

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

3
4 Jingqun Ao^{1†}, Yinnan Mu^{1†}, Li-Xin Xiang^{2†}, DingDing Fan^{3†}, MingJi Feng^{3†}, Shicui Zhang^{4†},
5 Qiong Shi³, Lv-Yun Zhu², Ting Li¹, Yang Ding¹, Li Nie², Qiuhua Li¹, Wei-ren Dong², Liang
6 Jiang⁵, Bing Sun⁴, XinHui Zhang³, Mingyu Li¹, Hai-Qi Zhang², ShangBo Xie³, YaBing Zhu³,
7 XuanTing Jiang³, Xianhui Wang¹, Pengfei Mu¹, Wei Chen¹, Zhen Yue³, Zhuo Wang³, Jun
8 Wang^{3*}, Jian-Zhong Shao^{2*} and Xinhua Chen^{1*}

9
10 ¹ Key Laboratory of Marine Biogenetics and Resources, Third Institute of Oceanography,
11 State Oceanic Administration; Fujian Collaborative Innovation Center for Exploitation and
12 Utilization of Marine Biological Resources; Key Laboratory of Marine Genetic Resources of
13 Fujian Province, Xiamen 361005, P.R. China

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

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

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

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

18
19 [†]These authors contributed equally to the work.

20 ^{*}To whom correspondence should be addressed: Xinhua Chen: Tel: +86-592-2195297, Fax:
21 +86-592-2195297, Email: chenxinhua@tio.org.cn; Jun Wang: Tel: +86-755-25274247, Fax:
22 +86-755-25274247, Email: wangj@genomics.org.cn; Jian-Zhong Shao: Tel:
23 +86-571-88206582, Fax: +86-571-88206582, Email: shaojz@zju.edu.cn.

24
25 **Keywords: Large yellow croaker, *Larimichthys crocea*; Genome sequencing; Hypoxia;**
26 **Air exposure; Transcriptome; Proteome.**

1 **Abstract**

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

22

1 **Introduction**

2 Teleost fish, nearly half of all living vertebrates, display an amazing level of diversity in body
3 forms, behaviors, physiologies, and environments that they occupy. Strategies for coping with
4 diverse environmental stresses have evolved in different teleost species. Therefore, teleost
5 fish are considered to be good models for investigating the adaptation and response to many
6 natural and anthropogenic environmental stressors (Gracey et al. 2001; Cossins and Crawford
7 2005; van der Meer et al. 2005). Recent genome-sequencing projects in several fish have
8 provided insights into the molecular and genetic mechanisms underlying their responses to
9 some environmental stressors (Star et al. 2011; Scharl et al. 2013; Chen et al. 2014).
10 However, to better clarify the conserved and differentiated features of the adaptive response
11 to specific stresses and to trace the evolutionary process of environmental adaptation and
12 response in teleost fish, insight from more teleost species with different evolutionary
13 positions, such as Perciformes, is required. Perciformes are by far the largest and most
14 diverse order of vertebrates, and thus offer a large number of models of adaptation and
15 response to various environmental stresses.

16 The large yellow croaker, *Larimichthys crocea* (*L. crocea*), is a temperate-water migratory
17 fish that belongs to the order Perciformes and the family Sciaenidae. It is mainly distributed
18 in the southern Yellow Sea, the East China Sea, and the northern South China Sea. *L. crocea*
19 is one of the most economically important marine fish in China and East Asian countries due
20 to its rich nutrients and trace elements, especially selenium. In China, the annual yield from *L.*
21 *crocea* aquaculture exceeds that of any other net-cage-farmed marine fish species (Liu et al.
22 2013; Liu et al. 2014). Recently, the basic studies on genetic improvement for growth and

1 disease resistance traits of *L. crocea* are increasingly performed for farming purpose (Ning et
2 al. 2007; Mu et al. 2010; Liu et al. 2013; Ye et al. 2014). *L. crocea* also exhibits peculiar
3 behavioral and physiological characteristics, such as loud sound production, high sensitivity
4 to sound, and well-developed photosensitive and olfactory systems (Su 2004; Zhou et al.
5 2011). Most importantly, *L. crocea* is especially sensitive to various environmental stresses,
6 such as hypoxia and air exposure. For example, the response of its brain to hypoxia is quick
7 and robust, and a large amount of mucus is secreted from its skin when it is exposed to air
8 (Su 2004; Gu and Xu 2011). These traits may render *L. crocea* a good model for investigating
9 the response mechanisms to environmental stress. Several studies have reported
10 transcriptomic and proteomic responses of *L. crocea* to pathogenic infections or immune
11 stimuli (Mu et al. 2010; Yu et al. 2010; Mu et al. 2014). The effect of hypoxia on the blood
12 physiology of *L. crocea* has been evaluated (Gu and Xu 2011). However, little is known
13 about the molecular response mechanisms of *L. crocea* against environmental stress.

