table S16**).

Unique genetic features of the *L. crocea*.

L. crocea is a migratory fish with good photosensitivity, olfactory detection, and sound perception, and it contains high levels of selenium (Su 2004). Our genomic analyses provide genetic basis for these behavioral and physiological characteristics. Several crystallin genes (*crygm2b*, *cryba1*, and *crybb3*), which encode proteins that maintain the transparency and refractive index of the lens (Chen et al. 2014), were significantly expanded in the genome of *L. crocea* relative to those of other sequenced teleosts (**Supplemental Table S17**). Phylogenetic analysis showed that the crystallin genes from *L. crocea* cluster together, indicating that these genes were specifically duplicated in *L. crocea* lineage (**Supplemental Fig. S6**). The specific expansion of these crystallin genes may be helpful for improving photosensitivity by increasing lens transparency, thereby enabling the fish to easily find food and avoid predation underwater.

We also identified 112 olfactory receptor (OR)-like genes from the *L. crocea* genome (**Supplemental Table S18**; **Fig. S7**), and almost all of them (111) have been reported to be expressed in the olfactory epithelial tissues of *L. crocea* (Zhou et al. 2011). The majority of these genes (66) were classified into the "delta" group, which is important for the perception of water-borne odorants (Niimura 2009). *L. crocea* also possessed the highest number of

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

genes that were classified into the "eta" group (30, $P < 0.001$), and these genes may contribute to the olfactory detection abilities, which could be useful for feeding and migration (Li et al. 1995).

L. crocea is named for its ability to generate strong repetitive drumming sounds, especially during reproduction (Su 2004). For good communication, fish have developed high sensitivities to environmental sound. Three important auditory genes, otoferlin (*OTOF*), *claudinj*, and otolin 1 (*OTOL1*), were significantly expanded in the *L. crocea* genome ($P < 0.01$, **Supplemental Table S19**). These expansions may contribute to the detection of sound signaling during communication, and thus to reproduction and survival (Eisen and Ryugo 2007).

Selenium is highly enriched in *L. crocea* (Su 2004), and it is mainly present as selenoproteins. We used the SelGenAmic-based selenoprotein prediction method (Jiang et al. 2010) to analyse the *L. crocea* genome and identified 40 selenoprotein genes, which is the highest number among all sequenced vertebrates (**Supplemental Table S20**). Interestingly, five copies of *MsrB1*, which encodes methionine sulfoxide reductase, were found in *L. crocea* (*MsrB1a, MsrB1b, MsrB1c*, *MsrB1d*, and *MsrB1e*), whereas only two copies (*MsrB1a* and *MsrB1b*) were found in other fish, thus suggesting its broader specificity to reduce all possible substrates (Vandermarliere et al. 2014).

Characterization of the *L. crocea* immune system

Approximately 2,524 immune-relevant genes were annotated in the *L. crocea* genome, including 819 innate immune-relevant genes and 1,705 adaptive immune-relevant genes (**Supplemental Table S21**). *L. crocea* has a relatively complete innate immune system,

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

whereas its adaptive immune system may possess unique characteristics. The $CD8^+$ T and $CD4^+$ T-helper type 1 (Th1) -type immune systems are well conserved in *L. crocea*, and almost all $CD8^+$ T and $CD4^+$ Th1 cell-related genes were found (**Fig. 2A**). Moreover, the genes related to Th17 cell- and γδ-T cell-mediated mucosal immune responses were conserved in *L. crocea*. These observations suggest that *L. crocea* may exhibit powerful cellular and mucosal immunity. However, the $CD4^+$ Th2-type immunity seemed to be weak in *L. crocea*, as suggested by the absence of many $CD4^+$ Th2-related genes and humoral immune effectors (**Fig. 2A**). We detected gene expansions in several of these immune-relevant genes, including those encoding lectin receptors (*CLEC17A*), a classical complement component (*C1q*), apoptosis regulator (*BAX*), and immunoglobulins (*IgHV*) ($P < 0.01$, **Supplemental Table S22**). Expansions were also observed in the genes encoding four key proteins for mammalian antiviral immunity: tripartite motif containing 25 (TRIM25), cyclic GMP-AMP synthase (cGAS), DDX41, and NOD-like receptor family CARD domain containing 3 (NLRC3) (**Fig. 2B**). However, retinoic acid-inducible gene-1 (*RIG-I*), which initiates antiviral signaling pathway in mammals, was not found in the *L. crocea* genome and transcriptome (Mu et al. 2010; Mu et al. 2014). The teleost *RIG-I* has been identified only in limited fish species, such as cyprinids and salmonids, and its absence suggests that it may have been lost from particular fish genomes (Hansen et al. 2011). Furthermore, laboratory of genetics and physiology 2 (LGP2) can serve as a suppressor to block RIG-I- and melanoma differentiation-associated protein 5 (MDA5)-elicited signaling in mammals, but LGP2 in fish is able to bind to poly(I:C) to trigger interferon production (Chang et al. 2011), thereby acting as a substitute for RIG-I (**Fig. 2B**). The expanded *TRIM25* (54 copies, **Supplemental Fig. S8**)

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

may trigger the ubiquitination of interferon-β promoter stimulator-1 (IPS-1), thus allowing for interferon regulatory factor 3 (IRF3) phosphorylation and antiviral signaling initiation (Castanier et al. 2012). *DDX41* and *cGAS* encode intracellular DNA sensors, which can activate stimulator of interferon genes (STING) and TANK-binding kinase 1 (TBK1) to induce type I interferons (Zhang et al. 2011; Gao et al. 2013). *L. crocea* contained 76 copies of *NLRC3* (**Supplemental Fig. S9**), which encodes regulators that prevent type I interferon overproduction (Zhang et al. 2014). The expansions of these virus-response genes suggest their enhanced roles in innate antiviral immunity, which may explain why *L. crocea* is less susceptible to viral infection.

Stress response under hypoxia

The brain allows rapid and coordinated responses to the environmental stress by driving the secretion of hormones. Therefore, we studied the response of the *L. crocea* brain to hypoxia. We sequenced seven transcriptomes of the brains at different times of hypoxia exposure and found that 8,402 genes were differentially expressed at one or more time points (false discovery rate [FDR] ≤ 0.001, fold change > 2; **Supplemental Fig. S10**). Hypoxia stress induced a response with the largest number of genes (4,535 genes) at 6 h (**Supplemental Fig. S11**), indicating that genes with regulated expression at 6 h may be critical for the response. Hypoxia stress can induce the response of the central neuroimmune system, in which brain neuropeptides, endocrine hormones, and inflammatory cytokines closely participate (Herman and Cullinan 1997; Yang et al. 2012; Lemos Vde et al. 2013). However, the precise regulatory networks among these factors have not yet been fully delineated. Our transcriptome analysis of *L. crocea* brains under hypoxia stress may outline a novel

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

hypothalamic-pituitary-adrenal (HPA) axis-endothelin-1 (ET-1)/adrenomedullin (ADM)-interleukin (IL)-6/tumor necrosis factor (TNF)-α feedback regulatory loop that is involved in the neuro-endocrine-immune network during hypoxia responses (**Fig. 3; Supplemental Table S23; Fig. S12).**

Results from transcriptome analyses show that the key HPA axis-relevant genes (corticotropin-releasing factor [*CRF*], CRF receptor 1 [*CRFR1*], pro-opiomelanocortin [*POMC*], and CRF-binding protein [*CRFBP*]) in the *L. crocea* brain displayed a down-up-down-up (W-type) dynamic expression pattern under hypoxia stress (**Supplemental Fig. S12)**. The HPA axis can strictly control the production of glucocorticoids (Nadeau and Rivest 2003; Sorrells and Sapolsky 2007), and glucocorticoids are suppressors of ET-1 and ADM, which are both involved in cerebral inflammation in mammals (Takahashi et al. 2003; Hayashi et al. 2004). Meanwhile, the dynamic expression levels of *ET-1* and *ADM* clearly showed a typical M-type pattern (up-down-up-down), and the time of inflexion point corresponded with that of *CRF*, *CRFR1*, *POMC*, and *CRFBP*. These observations suggest the existence of a feedback regulatory pathway between the HPA axis and ET-1/ADM under hypoxia stimulation. Notably, the expression of *IL-6/TNF-α* showed the M-type pattern and was consistent with that of *ET-1/ADM* (**Supplemental Fig. S12)**. These coordinated and fluctuating expression patterns indicate that hypoxia may induce the expression of *ET-1/ADM* and *IL-6/TNF-α* and trigger a positive feedback loop between them (**Fig. 3**). Furthermore, ET-1/ADM-IL-6/TNF-α may activate the HPA axis, and the latter subsequently induces glucocorticoids and generates a negative feedback to inhibit *ET-1/ADM* and *IL-6/TNF-α* expression to reduce inflammatory response in brain. This suggestion could be supported by

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

previous reports in mammals (Mastorakos et al. 1993; Kitamuro et al. 2000; Earley et al. 2002; Takahashi et al. 2003).

L. crocea also exhibits other protective mechanisms, such as the suppressors of cytokine signaling (SOCS)-dependant regulatory mechanism, to avoid inflammation-induced cerebral injury. Both *SOCS-1* and *SOCS-3* in the *L. crocea* brain display opposite expression patterns against *IL-6* and *TNF-α* (**Supplemental Fig. S12)**. Thus, SOCS-1 and SOCS-3 may have complementary roles in down-regulating *IL-6* and *TNF-α*, and both IL-6 and TNF-α have reciprocal functions to induce *SOCS-1* and *SOCS-3* expression (**Fig. 3**). These results suggest that a SOCS-1/3-dependent feedback regulation may exist in the process against hypoxia-induced cerebral inflammation in *L. crocea*.

Hypoxia can influence the hypothalamic-pituitary-thyroid (HPT) axis (Hou and Du 2005). HPT axis was found to regulate protein synthesis and glucose metabolism by production of thyroid hormones (Yen 2001). Here, the major HPT axis-related genes (thyrotropin-releasing hormone [*TRH*], TRH receptor [*TRHR*], thyroid-stimulating hormone [*TSH*], and TSH receptor [*TSHR*]) were significantly down-regulated in the *L. crocea* brain at 1 h to 6 h under hypoxia (**Supplemental Table S24**), thus indicating that the HPT axis may be inhibited during the early period of hypoxia. Inhibition of the HPT axis leads to a decrease in the production of thyroid hormones. Furthermore, thyroid hormones can regulate ribosomal biogenesis and protein translation by the PI3K-Akt-mTOR-S6K signaling pathway (Kenessey and Ojamaa 2006). In this study, the mRNA levels of *PI3K*, *S6K*, and most of the components of the protein translation machinery, including the ribosomal proteins and eukaryotic translation initiation factors (*eIF-1*, *-2*, *-3*, *-5* and *-6*), were all down-regulated under hypoxia

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(**Supplemental Table S25**). This suggests that the HPT axis may inhibit protein synthesis under hypoxia by decreasing the production of thyroid hormones (**Fig. 3**), which is beneficial for saving energy during hypoxia stress. Thyroid hormones can also accelerate the oxidative metabolism of glucose and inhibit the glycolytic anaerobic pathway (Sabell et al. 1985). Our transcriptome analyses show that genes involved in the tricarboxylic acid (TCA) cycle (pyruvate dehydrogenase complex [*PDC-E1*], succinyl-CoA synthetase [*SCS*], and fumarate hydratase [*FH*]) were down-regulated 12 h later under hypoxia, whereas glycolysis-related genes, such as pyruvate kinase (*PKM*), glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*), *GPI*, and aldolase A (*ALDOA*), were greatly increased at 1 h (280-, 130-, 73- and 12-fold, respectively) (**Supplemental Table S24**). The down-regulation of HPT axis-thyroid hormones may inhibit the TCA cycle and accelerate the anaerobic glycolytic pathway in the brain during hypoxia exposure (**Fig. 3**). The repression of the TCA cycle and the strong induction of the anaerobic glycolytic pathway resulted in a physiological shift from aerobic to anaerobic metabolism, where fish utilise O_2-independent mechanisms to produce adenosine triphosphate (ATP). However, the mRNA levels of hypoxia-inducible factor (HIF)-1α, which are significantly up-regulated under hypoxia in mammals (Dayan et al. 2006; Benita et al. 2009), were not significantly changed in the *L. crocea* brain (**Supplemental Table S24**). It is possible that the HIF-1α-mediated mechanism may not be essential for the hypoxia response in the *L. crocea* brain during the early period of hypoxia. These results suggest that the HPT axis-mediated effects may play major roles in response to hypoxia by reorganizing energy consumption and energy generation.

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Mucus components and function

The skin mucus is considered as the first defensive barrier between fish and its aquatic environment, and it plays a role in a number of functions, including locomotion, antioxidant responses, respiration, disease resistance, communication, ionic and osmotic regulation (Shephard 1994). However, the exact mechanisms underlying these functions remain unknown. Mucus is composed mainly of the gel-forming macromolecule mucin and water (Subramanian et al. 2008). We identified 159 genes that are implicated in mucin biosynthesis and mucus production in the *L. crocea* genome (**Supplemental Table S26**), based on previous studies in mammals (Pluta et al. 2012). This indicates that the mucin synthetic pathway is conserved between fish and mammals. Among these gene families, GALNT, which encodes N-acetylgalactosaminyl transferases (Guzman-Aranguez et al. 2009), was significantly expanded in *L. crocea* (27 copies versus 15–20 copies in other fish) (**Supplemental Fig. S13**). Syntaxin-11 was also expanded. Additionally, genes encoding syntaxin-binding protein 1 and syntaxin-binding protein 5, which are related to mucus secretion, were positively selected in the *L. crocea* genome (**Supplemental Table S16**). The expansion and positive selection of these genes may explain why the *L. crocea* secretes more mucus than other fish under stress.

We identified 22,054 peptides belonging to 3,209 genes in the *L. crocea* skin mucus proteome, and this accounted for more than 12% of the protein-coding genes in the genome (**Supplemental Table S27**). The complexity of the *L. crocea* mucus presumably relates to the multitude of its biological functions that allow the fish to survive and adapt to environmental changes. The over-represented functional categories were oxidoreductase activity

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(GO:0016491, $P=1.58\times10^{-35}$, 223 proteins), peroxidase activity (GO:0004601, $P=0.0075$, nine proteins), oxygen binding (GO:0019825, $P=0.0011$, eight proteins), and ion binding (GO:0043167, $P=2.21\times10^{-6}$, 347 proteins) (**Fig. 4A; Supplemental Fig. S14**). Two hundred and thirty-two antioxidant proteins that were related to oxidoreductase activity and peroxidase activity were highly enriched in the *L. crocea* mucus, and they included peroxiredoxins, glutathione peroxidase, and thioredoxin (**Supplemental Table S28**). These proteins intercept and degrade environmental peroxyl and hydroxyl radicals from aqueous environments (Cross et al. 1984). Therefore, the presence of high-abundance antioxidant proteins in the skin mucus may have the potential to protect fish against air exposure-induced oxidative damage (**Fig. 4B**). Eight proteins related to oxygen transport, including hemoglobin subunits α1, αA, αD, β, and β1, and cytoglobin-1, were identified in the *L. crocea* skin mucus (**Supplemental Table S29**). The abundant expression of hemoglobin may contribute to the binding and holding of oxygen for respiration. Various immune molecules that provide immediate protection to fish from potential pathogens, such as lectins, lysozymes, C-reactive proteins, complement components, immunoglobulins, and chemokines, were also found in the *L. crocea* skin mucus (**Supplemental Table S30**). To date, the mechanisms of osmotic and ionic regulation of the skin mucus have not been confirmed (Shephard 1994). In this study, a large number of ion-binding proteins were identified in the *L. crocea* mucus (**Supplemental Table S31**). These proteins and the layer of mucus may have a role in limiting the diffusion of ions on the surface of the fish (**Fig. 4B**). However, a substantial proportion of the proteins, which are highly present in the skin mucus of fish under air exposure, play an unknown role in the mucus response.

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Discussion

We sequenced and assembled the genome of the large yellow croakerr (*L. crocea*) using BACs and the WGS hierarchical assembly strategy. This methodology is effective for high-polymorphism genomes and produces a high quality genome assembly, with the 63.11 kb contig N50 and 1.03 Mb scaffold N50 (**Table 1**). Support from the 563-fold coverage of genome yields high single-base resolution and 98.80% completeness of the coding region (**Supplemental Table S6**). Further genomic analyses showed the significant expansion of several gene families, such as vision-related crystallins, olfactory receptors, and auditory sense-related genes, and provided a genetic basis for the peculiar behavioral and physiological characteristics of *L. crocea*.

During the early stages of hypoxia stress, the induction of ET-1/ADM and IL-6/TNF-α generates the primary protective effect to increase blood pressure, enhance vascular permeability and trigger inflammatory response (Bona et al. 1999; Taylor et al. 2005). These mechanisms maintain the brain oxygen supply and resist pathogen infection when the blood brain barrier is disrupted by hypoxia (Kaur and Ling 2008). As the stress response progresses, several natural brakes, including HPA axis-Glucocorticoids and SOCS family members, exhibit secondary protection effects to avoid excessive inflammatory responses in the brain. Our transcriptome results show that a novel HPA axis-ET-1/ADM-IL-6/TNF-α feedback regulatory loop in neuro-endocrine-immune networks contributed to the protective effect and regulated moderate inflammation under hypoxia stress (**Fig. 3**). On the other hand, the hypoxia-induced down-regulation of the HPT axis may lead to the inhibition of protein synthesis and the activation of anaerobic metabolism (**Fig. 3; Supplemental Tables S24-S25**).

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Inhibition of protein synthesis principally contributes to the reduction in cellular energy consumption during hypoxia (Gracey et al. 2001; Richards 2011). Activation of anaerobic metabolism facilitates O_2-independent ATP production under hypoxia, albeit with low ATP yield (Richards 2011). Therefore the reduction in ATP consumption through the HPT axis-mediated inhibition of protein synthesis matched the lower ATP yield by the HPT axis-activated anaerobic metabolism, which may aid to maintain cellular energy balance under hypoxia, thus extending fish survival. Hence, our results reveal new aspects of neuro-endocrine-immune/metabolism regulatory networks that may help the fish to avoid cerebral inflammatory injury and maintain energy balance under hypoxia stress. These discoveries will help to improve current understanding of neuro-endocrine-immune/metabolism regulatory networks and protective mechanisms against hypoxia-induced cerebral injury in vertebrates, providing clues for research on the pathogenesis and treatment of hypoxia-induced cerebral diseases.