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

19

20 **Results**

21 **Genome features**

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

1 hierarchical assembly strategy for the *L. crocea* genome to overcome the high levels of
2 genome heterozygosity (**Table 1; Supplemental Fig. S1-S2**). The 42,528 BACs were
3 sequenced by the HiSeq 2000 platform and each BAC was assembled by SOAPdenovo (Luo
4 et al. 2012) (**Supplemental Table S1**). The total length of all combined BACs was 3,006
5 megabases (Mb), which corresponded to approximately 4.3-fold genome coverage
6 (**Supplemental Tables S2-S3**). All BAC assemblies were then merged into super-contigs and
7 oriented to super-scaffolds with large mate-paired libraries (2-40 kb). Gap filling was made
8 with reads from short insert-sized libraries (170-500 bp) (**Supplemental Tables S3-S4**). In
9 total, we sequenced 563-fold coverage bases of the estimated 691 Mb genome size. The final
10 assembly was 679 Mb, with a contig N50 of 63.11 kb and a scaffold N50 of 1.03 Mb (**Table**
11 **1**). The 672 longest scaffolds (11.2% of all scaffolds) covered more than 90% of the assembly
12 (**Supplemental Table S5**). To assess the completeness of the *L. crocea* assembly, 52-fold
13 coverage paired-end high-quality reads were aligned against the assembly (**Supplemental**
14 **Fig. S3**). More than 95.63% of the generated reads could be mapped to the assembly.
15 Furthermore, the integrity of the assembly was validated by the successful mapping of
16 98.80% of the transcripts from the mixed-tissue transcriptomes (**Supplemental Table S6**).
17 These results indicate that the genome assembly of *L. crocea* has high coverage and is of high
18 quality (**Supplemental Table S7**).

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

1 We identified 25,401 protein-coding genes based on *ab initio* gene prediction and
2 evidence-based searches from the reference proteomes of six other teleost fish and
3 humans (**Supplemental Fig. S4; Table S11**), in which 24,941 genes (98.20% of the whole
4 gene set) were supported by homology or RNAseq evidence (**Supplemental Fig. S5**). Over
5 97.35% of the inferred proteins matched entries in the InterPro, SWISS-PROT, KEGG or
6 TrEMBL database (**Supplemental Table S12**).

7 **Phylogenetic relationships and genomic comparison**

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

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

1 the phylogenetic analysis (**Supplemental Tables S14-S15**). Based on the ratios of the number
2 of nonsynonymous substitutions per nonsynonymous site (K_a) to the number of synonymous
3 substitutions per synonymous site (K_s ; K_a/K_s ratios) in a branch-site model of PAML (Yang
4 1997), 92 genes were found to be positively selected in *L. crocea* compared with their
5 orthologues in the other six teleost species ($P < 0.001$, **Supplemental Table S16**).

6 **Unique genetic features of the *L. crocea*.**

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

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

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

4 *L. crocea* is named for its ability to generate strong repetitive drumming sounds, especially
5 during reproduction (Su 2004). For good communication, fish have developed high
6 sensitivities to environmental sound. Three important auditory genes, otoferlin (*OTOF*),
7 *claudinj*, and otolin 1 (*OTOL1*), were significantly expanded in the *L. crocea* genome ($P <$
8 0.01 , **Supplemental Table S19**). These expansions may contribute to the detection of sound
9 signaling during communication, and thus to reproduction and survival (Eisen and Ryugo
10 2007).

11 Selenium is highly enriched in *L. crocea* (Su 2004), and it is mainly present as
12 selenoproteins. We used the SelGenAmic-based selenoprotein prediction method (Jiang et al.
13 2010) to analyse the *L. crocea* genome and identified 40 selenoprotein genes, which is the
14 highest number among all sequenced vertebrates (**Supplemental Table S20**). Interestingly,
15 five copies of *MsrB1*, which encodes methionine sulfoxide reductase, were found in *L.*
16 *crocea* (*MsrB1a*, *MsrB1b*, *MsrB1c*, *MsrB1d*, and *MsrB1e*), whereas only two copies (*MsrB1a*
17 and *MsrB1b*) were found in other fish, thus suggesting its broader specificity to reduce all
18 possible substrates (Vandermarliere et al. 2014).