Amazingly, 3,209 different proteins were identified in the *L. crocea* skin mucus under air exposure. Of these, oxidoreductase activity-, oxygen binding-, immunity-, and ion binding-related proteins were enriched (**Fig. 4A; Supplemental Fig. S14**). The increase in secretion of the skin mucus of *L. crocea* under air exposure may reflect a physiological adjustment of the fish to cope with environmental changes, and the complex components suggest that the skin mucus exerts multiple protective mechanisms, which are involved in antioxidant functions, oxygen transport, immune defence, and osmotic and ionic regulation (**Fig. 4B**). These results expand our knowledge of skin mucus secretion and function in fish, highlighting its importance in response to stress. In addition, the mucus proteome shares

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

many proteins with the mucus from humans and other animals (Lee et al. 2011; Rodriguez-Pineiro et al. 2013). These characteristics thus make *L. crocea* a pertinent model for studying mucus biology.

In summary, our sequencing of the genome of the large yellow croaker provided the genetic basis for its peculiar behavioral and physiological characteristics. Results from transcriptome analyses revealed new aspects of neuro-endocrine-immune/metabolism regulatory networks that may help the fish to avoid cerebral inflammatory injury and maintain energy balance under hypoxia stress. Proteomic profiling suggested that the skin mucus of the fish exhibits multiple protective mechanisms in response to air-exposure stress. Overall, our results revealed the molecular and genetic basis of fish adaptation and response to hypoxia and air exposure. In addition, the data generated by this study will facilitate the genetic dissection of aquaculture traits in this species and provide valuable resources for the genetic improvement of the meat quality and production of *L. crocea*.

Materials and Methods

Genome assembly annotation

The wild *L. crocea* individuals were collected from the Sanduao sea area in Ningde, Fujian, China. Genomic DNA was isolated from the blood of a female fish by using standard molecular biology techniques for BAC library construction and sequencing by the HiSeq 2000 Sequencing System in BGI (Beijing Genomics Institute, Shenzhen, China). Subsequently, low-quality and duplicated reads were filtered out, and sequencing errors were removed. The BACs of *L. crocea* were assembled by using SOAPdenovo2 (Li et al. 2009)

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

(http://soap.genomics.org.cn) with k-mers that ranged from 25 to 63 in size. Then, we selected the assembly with the longest scaffold N50 for gap filling. The BACs were merged together based on the overlap found by BLAT, using the custom script: Rabbit (ftp://ftp.genomics.org.cn/pub/Plutellaxylostella/Rabbit_linux-2.6.18-194.blc.tar.gz). The redundant sequences that were produced by high polymorphisms were removed by sequence depth and shared k-mer percentage. Assembly was performed by scaffolding with mate-paired libraries (2–40 kb) using SSPACE v2 (Boetzer et al. 2011), and gap filling was made by Gapcloser (http://sourceforge.net/projects/soapdenovo2/files/GapCloser/) with small-insert libraries (170–500 bp).

Genome annotation

For the annotation of repetitive elements, we used a combination of homology-based and *ab initio* predictions. RepeatMasker (Smit 1996-2010) and Protein-based RepeatMasking (Smit 1996-2010) were used to search Repbase, which contains a vast amount of known transcriptional elements at the DNA and protein levels. During the process of *ab initio* prediction, RepeatScout (Price et al. 2005) was used to build the *ab initio* repeat library based on k-mer, using the fit-preferred alignment score on the *L. crocea* genome. Contamination and multi-copy genes in the library were filtered out before the RepeatScout library was used to find homologs in the genome and to categorise the found repeats by RepeatMasker (Smit 1996-2010).

Gene models were integrated based on *ab initio* predictions, homologue prediction, and transcription evidence.

Homology-based prediction

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

The protein sequences of seven species (*Danio rerio, Gasterosteus aculeatus, Oreochromis niloticus, Oryzias latipes, Takifugu rubripes, Tetraodon nigroviridis*, and *Homo sapiens*) were aligned to the *L. crocea* assembly using BLAST (E-value ≤ 1e-5), and the matches with length coverage > 30% of the homologous proteins were considered as gene-model candidates. The corresponding homologous genome sequences were then aligned against the matching proteins by using Genewise (Birney et al. 2004) to improve gene models.

Ab initio prediction

Augustus (Stanke and Morgenstern 2005), SNAP (Korf 2004), and GENESCAN (Burge and Karlin 1997) were used for the *ab initio* predictions of gene structures on the repeat-masked assembly.

Transcriptome-based prediction

RNAseq reads from the transcriptomes of the mixed tissues of a female and a male (eleven tissues each) were aligned to the genome assembly by Tophat (Trapnell et al. 2009), which can identify splice junctions between exons. Cufflinks (Mortazavi et al. 2008) was used to obtain transcript structures.

Homology-based, *ab initio* derived and transcript gene sets were integrated to form a comprehensive and non-redundant gene set. The overlap length of each gene was verified by different methods, and genes showing 50% overlap by at least one method were selected. To eliminate false positives (genes only supported by *ab initio* methods), novel genes with the reads per kb of gene model per million of reads (rpkm) ≤ 1 were removed.

Evolutionary and Comparative Analyses

To detect variations in the *L. crocea* genome, we chose nine species (*Larimichthys crocea,*

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Gasterosteus aculeatus, *Takifugu rubripes*, *Tetraodon nigroviridis*, *Oryzias latipes*, *Gadus morhua*, *Danio rerio*, *Gallus gallus*, and *Homo sapiens*). Proteins that were greater than 50 amino acids in size were aligned by BLAST (-p blastp-e 1e-7), and Treefam (Ruan et al. 2008) was used to construct gene families for comparison.

The 2,257 single-copy genes from the gene family analysis were aligned using MUSCLE (Edgar 2004), and alignments were concatenated as a single data set. To reduce the error topology of phylogeny by alignment inaccuracies, we used Gblock (Castresana 2000) (codon model) to remove unreliably aligned sites and gaps in the alignments. The phylogenetic tree and divergence time were calculated using the PAML 3.0 (Yang 1997) package.

Gene family expansion and contraction analyses were performed by cafe (De Bie et al. 2006). For optical, olfactory receptor, and auditory system-related genes, we downloaded the genes from Swissprot or Genebank and predicted their candidates using BLAST and Genewise to determinate copy numbers. Pseudogenes produced by frame shift were removed. Phylogenetic analysis of the expanded gene families was based on maximum likelihood methods by PAML 3.0 (Yang 1997), and the phylogenetic tree was represented by EvolView (Zhang et al. 2012b).

Amino acid sequences from six representative teleosts (*Larimichthys crocea, Gasterosteus aculeatus, Danio rerio, Oryzias latipes, Takifugu rubripes,* and *Tetraodon nigroviridis*) were aligned by BLAST (-p blastp -e 1e-5 -m 8), and reciprocal-best-BLAST-hit methods were used to define orthologous genes in six teleost fish. Because alignment errors are an important concern in molecular data analysis, we made alignments of codon sequences, which are nucleotide sequences that code for proteins, using the PRANK (Loytynoja and

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Goldman 2010) aligner. Positive selection was inferred, based on the branch-site K_a/K_s test by codeml in the PAML 3.0 package (Yang 1997).

Transcriptome under hypoxia

L. crocea (90–100 g) individuals were purchased from a mariculture farm in Ningde, Fujian, China. The fish were maintained at 25 ℃ in aerated water tanks (dissolved oxygen [DO] concentration: 7.8±0.5 mg/L) with a flow-through seawater supply. After 7 days of acclimation, hypoxia-exposure experiments were conducted at 25 ℃ using published methods (Gracey et al. 2001) by bubbling nitrogen gas into an aquarium. The desired concentration of DO was detected by using a DO meter (YSI, Canada). *L. crocea* cannot maintain the aerobic pathway at DO levels below 2.0 mg/L, and it resorts to anaerobic metabolism (Gu and Xu 2011). Therefore, at the onset of hypoxia, the oxygen content in the tank was lowered from 7.8±0.5 mg/L to 1.6±0.2 mg/L over a 10-min period. Brains were harvested from six fish at the 1-, 3-, 6-, 12-, 24-, and 48-h time points and frozen immediately in liquid nitrogen until RNA extraction and transcriptome analyses were performed.

Total RNA was extracted from the tissues of *L. crocea* using the guanidinium thiocyanate-phenol-chloroform extraction method (Trizol, Invitrogen, USA), according to the manufacturer's protocol. The libraries were sequenced by using the Illumina HiSeq 2000 sequencing platform with the paired-end sequencing module (Zhang et al. 2012a). After removing low-quality reads, RNAseq reads were aligned to the *L. crocea* genome with SOAPaligner/SOAP2 (Li et al. 2009). The alignment was utilised to calculate the distribution of reads on reference genes and to perform coverage analysis. If an alignment result passed quality control (alignment ratio > 70%), we proceeded in gene expression calculations and

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

differential expression comparisons.

LC–MS/MS analyses and mucus protein identification

Skin mucus was collected from six healthy *L. crocea* individuals under air exposure as previously described (Subramanian et al. 2008). Briefly, the fish were anesthetised with a sub-lethal dose of Tricaine-S (100 mg/L), and transferred gently to a sterile plastic bag for 3 min to slough off the mucus under air exposure. To exclude the cell contamination, mucus was diluted in fresh, cold phosphate-buffered saline and drop-splashed onto slides, which were then air-dried. After staining with 10% Giemsa dye (Sigma, St Louis, MO, USA) for 20 min, the mucus was observed under a Nikon microscope with a 20 × objective. No cell was observed.

Proteins were extracted from a pool of skin mucus of six fish by the trichloroacetic acid-acetone precipitation method and digested by the trypsin gold (Promega, USA). The peptides were then separated by the strong cation exchange chromatography using a Shimadzu LC-20AB HPLC Pump system (Kyoto, Japan). Data acquisition was performed with a Triple TOF 5600 System (AB SCIEX, Concord, ON) fitted with a Nanospray III source (AB SCIEX, Concord, ON). All spectra were mapped by MASCOT server version 2.3.02 against the database of the *L. crocea* genome with the parameters as follows: peptide mass tolerance 0.05 Da; fragment mass tolerance 0.1 Da; fixed modifications "Carbamidomethyl (C)"; and variable modifications "Gln->pyro-Glu (N-term Q), Oxidation (M), Deamidated (NQ)". For further analyses of the function of the mucus proteome, we selected proteins with more than two unique peptides.

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Data access

The large yellow croaker whole-genome sequence has been deposited at the DNA Data Bank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL) nucleotide sequencing database and GenBank under the same accession XXX (The data have been submitted and we are waiting for return of the accession numbers). All short-read data of WGS and BAC have been deposited in the Short Read Archive (SRA) under accession SRA159210 and SRA159209 respectively. Raw sequencing data for the transcriptome have been deposited in the Gene Expression Omnibus (GEO) under accession GSE57608.

Acknowledgements

This work was supported by the Nation '863' Project (2012AA092202); National Basic Research Program of China (2012CB114402); National Natural Science Foundation of China (31125027, 31372556), and fund of Xiamen south ocean research center (13GZP002NF08). We thank XinXin You, Yue Feng, GuanXing Chen, RiBei Fu, JinTu Wang, Ying Qiu and Jie Bai (BGI-Shenzhen, China) very much for their hard work in preparing the manuscript and analyses. We thank Jiafu Liu (Fujian Ningde Municipal Station of Fishery Technical Extension) for sample collection. We also thank Qiong Liu and JiaZan Ni (College of Life Sciences, Shenzhen University, China) for help with selenoprotein prediction.

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Figure legends

Figure 1. Phylogenetic tree of and orthologous genes in *L. crocea* and other vertebrates. (**A**) The phylogenetic tree was constructed from 2,257 single-copy genes with 3.18 M reliable sites by maximum likelihood methods. The red points on six of the internal nodes indicate fossil calibration times in the analysis. Blue numbers indicate the divergence time (Myr, million years ago), and the green and red numbers represent the expanded and extracted gene families, respectively, in *L. crocea*. (**B**) The different types of orthologous relationships are shown. "1:1:1" = universal single-copy genes; "N:N:N" = orthologues exist in all genomes; "Fish" = fish-specific genes; "SD" = genes that have undergone species-specific duplication; "Homology" = genes with an e-value less than 1e-5 by BLAST but do not cluster to a gene family; "ND" = species-specific genes; and "Others" = orthologues that do not fit into the other categories. (**C**) The shared and unique gene families in five teleost fish are shown in the Venn diagram. (**D**) Distribution of the identity values of orthologous genes is compared among *L. crocea* and other teleosts.

Figure 2. Characterisation of the T-cell lineages in *L. crocea* adaptive immunity and the expanded genes in antiviral immunity.

(**A**) A schematic diagram summarising genes related to different T-cell lineages in *L. crocea* is shown. The inducible factors, the main regulatory transcriptional factors, and the immune effectors of T cells are present in green, blue, and orange backgrounds, respectively. The genes that have been annotated by genome survey are shown in black, and the unannotated genes are shown in red. The dashed square outlines the possible incomplete Th2 cell-mediated humoral immunity of *L. crocea*. (**B**) Several key genes are expanded in the

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

antiviral immunity pathways in *L. crocea*. The genes that have been identified in the *L. crocea* genome are shown in orange boxes, and the lost gene (*RIG-I*) is shown in the grey box. LGP2 is able to bind to double-stranded RNA (dsRNA) to trigger interferon production, but the adaptor molecule of LGP2 is still unknown in fish. The red boxes indicate gene families (*TRIM25*, *cGAS*, *DDX41*, and *NLRC3*) that are expanded in *L. crocea*. The arrow represents induction, and the interrupted line represents inhibition.

Figure 3. Hypoxia stress exerts responses involving the HPA and HPT axes.

Under hypoxia, a potential neuro-endocrine-immune/metabolism network contributes to the regulation of moderate inflammation and the maintenance of energy balance. Hypoxia can initially promote *ET-1* and *ADM* expression, after which it increases pivotal inflammatory cytokines, such as IL-6 and TNF-α in the brain, to induce cerebral inflammation. ET-1/ADM and IL-6/TNF-α form a positive feedback loop to amplify cerebral inflammation. Afterwards, the hypothalamic-pituitary-adrenal (HPA) axis-glucocorticoids pathway and SOCS family members (SOCS-1 and SOCS-3) can inhibit *IL-6/TNF-α* expression, which constitutes the negative feedback loops with IL-6/TNF-α to modify cerebral inflammation. The HPT axis was inhibited in *L. crocea* brains during the early period of hypoxia, thus leading to a decrease in thyroid hormone production. Thyroid hormones subsequently inhibited ribosomal biogenesis and protein translation by the PI3K-Akt-mTOR-S6K signaling pathway. Down-regulation of HPT axis-thyroid hormones also repressed the tricarboxylic acid (TCA) cycle and accelerated the anaerobic glycolytic pathway in the brain, along with increases in the exposure to hypoxia. Genes related to the neuro-endocrine system (orange), immunity (red), and metabolic system and protein synthesis (blue) are indicated. The outer border

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

indicates the brain of *L. crocea*. The arrow represents promotion, and the interrupted line represents inhibition. Solid lines indicate direct relationships between genes. Dashed lines indicate that more than one step is involved in the process.

Figure 4. Skin mucus proteins are overexpressed in air-exposed *L. crocea*.

(**A**) The distribution of mucus proteins in the molecular function class of Gene Ontology is shown. The over-represented functional categories are indicated in the pie chart. (**B**) A representation of the functional mechanisms of the mucus barrier is shown. The continuously replenished thick mucus layer can retain a large number of antioxidant, immune, oxygen-binding, and ion-binding molecules, which are involved in antioxidant functions, immune defence, oxygen transport, and osmotic and ionic regulation, respectively.

References

Benita Y, Kikuchi H, Smith AD, Zhang MQ, Chung DC, Xavier RJ. 2009. An integrative genomics approach identifies Hypoxia Inducible Factor-1 (HIF-1)-target genes that form the core response to hypoxia. *Nucleic Acids Res* **37**(14): 4587-4602.

Birney E, Clamp M, Durbin R. 2004. GeneWise and Genomewise. *Genome research* **14**(5): 988-995.

Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* **27**(4): 578-579.

Bona E, Andersson AL, Blomgren K, Gilland E, Puka-Sundvall M, Gustafson K, Hagberg H. 1999. Chemokine and inflammatory cell response to hypoxia-ischemia in immature rats. *Pediatr Res* **45**(4 Pt 1): 500-509.

Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA. *Journal of molecular biology* **268**(1): 78-94.

Castanier C, Zemirli N, Portier A, Garcin D, Bidere N, Vazquez A, Arnoult D. 2012. MAVS ubiquitination by the E3 ligase TRIM25 and degradation by the proteasome is involved in type I interferon production after activation of the antiviral RIG-I-like receptors. *BMC Biol* **10**: 44.

Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular biology and evolution* **17**(4): 540-552.

Chang M, Collet B, Nie P, Lester K, Campbell S, Secombes CJ, Zou J. 2011. Expression and functional characterization of the RIG-I-like receptors MDA5 and LGP2 in Rainbow

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

trout (*Oncorhynchus mykiss*). *J Virol* **85**(16): 8403-8412.

Chen S, Zhang G, Shao C, Huang Q, Liu G, Zhang P, Song W, An N, Chalopin D, Volff JN et al. 2014. Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nat Genet* **46**(3): 253-260.

Cossins AR, Crawford DL. 2005. Fish as models for environmental genomics. *Nat Rev Genet* **6**(4): 324-333.

Cross CE, Halliwell B, Allen A. 1984. Antioxidant protection: a function of tracheobronchial and gastrointestinal mucus. *Lancet* **1**(8390): 1328-1330.