19 **Characterization of the *L. crocea* immune system**

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

1 whereas its adaptive immune system may possess unique characteristics. The CD8⁺ T and
2 CD4⁺ T-helper type 1 (Th1) -type immune systems are well conserved in *L. crocea*, and
3 almost all CD8⁺ T and CD4⁺ Th1 cell-related genes were found (**Fig. 2A**). Moreover, the
4 genes related to Th17 cell- and $\gamma\delta$ -T cell-mediated mucosal immune responses were
5 conserved in *L. crocea*. These observations suggest that *L. crocea* may exhibit powerful
6 cellular and mucosal immunity. However, the CD4⁺ Th2-type immunity seemed to be weak in
7 *L. crocea*, as suggested by the absence of many CD4⁺ Th2-related genes and humoral
8 immune effectors (**Fig. 2A**). We detected gene expansions in several of these
9 immune-relevant genes, including those encoding lectin receptors (*CLECI7A*), a classical
10 complement component (*Clq*), apoptosis regulator (*BAX*), and immunoglobulins (*IgHV*) ($P <$
11 0.01 , **Supplemental Table S22**). Expansions were also observed in the genes encoding four
12 key proteins for mammalian antiviral immunity: tripartite motif containing 25 (*TRIM25*),
13 cyclic GMP-AMP synthase (*cGAS*), *DDX41*, and NOD-like receptor family CARD domain
14 containing 3 (*NLRC3*) (**Fig. 2B**). However, retinoic acid-inducible gene-1 (*RIG-I*), which
15 initiates antiviral signaling pathway in mammals, was not found in the *L. crocea* genome and
16 transcriptome (Mu et al. 2010; Mu et al. 2014). The teleost *RIG-I* has been identified only in
17 limited fish species, such as cyprinids and salmonids, and its absence suggests that it may
18 have been lost from particular fish genomes (Hansen et al. 2011). Furthermore, laboratory of
19 genetics and physiology 2 (*LGP2*) can serve as a suppressor to block *RIG-I*- and melanoma
20 differentiation-associated protein 5 (*MDA5*)-elicited signaling in mammals, but *LGP2* in fish
21 is able to bind to poly(I:C) to trigger interferon production (Chang et al. 2011), thereby acting
22 as a substitute for *RIG-I* (**Fig. 2B**). The expanded *TRIM25* (54 copies, **Supplemental Fig. S8**)

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

10 **Stress response under hypoxia**

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

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

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

1 previous reports in mammals (Mastorakos et al. 1993; Kitamuro et al. 2000; Earley et al.
2 2002; Takahashi et al. 2003).

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

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

1 **(Supplemental Table S25)**. This suggests that the HPT axis may inhibit protein synthesis
2 under hypoxia by decreasing the production of thyroid hormones (**Fig. 3**), which is beneficial
3 for saving energy during hypoxia stress. Thyroid hormones can also accelerate the oxidative
4 metabolism of glucose and inhibit the glycolytic anaerobic pathway (Sabell et al. 1985). Our
5 transcriptome analyses show that genes involved in the tricarboxylic acid (TCA) cycle
6 (pyruvate dehydrogenase complex [*PDC-E1*], succinyl-CoA synthetase [*SCS*], and fumarate
7 hydratase [*FH*]) were down-regulated 12 h later under hypoxia, whereas glycolysis-related
8 genes, such as pyruvate kinase (*PKM*), glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*),
9 *GPI*, and aldolase A (*ALDOA*), were greatly increased at 1 h (280-, 130-, 73- and 12-fold,
10 respectively) (**Supplemental Table S24**). The down-regulation of HPT axis-thyroid
11 hormones may inhibit the TCA cycle and accelerate the anaerobic glycolytic pathway in the
12 brain during hypoxia exposure (**Fig. 3**). The repression of the TCA cycle and the strong
13 induction of the anaerobic glycolytic pathway resulted in a physiological shift from aerobic to
14 anaerobic metabolism, where fish utilise O₂-independent mechanisms to produce adenosine
15 triphosphate (ATP). However, the mRNA levels of hypoxia-inducible factor (HIF)-1 α , which
16 are significantly up-regulated under hypoxia in mammals (Dayan et al. 2006; Benita et al.
17 2009), were not significantly changed in the *L. crocea* brain (**Supplemental Table S24**). It is
18 possible that the HIF-1 α -mediated mechanism may not be essential for the hypoxia response
19 in the *L. crocea* brain during the early period of hypoxia. These results suggest that the HPT
20 axis-mediated effects may play major roles in response to hypoxia by reorganizing energy
21 consumption and energy generation.

22

1 **Mucus components and function**

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

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

1 (GO:0016491, $P=1.58 \times 10^{-35}$, 223 proteins), peroxidase activity (GO:0004601, $P=0.0075$,
2 nine proteins), oxygen binding (GO:0019825, $P=0.0011$, eight proteins), and ion binding
3 (GO:0043167, $P=2.21 \times 10^{-6}$, 347 proteins) (**Fig. 4A; Supplemental Fig. S14**). Two hundred
4 and thirty-two antioxidant proteins that were related to oxidoreductase activity and
5 peroxidase activity were highly enriched in the *L. crocea* mucus, and they included
6 peroxiredoxins, glutathione peroxidase, and thioredoxin (**Supplemental Table S28**). These
7 proteins intercept and degrade environmental peroxy and hydroxyl radicals from aqueous
8 environments (Cross et al. 1984). Therefore, the presence of high-abundance antioxidant
9 proteins in the skin mucus may have the potential to protect fish against air exposure-induced
10 oxidative damage (**Fig. 4B**). Eight proteins related to oxygen transport, including hemoglobin
11 subunits $\alpha 1$, αA , αD , β , and $\beta 1$, and cytoglobin-1, were identified in the *L. crocea* skin mucus
12 (**Supplemental Table S29**). The abundant expression of hemoglobin may contribute to the
13 binding and holding of oxygen for respiration. Various immune molecules that provide
14 immediate protection to fish from potential pathogens, such as lectins, lysozymes, C-reactive
15 proteins, complement components, immunoglobulins, and chemokines, were also found in
16 the *L. crocea* skin mucus (**Supplemental Table S30**). To date, the mechanisms of osmotic
17 and ionic regulation of the skin mucus have not been confirmed (Shephard 1994). In this
18 study, a large number of ion-binding proteins were identified in the *L. crocea* mucus
19 (**Supplemental Table S31**). These proteins and the layer of mucus may have a role in
20 limiting the diffusion of ions on the surface of the fish (**Fig. 4B**). However, a substantial
21 proportion of the proteins, which are highly present in the skin mucus of fish under air
22 exposure, play an unknown role in the mucus response.