Dayan F, Roux D, Brahimi-Horn MC, Pouyssegur J, Mazure NM. 2006. The oxygen sensor factor-inhibiting hypoxia-inducible factor-1 controls expression of distinct genes through the bifunctional transcriptional character of hypoxia-inducible factor-1alpha. *Cancer Res* **66**(7): 3688-3698.

De Bie T, Cristianini N, Demuth JP, Hahn MW. 2006. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**(10): 1269-1271.

Earley S, Nelin LD, Chicoine LG, Walker BR. 2002. Hypoxia-induced pulmonary endothelin-1 expression is unaltered by nitric oxide. *J Appl Physiol (1985)* **92**(3): 1152-1158.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research* **32**(5): 1792-1797.

Eisen MD, Ryugo DK. 2007. Hearing molecules: contributions from genetic deafness. *Cell Mol Life Sci* **64**(5): 566-580.

Gao D, Wu J, Wu YT, Du F, Aroh C, Yan N, Sun L, Chen ZJ. 2013. Cyclic GMP-AMP synthase is an innate immune sensor of HIV and other retroviruses. *Science* **341**(6148): 903-906.

Gracey AY, Troll JV, Somero GN. 2001. Hypoxia-induced gene expression profiling in the euryoxic fish *Gillichthys mirabilis*. *Proc Natl Acad Sci U S A* **98**(4): 1993-1998.

Gu X, Xu Z. 2011. Effect of hypoxia on the blood of large yellow croaker (*Pseudosciaena crocea*). *Chinese Journal of Oceanology and Limnology* **29**(3): 524.

Guzman-Aranguez A, Mantelli F, Argueso P. 2009. Mucin-type O-glycans in tears of normal subjects and patients with non-Sjogren's dry eye. *Invest Ophthalmol Vis Sci* **50**(10): 4581-4587.

Hansen JD, Vojtech LN, Laing KJ. 2011. Sensing disease and danger: a survey of vertebrate PRRs and their origins. *Dev Comp Immunol* **35**(9): 886-897.

Hayashi R, Wada H, Ito K, Adcock IM. 2004. Effects of glucocorticoids on gene transcription. *Eur J Pharmacol* **500**(1-3): 51-62.

Herman JP, Cullinan WE. 1997. Neurocircuitry of stress: central control of the hypothalamo-pituitary-adrenocortical axis. *Trends Neurosci* **20**(2): 78-84.

Hou TD, Du JZ. 2005. Norepinephrine attenuates hypoxia-inhibited thyrotropin-releasing hormone release in median eminence and paraventricular nucleus of rat hypothalamus. *Neuro Endocrinol Lett* **26**(1): 43-49.

Jiang L, Liu Q, Ni J. 2010. In silico identification of the sea squirt selenoproteome. *BMC Genomics* **11**: 289.

Kaur C, Ling EA. 2008. Blood brain barrier in hypoxic-ischemic conditions. *Curr Neurovasc*

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Res **5**(1): 71-81.

Kenessey A, Ojamaa K. 2006. Thyroid hormone stimulates protein synthesis in the cardiomyocyte by activating the Akt-mTOR and p70S6K pathways. *J Biol Chem* **281**(30): 20666-20672.

Kitamuro T, Takahashi K, Nakayama M, Murakami O, Hida W, Shirato K, Shibahara S. 2000. Induction of adrenomedullin during hypoxia in cultured human glioblastoma cells. *J Neurochem* **75**(5): 1826-1833.

Korf I. 2004. Gene finding in novel genomes. *BMC bioinformatics* **5**: 59.

Lee DC, Hassan SS, Romero R, Tarca AL, Bhatti G, Gervasi MT, Caruso JA, Stemmer PM, Kim CJ, Hansen LK et al. 2011. Protein profiling underscores immunological functions of uterine cervical mucus plug in human pregnancy. *J Proteomics* **74**(6): 817-828.

Lemos Vde A, dos Santos RV, Lira FS, Rodrigues B, Tufik S, de Mello MT. 2013. Can high altitude influence cytokines and sleep? *Mediators Inflamm* **2013**: 279365.

Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. 2009. SOAP2: an improved ultrafast tool for short read alignment. *Bioinformatics* **25**(15): 1966-1967.

Li W, Sorensen PW, Gallaher DD. 1995. The olfactory system of migratory adult sea lamprey (*Petromyzon marinus*) is specifically and acutely sensitive to unique bile acids released by conspecific larvae. *J Gen Physiol* **105**(5): 569-587.

Liu XD, Zhao GT, Cai MY, Wang ZY. 2013. Estimated genetic parameters for growth-related traits in large yellow croaker *Larimichthys crocea* using microsatellites to assign parentage. *Journal of fish biology* **82**(1): 34-41.

Liu YY, Cao MJ, Zhang ML, Hu JW, Zhang YX, Zhang LJ, Liu GM. 2014. Purification, characterization and immunoreactivity of beta'-component, a major allergen from the roe of large yellow croaker (*Pseudosciaena crocea*). *Food and chemical toxicology : an international journal published for the British Industrial Biological Research Association* **72C**: 111-121.

Loytynoja A, Goldman N. 2010. webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC bioinformatics* **11**: 579.

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y et al. 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. *GigaScience* **1**(1): 18.

Mastorakos G, Chrousos GP, Weber JS. 1993. Recombinant interleukin-6 activates the hypothalamic-pituitary-adrenal axis in humans. *J Clin Endocrinol Metab* **77**(6): 1690-1694.

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nature methods* **5**(7): 621-628.

Mu Y, Ding F, Cui P, Ao J, Hu S, Chen X. 2010. Transcriptome and expression profiling analysis revealed changes of multiple signaling pathways involved in immunity in the large yellow croaker during *Aeromonas hydrophila* infection. *BMC Genomics* **11**: 506.

Mu Y, Li M, Ding F, Ding Y, Ao J, Hu S, Chen X. 2014. De novo characterization of the spleen transcriptome of the large yellow croaker (*Pseudosciaena crocea*) and analysis of the immune relevant genes and pathways involved in the antiviral response. *PLoS One* **9**(5): e97471.

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Nadeau S, Rivest S. 2003. Glucocorticoids play a fundamental role in protecting the brain during innate immune response. *J Neurosci* **23**(13): 5536-5544.

Niimura Y. 2009. On the origin and evolution of vertebrate olfactory receptor genes: comparative genome analysis among 23 chordate species. *Genome Biol Evol* **1**: 34-44.

Ning Y, Liu X, Wang ZY, Gou W, Li Y, Xie F. 2007. A genetic map of large yellow croaker *Pseudosciaena crocea*. *Aquaculture* **264** (1): 16-26.

Pluta K, McGettigan PA, Reid CJ, Browne JA, Irwin JA, Tharmalingam T, Corfield A, Baird A, Loftus BJ, Evans AC et al. 2012. Molecular aspects of mucin biosynthesis and mucus formation in the bovine cervix during the periestrous period. *Physiol Genomics* **44**(24): 1165-1178.

Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large genomes. *Bioinformatics* **21 Suppl 1**: i351-358.

Richards JG. 2011. Physiological, behavioral and biochemical adaptations of intertidal fishes to hypoxia. *J Exp Biol* **214**(Pt 2): 191-199.

Rodriguez-Pineiro AM, Bergstrom JH, Ermund A, Gustafsson JK, Schutte A, Johansson ME, Hansson GC. 2013. Studies of mucus in mouse stomach, small intestine, and colon. II. Gastrointestinal mucus proteome reveals Muc2 and Muc5ac accompanied by a set of core proteins. *Am J Physiol Gastrointest Liver Physiol* **305**(5): G348-356.

Ruan J, Li H, Chen Z, Coghlan A, Coin LJ, Guo Y, Heriche JK, Hu Y, Kristiansen K, Li R et al. 2008. TreeFam: 2008 Update. *Nucleic acids research* **36**(Database issue): D735-740.

Sabell I, Morata P, Quesada J, Morell M. 1985. Effect of thyroid hormones on the glycolytic enzyme activity in brain areas of the rat. *Enzyme* **34**(1): 27-32.

Schartl M, Walter RB, Shen Y, Garcia T, Catchen J, Amores A, Braasch I, Chalopin D, Volff JN, Lesch KP et al. 2013. The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. *Nat Genet* **45**(5): 567-572.

Shephard KL. 1994. Functions for fish mucus. *Reviews in Fish Biology and Fisheries* **4**(4): 401-429.

Smit A, Hubley, R & Green, P. RepeatMasker Open-3.0. 1996-2010 http://www.repeatmasker.org.

Sorrells SF, Sapolsky RM. 2007. An inflammatory review of glucocorticoid actions in the CNS. *Brain Behav Immun* **21**(3): 259-272.

Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints. *Nucleic acids research* **33**(Web Server issue): W465-467.

Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrom M, Gregers TF, Rounge TB, Paulsen J, Solbakken MH, Sharma A et al. 2011. The genome sequence of Atlantic cod reveals a unique immune system. *Nature* **477**(7363): 207-210.

Su Y. 2004. Breeding and Farming of *Pseudosciaena Crocea*. 1st ed. China Ocean Press, Beijing.

Subramanian S, Ross NW, MacKinnon SL. 2008. Comparison of antimicrobial activity in the epidermal mucus extracts of fish. *Comp Biochem Physiol B Biochem Mol Biol* **150**(1): 85-92.

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Takahashi K, Udono-Fujimori R, Totsune K, Murakami O, Shibahara S. 2003. Suppression of cytokine-induced expression of adrenomedullin and endothelin-1 by dexamethasone in T98G human glioblastoma cells. *Peptides* **24**(7): 1053-1062.

Taylor MM, Bagley SL, Samson WK. 2005. Intermedin/adrenomedullin-2 acts within central nervous system to elevate blood pressure and inhibit food and water intake. *Am J Physiol Regul Integr Comp Physiol* **288**(4): R919-927.

Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**(9): 1105-1111.

van der Meer DL, van den Thillart GE, Witte F, de Bakker MA, Besser J, Richardson MK, Spaink HP, Leito JT, Bagowski CP. 2005. Gene expression profiling of the long-term adaptive response to hypoxia in the gills of adult zebrafish. *Am J Physiol Regul Integr Comp Physiol* **289**(5): R1512-1519.

Vandermarliere E, Ghesquiere B, Jonckheere V, Gevaert K, Martens L. 2014. Unraveling the specificities of the different human methionine sulfoxide reductases. *Proteomics*, doi: 10.1002/pmic.201300357.

Yang N, Ray DW, Matthews LC. 2012. Current concepts in glucocorticoid resistance. *Steroids* **77**(11): 1041-1049.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. *Computer applications in the biosciences : CABIOS* **13**(5): 555-556.

Ye H, Liu Y, Liu X, Wang X, Wang Z. 2014. Genetic Mapping and QTL Analysis of Growth Traits in the Large Yellow Croaker *Larimichthys crocea*. *Marine biotechnology*, doi: 10.1007/s10126-014-9590-z.

Yen PM. 2001. Physiological and molecular basis of thyroid hormone action. *Physiol Rev* **81**(3): 1097-1142.

Yu S, Mu Y, Ao J, Chen X. 2010. Peroxiredoxin IV regulates pro-inflammatory responses in large yellow croaker (*Pseudosciaena crocea*) and protects against bacterial challenge. *J Proteome Res* **9**(3): 1424-1436.

Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H et al. 2012a. The oyster genome reveals stress adaptation and complexity of shell formation. *Nature* **490**(7418): 49-54.

Zhang H, Gao S, Lercher MJ, Hu S, Chen WH. 2012b. EvolView, an online tool for visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Res* **40**(Web Server issue): W569-572.

Zhang L, Mo J, Swanson KV, Wen H, Petrucelli A, Gregory SM, Zhang Z, Schneider M, Jiang Y, Fitzgerald KA et al. 2014. NLRC3, a Member of the NLR Family of Proteins, Is a Negative Regulator of Innate Immune Signaling Induced by the DNA Sensor STING. *Immunity* **40**(3): 329-341.

Zhang Z, Yuan B, Bao M, Lu N, Kim T, Liu YJ. 2011. The helicase DDX41 senses intracellular DNA mediated by the adaptor STING in dendritic cells. *Nat Immunol* **12**(10): 959-965.

Zhou Y, Yan X, Xu S, Zhu P, He X, Liu J. 2011. Family structure and phylogenetic analysis of odorant receptor genes in the large yellow croaker (*Larimichthys crocea*). *BMC Evol Biol* **11**: 237.

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

Table 1. Summary of the *Larimichthys crocea* genome

Sequencing	Insert Size (bp)	Total Data (Gbp)	Coverage (×)
BAC	120 k	324.73	469.94
WGS	170–500	36.22	52.42
	2 k–40 k	34.26	49.58
Assembly	**N50 (bp)**	**Longest (kbp)**	**Size (Mbp)**
Contig	63.11 k	716.89	661.33
Scaffold	1.03 M	4,914.79	678.96
Annotation	**Number**	**Total Length (Mbp)**	**Percentage Of Genome (%)**
Repeat	1,454,906	122.92	18.10
Gene	25,401	350.95	51.69
Exon	251,617	44.86	6.61

BAC = bacterial artificial chromosome; WGS = whole genome shotgun.

Figure 1

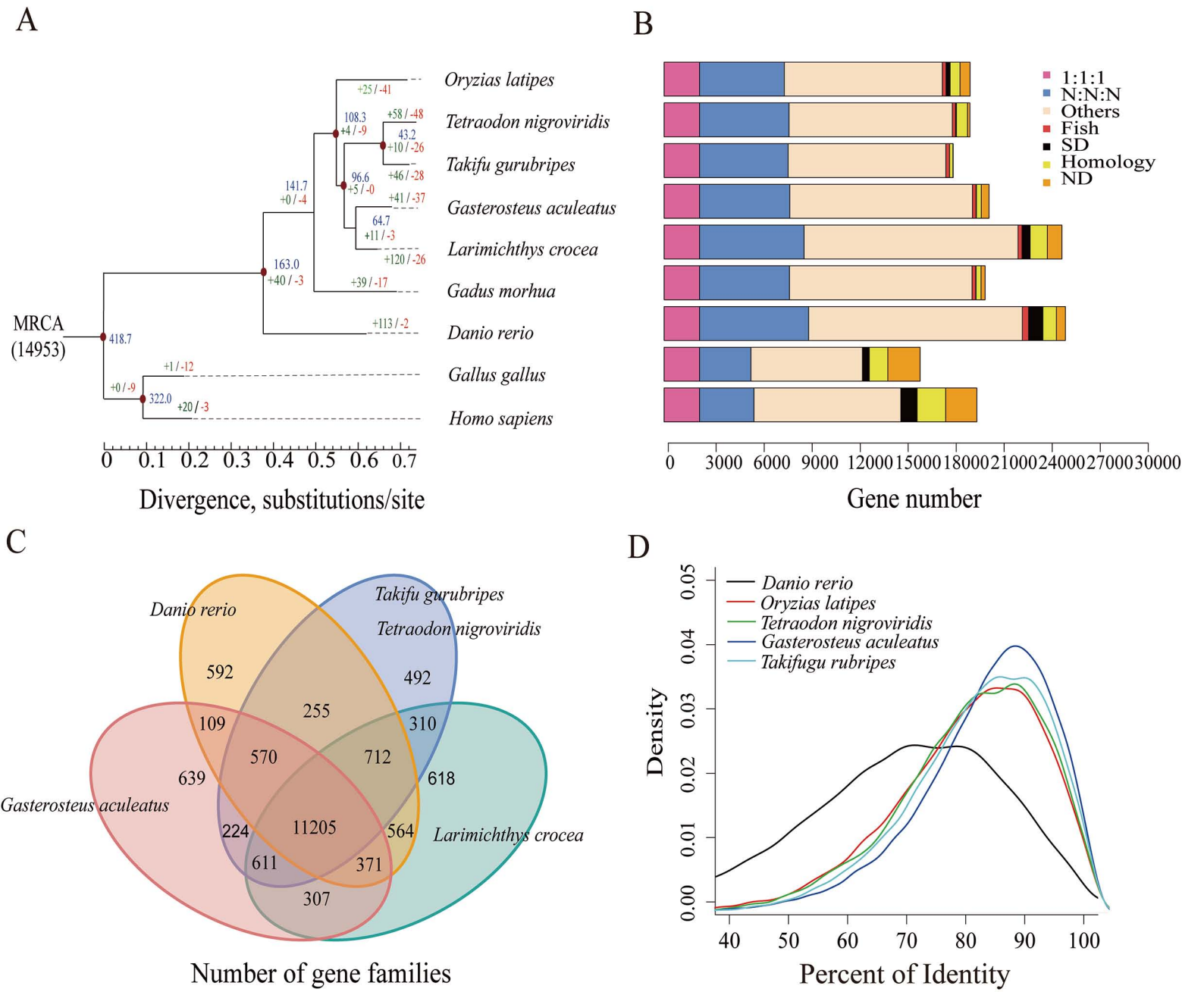
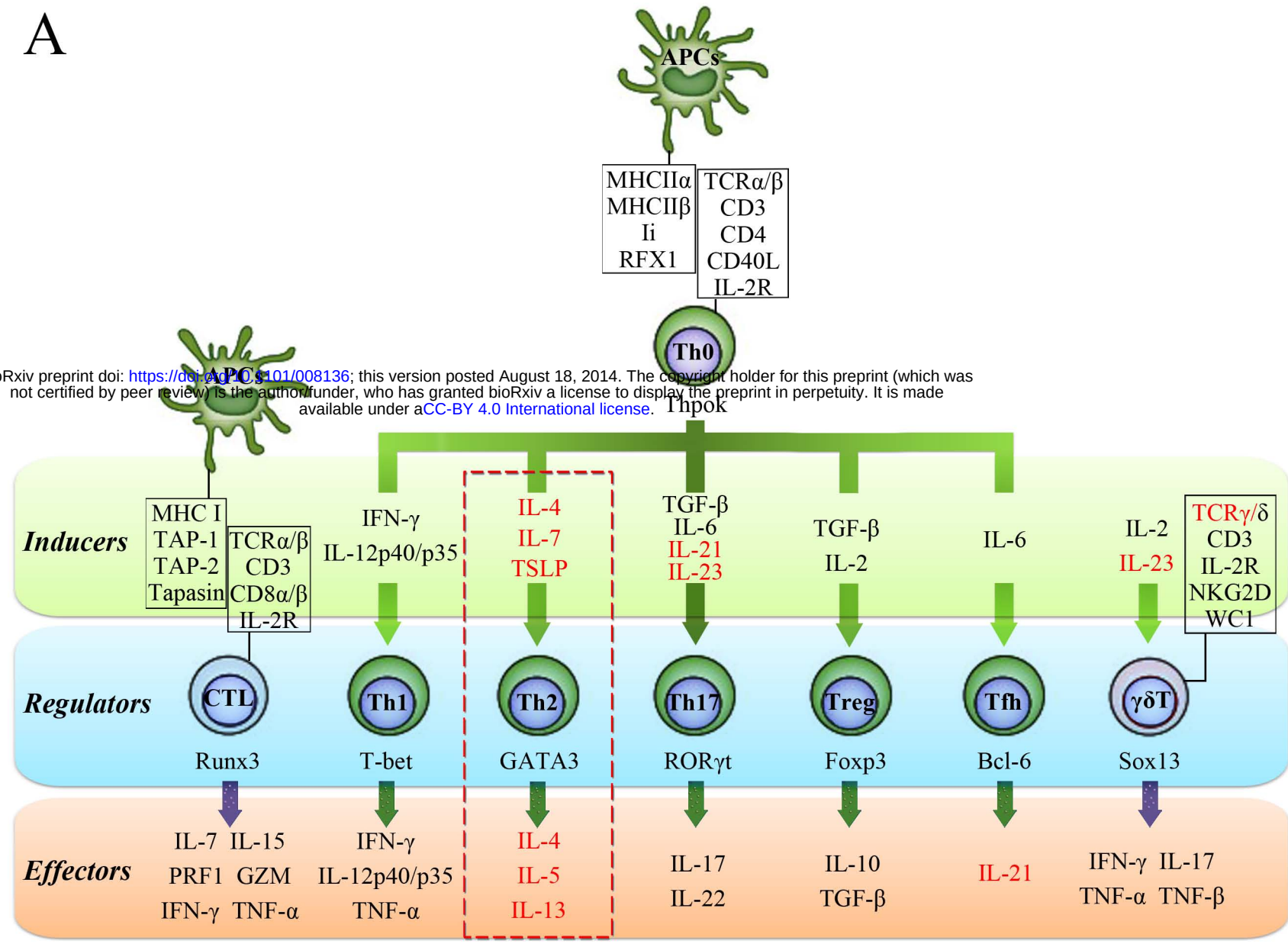