1 **Discussion**

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

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

1 Inhibition of protein synthesis principally contributes to the reduction in cellular energy
2 consumption during hypoxia (Gracey et al. 2001; Richards 2011). Activation of anaerobic
3 metabolism facilitates O₂-independent ATP production under hypoxia, albeit with low ATP
4 yield (Richards 2011). Therefore the reduction in ATP consumption through the HPT
5 axis-mediated inhibition of protein synthesis matched the lower ATP yield by the HPT
6 axis-activated anaerobic metabolism, which may aid to maintain cellular energy balance
7 under hypoxia, thus extending fish survival. Hence, our results reveal new aspects of
8 neuro-endocrine-immune/metabolism regulatory networks that may help the fish to avoid
9 cerebral inflammatory injury and maintain energy balance under hypoxia stress. These
10 discoveries will help to improve current understanding of
11 neuro-endocrine-immune/metabolism regulatory networks and protective mechanisms against
12 hypoxia-induced cerebral injury in vertebrates, providing clues for research on the
13 pathogenesis and treatment of hypoxia-induced cerebral diseases.

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

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

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

14

15 **Materials and Methods**

16 **Genome assembly annotation**

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

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

10 **Genome annotation**

11 For the annotation of repetitive elements, we used a combination of homology-based and *ab*
12 *initio* predictions. RepeatMasker (Smit 1996-2010) and Protein-based RepeatMasking (Smit
13 1996-2010) were used to search Repbase, which contains a vast amount of known
14 transcriptional elements at the DNA and protein levels. During the process of *ab initio*
15 prediction, RepeatScout (Price et al. 2005) was used to build the *ab initio* repeat library based
16 on k-mer, using the fit-preferred alignment score on the *L. crocea* genome. Contamination
17 and multi-copy genes in the library were filtered out before the RepeatScout library was used
18 to find homologs in the genome and to categorise the found repeats by RepeatMasker (Smit
19 1996-2010).

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

22 **Homology-based prediction**

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

7 ***Ab initio* prediction**

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

11 **Transcriptome-based prediction**

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

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

21 **Evolutionary and Comparative Analyses**

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

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

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

10 Gene family expansion and contraction analyses were performed by cafe (De Bie et al.
11 2006). For optical, olfactory receptor, and auditory system-related genes, we downloaded the
12 genes from Swissprot or Genebank and predicted their candidates using BLAST and
13 Genewise to determinate copy numbers. Pseudogenes produced by frame shift were removed.
14 Phylogenetic analysis of the expanded gene families was based on maximum likelihood
15 methods by PAML 3.0 (Yang 1997), and the phylogenetic tree was represented by EvolView
16 (Zhang et al. 2012b).

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

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

3 **Transcriptome under hypoxia**

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

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

1 differential expression comparisons.

2 **LC–MS/MS analyses and mucus protein identification**

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

11 Proteins were extracted from a pool of skin mucus of six fish by the trichloroacetic
12 acid-acetone precipitation method and digested by the trypsin gold (Promega, USA). The
13 peptides were then separated by the strong cation exchange chromatography using a
14 Shimadzu LC-20AB HPLC Pump system (Kyoto, Japan). Data acquisition was performed
15 with a Triple TOF 5600 System (AB SCIEX, Concord, ON) fitted with a Nanospray III
16 source (AB SCIEX, Concord, ON). All spectra were mapped by MASCOT server version
17 2.3.02 against the database of the *L. crocea* genome with the parameters as follows: peptide
18 mass tolerance 0.05 Da; fragment mass tolerance 0.1 Da; fixed modifications
19 “Carbamidomethyl (C)”; and variable modifications “Gln->pyro-Glu (N-term Q), Oxidation
20 (M), Deamidated (NQ)”. For further analyses of the function of the mucus proteome, we
21 selected proteins with more than two unique peptides.

22

1 **Data access**

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

9

10 **Acknowledgements**

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

19

20

21

22

1 **Figure legends**

2 **Figure 1. Phylogenetic tree of and orthologous genes in *L. crocea* and other vertebrates.**

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

15 **Figure 2. Characterisation of the T-cell lineages in *L. crocea* adaptive immunity and the** 16 **expanded genes in antiviral immunity.**

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

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

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

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

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

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

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

11

12 **References**

- 13 Benita Y, Kikuchi H, Smith AD, Zhang MQ, Chung DC, Xavier RJ. 2009. An integrative
14 genomics approach identifies Hypoxia Inducible Factor-1 (HIF-1)-target genes that
15 form the core response to hypoxia. *Nucleic Acids Res* **37**(14): 4587-4602.
- 16 Birney E, Clamp M, Durbin R. 2004. GeneWise and Genomewise. *Genome research* **14**(5):
17 988-995.
- 18 Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W. 2011. Scaffolding pre-assembled
19 contigs using SSPACE. *Bioinformatics* **27**(4): 578-579.
- 20 Bona E, Andersson AL, Blomgren K, Gilland E, Puka-Sundvall M, Gustafson K, Hagberg H.
21 1999. Chemokine and inflammatory cell response to hypoxia-ischemia in immature
22 rats. *Pediatr Res* **45**(4 Pt 1): 500-509.
- 23 Burge C, Karlin S. 1997. Prediction of complete gene structures in human genomic DNA.
24 *Journal of molecular biology* **268**(1): 78-94.
- 25 Castanier C, Zemirli N, Portier A, Garcin D, Bidere N, Vazquez A, Arnoult D. 2012. MAVS
26 ubiquitination by the E3 ligase TRIM25 and degradation by the proteasome is
27 involved in type I interferon production after activation of the antiviral RIG-I-like
28 receptors. *BMC Biol* **10**: 44.
- 29 Castresana J. 2000. Selection of conserved blocks from multiple alignments for their use in
30 phylogenetic analysis. *Molecular biology and evolution* **17**(4): 540-552.
- 31 Chang M, Collet B, Nie P, Lester K, Campbell S, Secombes CJ, Zou J. 2011. Expression and
32 functional characterization of the RIG-I-like receptors MDA5 and LGP2 in Rainbow

- 1 trout (*Oncorhynchus mykiss*). *J Virol* **85**(16): 8403-8412.
- 2 Chen S, Zhang G, Shao C, Huang Q, Liu G, Zhang P, Song W, An N, Chalopin D, Volff JN et
3 al. 2014. Whole-genome sequence of a flatfish provides insights into ZW sex
4 chromosome evolution and adaptation to a benthic lifestyle. *Nat Genet* **46**(3):
5 253-260.
- 6 Cossins AR, Crawford DL. 2005. Fish as models for environmental genomics. *Nat Rev Genet*
7 **6**(4): 324-333.
- 8 Cross CE, Halliwell B, Allen A. 1984. Antioxidant protection: a function of tracheobronchial
9 and gastrointestinal mucus. *Lancet* **1**(8390): 1328-1330.
- 10 Dayan F, Roux D, Brahimi-Horn MC, Pouyssegur J, Mazure NM. 2006. The oxygen sensor
11 factor-inhibiting hypoxia-inducible factor-1 controls expression of distinct genes
12 through the bifunctional transcriptional character of hypoxia-inducible factor-1alpha.
13 *Cancer Res* **66**(7): 3688-3698.
- 14 De Bie T, Cristianini N, Demuth JP, Hahn MW. 2006. CAFE: a computational tool for the
15 study of gene family evolution. *Bioinformatics* **22**(10): 1269-1271.
- 16 Earley S, Nelin LD, Chicoine LG, Walker BR. 2002. Hypoxia-induced pulmonary
17 endothelin-1 expression is unaltered by nitric oxide. *J Appl Physiol (1985)* **92**(3):
18 1152-1158.
- 19 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
20 throughput. *Nucleic acids research* **32**(5): 1792-1797.
- 21 Eisen MD, Ryugo DK. 2007. Hearing molecules: contributions from genetic deafness. *Cell*
22 *Mol Life Sci* **64**(5): 566-580.
- 23 Gao D, Wu J, Wu YT, Du F, Aroh C, Yan N, Sun L, Chen ZJ. 2013. Cyclic GMP-AMP
24 synthase is an innate immune sensor of HIV and other retroviruses. *Science*
25 **341**(6148): 903-906.
- 26 Gracey AY, Troll JV, Somero GN. 2001. Hypoxia-induced gene expression profiling in the
27 euryoxic fish *Gillichthys mirabilis*. *Proc Natl Acad Sci U S A* **98**(4): 1993-1998.
- 28 Gu X, Xu Z. 2011. Effect of hypoxia on the blood of large yellow croaker (*Pseudosciaena*
29 *crocea*). *Chinese Journal of Oceanology and Limnology* **29**(3): 524.
- 30 Guzman-Aranguiz A, Mantelli F, Argueso P. 2009. Mucin-type O-glycans in tears of normal
31 subjects and patients with non-Sjogren's dry eye. *Invest Ophthalmol Vis Sci* **50**(10):
32 4581-4587.
- 33 Hansen JD, Vojtech LN, Laing KJ. 2011. Sensing disease and danger: a survey of vertebrate
34 PRRs and their origins. *Dev Comp Immunol* **35**(9): 886-897.
- 35 Hayashi R, Wada H, Ito K, Adcock IM. 2004. Effects of glucocorticoids on gene transcription.
36 *Eur J Pharmacol* **500**(1-3): 51-62.
- 37 Herman JP, Cullinan WE. 1997. Neurocircuitry of stress: central control of the
38 hypothalamo-pituitary-adrenocortical axis. *Trends Neurosci* **20**(2): 78-84.
- 39 Hou TD, Du JZ. 2005. Norepinephrine attenuates hypoxia-inhibited thyrotropin-releasing
40 hormone release in median eminence and paraventricular nucleus of rat hypothalamus.
41 *Neuro Endocrinol Lett* **26**(1): 43-49.
- 42 Jiang L, Liu Q, Ni J. 2010. In silico identification of the sea squirt selenoproteome. *BMC*
43 *Genomics* **11**: 289.
- 44 Kaur C, Ling EA. 2008. Blood brain barrier in hypoxic-ischemic conditions. *Curr Neurovasc*