A
Oryzias latipes
Tetraodon nigroviridis
Takifu gurubripes
Gasterosteus aculeatus
Larimichthys crocea
Gadus morhua
Danio rerio
Gallus gallus
Homo sapiens
MRCA
(14953)
418.7
322.0
163.0
+40 / -3
141.7
+0 / -4
108.3
+4 / -9
96.6
+5 / -0
64.7
+11 / -3
43.2
+10 / -26
+25 / -41
+58 / -48
+46 / -28
+41 / -37
+120 / -26
+39 / -17
+113 / -2
+1 / -12
+0 / -9
+20 / -3
0 0.1 0.2 0.3 0.4 0.5 0.6 0.7
Divergence, substitutions/site
B
1:1:1
N:N:N
Others
Fish
SD
Homology
ND
0 3000 6000 9000 12000 15000 18000 21000 24000 27000 30000
Gene number
C
Danio rerio
Takifu gurubripes
Tetraodon nigroviridis
Gasterosteus aculeatus
Larimichthys crocea
592
492
255
109
310
570
712
639
618
224
11205
564
611
371
307
Number of gene families
D
Danio rerio
Oryzias latipes
Tetraodon nigroviridis
Gasterosteus aculeatus
Takifugu rubripes
Density
0.00 0.01 0.02 0.03 0.04 0.05
40 50 60 70 80 90 100
Percent of Identity

Figure 2

bioRxiv preprint doi: https://doi.org/10.1101/008136; this version posted August 18, 2014. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.