- 1 *Res* **5**(1): 71-81.
- 2 Kenessey A, Ojamaa K. 2006. Thyroid hormone stimulates protein synthesis in the
3 cardiomyocyte by activating the Akt-mTOR and p70S6K pathways. *J Biol Chem*
4 **281**(30): 20666-20672.
- 5 Kitamuro T, Takahashi K, Nakayama M, Murakami O, Hida W, Shirato K, Shibahara S. 2000.
6 Induction of adrenomedullin during hypoxia in cultured human glioblastoma cells. *J*
7 *Neurochem* **75**(5): 1826-1833.
- 8 Korf I. 2004. Gene finding in novel genomes. *BMC bioinformatics* **5**: 59.
- 9 Lee DC, Hassan SS, Romero R, Tarca AL, Bhatti G, Gervasi MT, Caruso JA, Stemmer PM,
10 Kim CJ, Hansen LK et al. 2011. Protein profiling underscores immunological
11 functions of uterine cervical mucus plug in human pregnancy. *J Proteomics* **74**(6):
12 817-828.
- 13 Lemos Vde A, dos Santos RV, Lira FS, Rodrigues B, Tufik S, de Mello MT. 2013. Can high
14 altitude influence cytokines and sleep? *Mediators Inflamm* **2013**: 279365.
- 15 Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. 2009. SOAP2: an improved
16 ultrafast tool for short read alignment. *Bioinformatics* **25**(15): 1966-1967.
- 17 Li W, Sorensen PW, Gallaher DD. 1995. The olfactory system of migratory adult sea lamprey
18 (*Petromyzon marinus*) is specifically and acutely sensitive to unique bile acids
19 released by conspecific larvae. *J Gen Physiol* **105**(5): 569-587.
- 20 Liu XD, Zhao GT, Cai MY, Wang ZY. 2013. Estimated genetic parameters for growth-related
21 traits in large yellow croaker *Larimichthys crocea* using microsatellites to assign
22 parentage. *Journal of fish biology* **82**(1): 34-41.
- 23 Liu YY, Cao MJ, Zhang ML, Hu JW, Zhang YX, Zhang LJ, Liu GM. 2014. Purification,
24 characterization and immunoreactivity of beta'-component, a major allergen from the
25 roe of large yellow croaker (*Pseudosciaena crocea*). *Food and chemical toxicology :
26 an international journal published for the British Industrial Biological Research
27 Association* **72C**: 111-121.
- 28 Loytynoja A, Goldman N. 2010. webPRANK: a phylogeny-aware multiple sequence aligner
29 with interactive alignment browser. *BMC bioinformatics* **11**: 579.
- 30 Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y et al. 2012.
31 SOAPdenovo2: an empirically improved memory-efficient short-read de novo
32 assembler. *GigaScience* **1**(1): 18.
- 33 Mastorakos G, Chrousos GP, Weber JS. 1993. Recombinant interleukin-6 activates the
34 hypothalamic-pituitary-adrenal axis in humans. *J Clin Endocrinol Metab* **77**(6):
35 1690-1694.
- 36 Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying
37 mammalian transcriptomes by RNA-Seq. *Nature methods* **5**(7): 621-628.
- 38 Mu Y, Ding F, Cui P, Ao J, Hu S, Chen X. 2010. Transcriptome and expression profiling
39 analysis revealed changes of multiple signaling pathways involved in immunity in the
40 large yellow croaker during *Aeromonas hydrophila* infection. *BMC Genomics* **11**: 506.
- 41 Mu Y, Li M, Ding F, Ding Y, Ao J, Hu S, Chen X. 2014. De novo characterization of the
42 spleen transcriptome of the large yellow croaker (*Pseudosciaena crocea*) and analysis
43 of the immune relevant genes and pathways involved in the antiviral response. *PLoS*
44 *One* **9**(5): e97471.

- 1 Nadeau S, Rivest S. 2003. Glucocorticoids play a fundamental role in protecting the brain
2 during innate immune response. *J Neurosci* **23**(13): 5536-5544.
- 3 Niimura Y. 2009. On the origin and evolution of vertebrate olfactory receptor genes:
4 comparative genome analysis among 23 chordate species. *Genome Biol Evol* **1**: 34-44.
- 5 Ning Y, Liu X, Wang ZY, Gou W, Li Y, Xie F. 2007. A genetic map of large yellow croaker
6 *Pseudosciaena crocea*. *Aquaculture* **264** (1): 16-26.
- 7 Pluta K, McGettigan PA, Reid CJ, Browne JA, Irwin JA, Tharmalingam T, Corfield A, Baird
8 A, Loftus BJ, Evans AC et al. 2012. Molecular aspects of mucin biosynthesis and
9 mucus formation in the bovine cervix during the peri-estrous period. *Physiol Genomics*
10 **44**(24): 1165-1178.
- 11 Price AL, Jones NC, Pevzner PA. 2005. De novo identification of repeat families in large
12 genomes. *Bioinformatics* **21 Suppl 1**: i351-358.
- 13 Richards JG. 2011. Physiological, behavioral and biochemical adaptations of intertidal fishes
14 to hypoxia. *J Exp Biol* **214**(Pt 2): 191-199.
- 15 Rodriguez-Pineiro AM, Bergstrom JH, Ermund A, Gustafsson JK, Schutte A, Johansson ME,
16 Hansson GC. 2013. Studies of mucus in mouse stomach, small intestine, and colon. II.
17 Gastrointestinal mucus proteome reveals Muc2 and Muc5ac accompanied by a set of
18 core proteins. *Am J Physiol Gastrointest Liver Physiol* **305**(5): G348-356.
- 19 Ruan J, Li H, Chen Z, Coghlan A, Coin LJ, Guo Y, Heriche JK, Hu Y, Kristiansen K, Li R et
20 al. 2008. TreeFam: 2008 Update. *Nucleic acids research* **36**(Database issue):
21 D735-740.
- 22 Sabell I, Morata P, Quesada J, Morell M. 1985. Effect of thyroid hormones on the glycolytic
23 enzyme activity in brain areas of the rat. *Enzyme* **34**(1): 27-32.
- 24 Schartl M, Walter RB, Shen Y, Garcia T, Catchen J, Amores A, Braasch I, Chalopin D, Volff
25 JN, Lesch KP et al. 2013. The genome of the platyfish, *Xiphophorus maculatus*,
26 provides insights into evolutionary adaptation and several complex traits. *Nat Genet*
27 **45**(5): 567-572.
- 28 Shephard KL. 1994. Functions for fish mucus. *Reviews in Fish Biology and Fisheries* **4**(4):
29 401-429.
- 30 Smit A, Hubley R & Green, P. RepeatMasker Open-3.0. 1996-2010
31 <http://www.repeatmasker.org>.
- 32 Sorrells SF, Sapolsky RM. 2007. An inflammatory review of glucocorticoid actions in the
33 CNS. *Brain Behav Immun* **21**(3): 259-272.
- 34 Stanke M, Morgenstern B. 2005. AUGUSTUS: a web server for gene prediction in
35 eukaryotes that allows user-defined constraints. *Nucleic acids research* **33**(Web Server
36 issue): W465-467.
- 37 Star B, Nederbragt AJ, Jentoft S, Grimholt U, Malmstrom M, Gregers TF, Rounge TB,
38 Paulsen J, Solbakken MH, Sharma A et al. 2011. The genome sequence of Atlantic
39 cod reveals a unique immune system. *Nature* **477**(7363): 207-210.
- 40 Su Y. 2004. Breeding and Farming of *Pseudosciaena Crocea*. 1st ed. China Ocean Press,
41 Beijing.
- 42 Subramanian S, Ross NW, MacKinnon SL. 2008. Comparison of antimicrobial activity in the
43 epidermal mucus extracts of fish. *Comp Biochem Physiol B Biochem Mol Biol* **150**(1):
44 85-92.

- 1 Takahashi K, Uono-Fujimori R, Totsune K, Murakami O, Shibahara S. 2003. Suppression of
2 cytokine-induced expression of adrenomedullin and endothelin-1 by dexamethasone
3 in T98G human glioblastoma cells. *Peptides* **24**(7): 1053-1062.
- 4 Taylor MM, Bagley SL, Samson WK. 2005. Intermedin/adrenomedullin-2 acts within central
5 nervous system to elevate blood pressure and inhibit food and water intake. *Am J*
6 *Physiol Regul Integr Comp Physiol* **288**(4): R919-927.
- 7 Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: discovering splice junctions with
8 RNA-Seq. *Bioinformatics* **25**(9): 1105-1111.
- 9 van der Meer DL, van den Thillart GE, Witte F, de Bakker MA, Besser J, Richardson MK,
10 Spaink HP, Leito JT, Bagowski CP. 2005. Gene expression profiling of the long-term
11 adaptive response to hypoxia in the gills of adult zebrafish. *Am J Physiol Regul Integr*
12 *Comp Physiol* **289**(5): R1512-1519.
- 13 Vandermarliere E, Ghesquiere B, Jonckheere V, Gevaert K, Martens L. 2014. Unraveling the
14 specificities of the different human methionine sulfoxide reductases. *Proteomics*, doi:
15 10.1002/pmic.201300357.
- 16 Yang N, Ray DW, Matthews LC. 2012. Current concepts in glucocorticoid resistance.
17 *Steroids* **77**(11): 1041-1049.
- 18 Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood.
19 *Computer applications in the biosciences : CABIOS* **13**(5): 555-556.
- 20 Ye H, Liu Y, Liu X, Wang X, Wang Z. 2014. Genetic Mapping and QTL Analysis of Growth
21 Traits in the Large Yellow Croaker *Larimichthys crocea*. *Marine biotechnology*, doi:
22 10.1007/s10126-014-9590-z.
- 23 Yen PM. 2001. Physiological and molecular basis of thyroid hormone action. *Physiol Rev*
24 **81**(3): 1097-1142.
- 25 Yu S, Mu Y, Ao J, Chen X. 2010. Peroxiredoxin IV regulates pro-inflammatory responses in
26 large yellow croaker (*Pseudosciaena crocea*) and protects against bacterial challenge.
27 *J Proteome Res* **9**(3): 1424-1436.
- 28 Zhang G, Fang X, Guo X, Li L, Luo R, Xu F, Yang P, Zhang L, Wang X, Qi H et al. 2012a.
29 The oyster genome reveals stress adaptation and complexity of shell formation.
30 *Nature* **490**(7418): 49-54.
- 31 Zhang H, Gao S, Lercher MJ, Hu S, Chen WH. 2012b. EvolView, an online tool for
32 visualizing, annotating and managing phylogenetic trees. *Nucleic Acids Res* **40**(Web
33 Server issue): W569-572.
- 34 Zhang L, Mo J, Swanson KV, Wen H, Petrucelli A, Gregory SM, Zhang Z, Schneider M,
35 Jiang Y, Fitzgerald KA et al. 2014. NLRC3, a Member of the NLR Family of Proteins,
36 Is a Negative Regulator of Innate Immune Signaling Induced by the DNA Sensor
37 STING. *Immunity* **40**(3): 329-341.
- 38 Zhang Z, Yuan B, Bao M, Lu N, Kim T, Liu YJ. 2011. The helicase DDX41 senses
39 intracellular DNA mediated by the adaptor STING in dendritic cells. *Nat Immunol*
40 **12**(10): 959-965.
- 41 Zhou Y, Yan X, Xu S, Zhu P, He X, Liu J. 2011. Family structure and phylogenetic analysis
42 of odorant receptor genes in the large yellow croaker (*Larimichthys crocea*). *BMC*
43 *Evol Biol* **11**: 237.
- 44