A



B

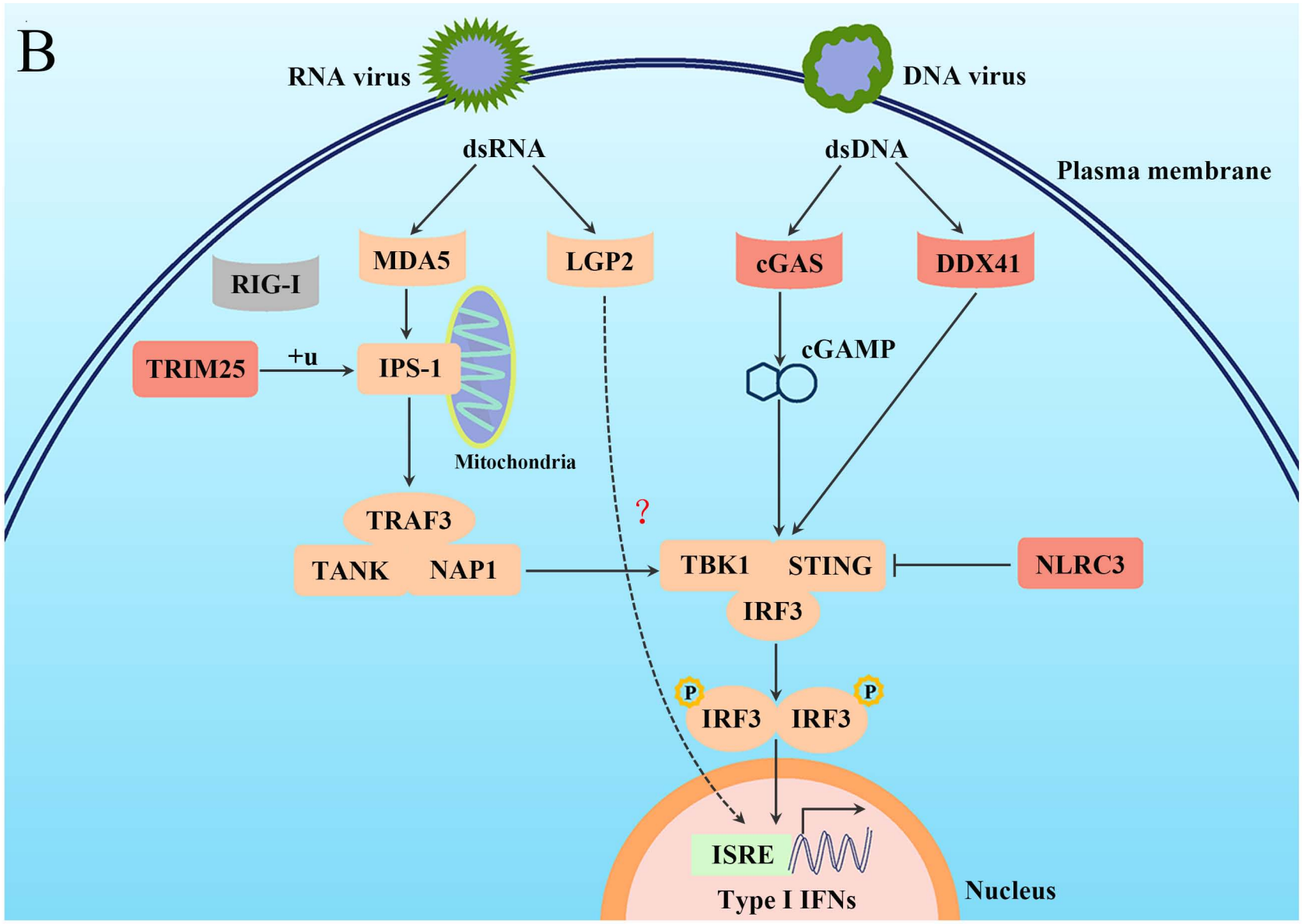


Figure 3

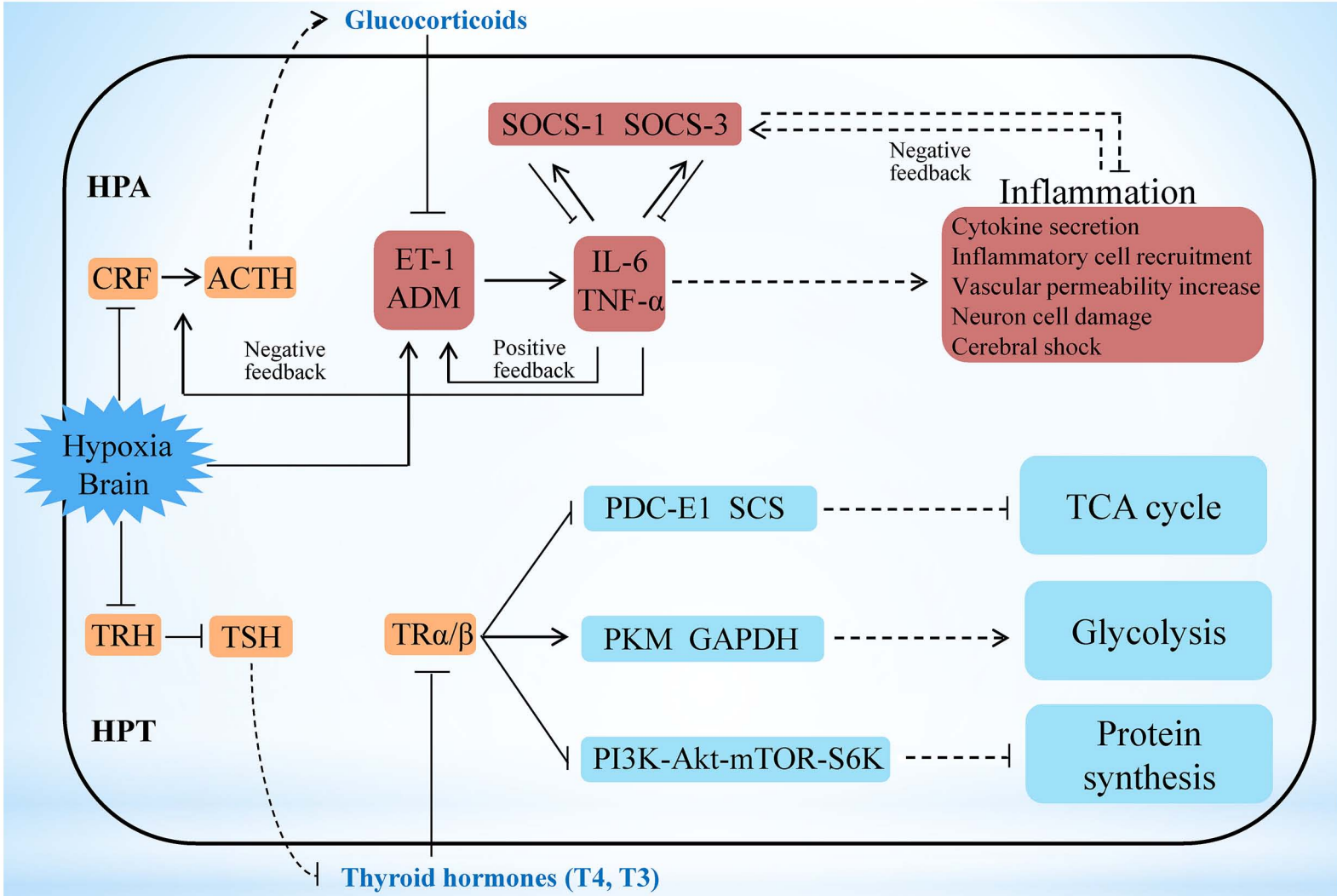
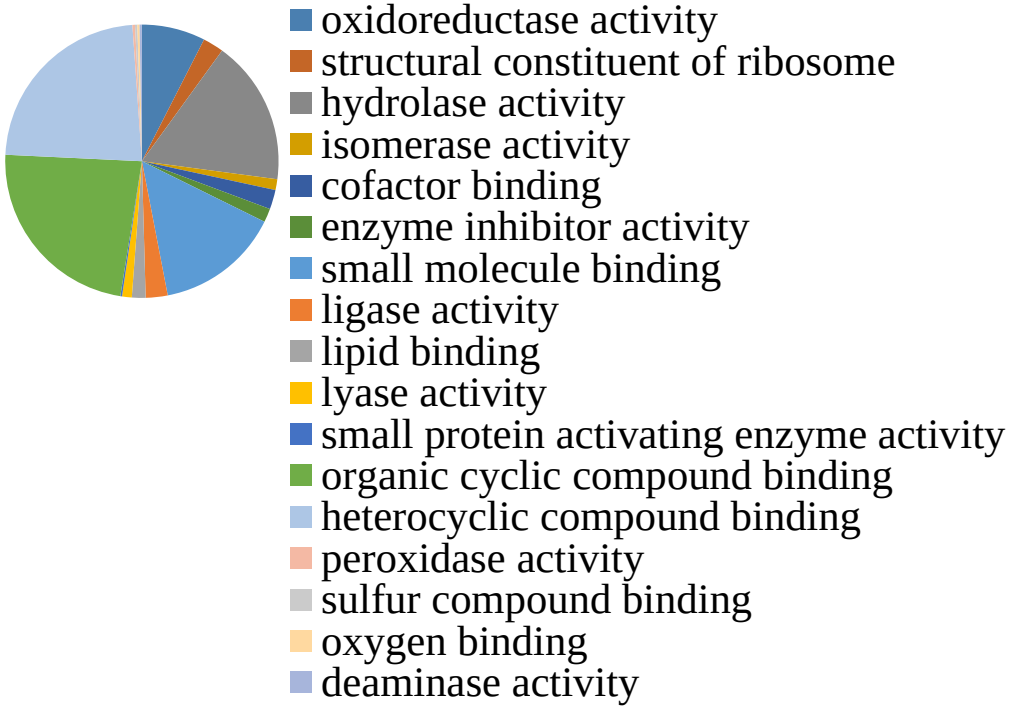
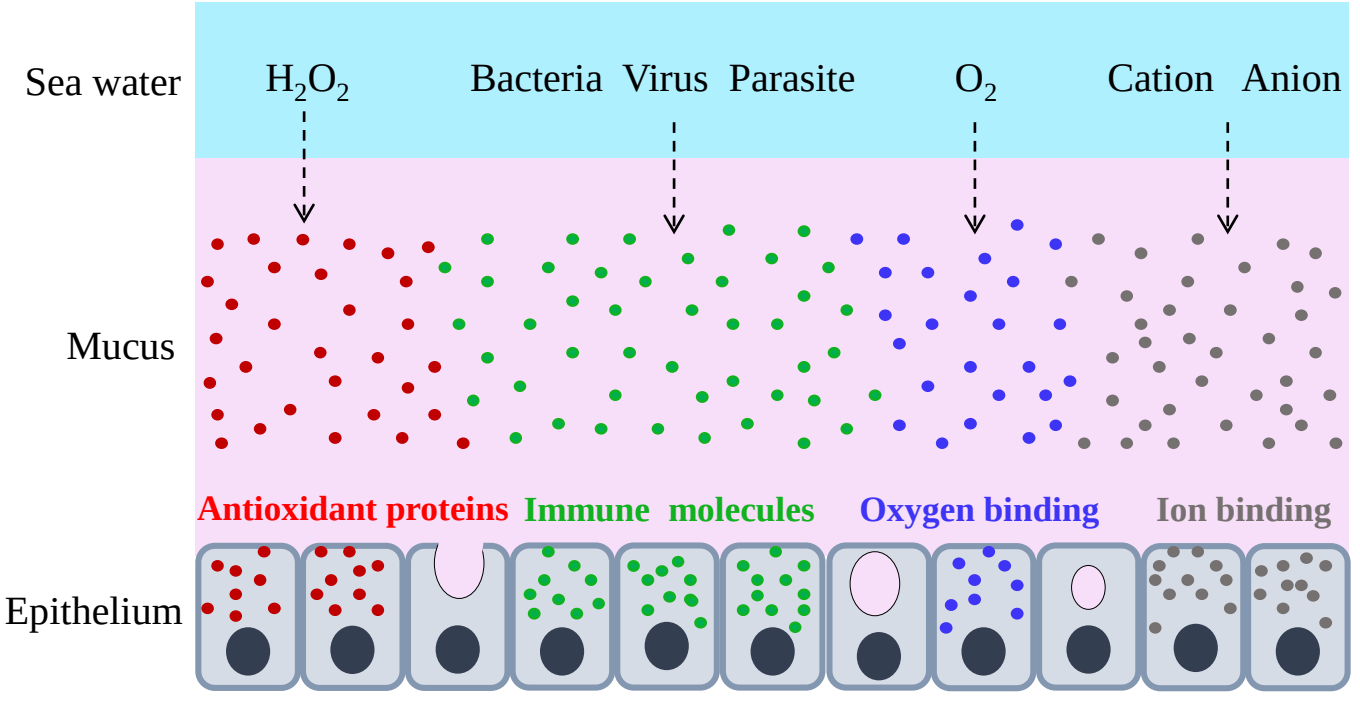

Glucocorticoids
SOCS-1 SOCS-3
Negative feedback
Inflammation
HPA
CRF
ACTH
ET-1
ADM
IL-6
TNF-α
Cytokine secretion
Inflammatory cell recruitment
Vascular permeability increase
Neuron cell damage
Cerebral shock
Negative feedback
Positive feedback
Hypoxia Brain
PDC-E1 SCS
TCA cycle
TRH
TSH
TRα/β
PKM GAPDH
Glycolysis
HPT
PI3K-Akt-mTOR-S6K
Protein synthesis
Thyroid hormones (T4, T3)

Figure 4

A


oxidoreductase activity
structural constituent of ribosome
hydrolase activity
isomerase activity
cofactor binding
enzyme inhibitor activity
small molecule binding
ligase activity
lipid binding
lyase activity
small protein activating enzyme activity
organic cyclic compound binding
heterocyclic compound binding
peroxidase activity
sulfur compound binding
oxygen binding
deaminase activity

B


Sea water
Mucus
Epithelium
H_2O_2
Bacteria
Virus
Parasite
O_2
Cation
Anion
Antioxidant proteins
Immune molecules
Oxygen binding
Ion binding