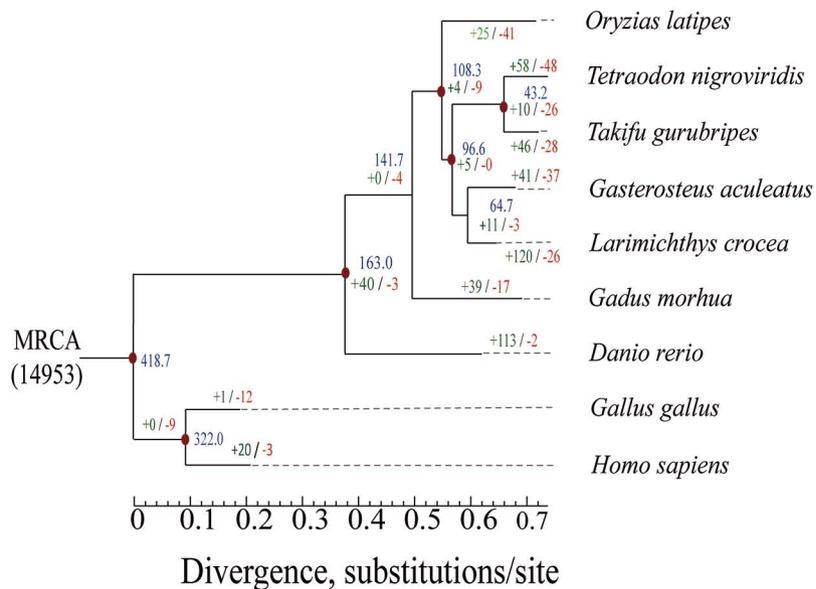
Table 1. Summary of the *Larimichthys crocea* genome

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

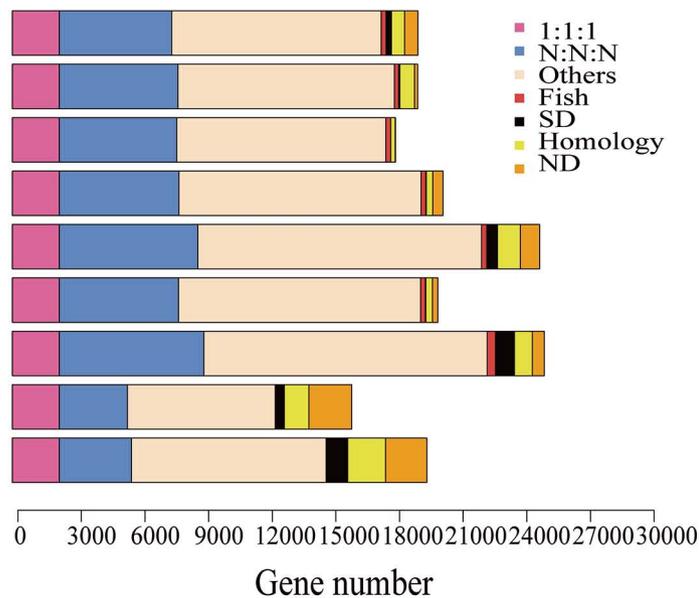
BAC = bacterial artificial chromosome; WGS = whole genome shotgun.

Figure 1

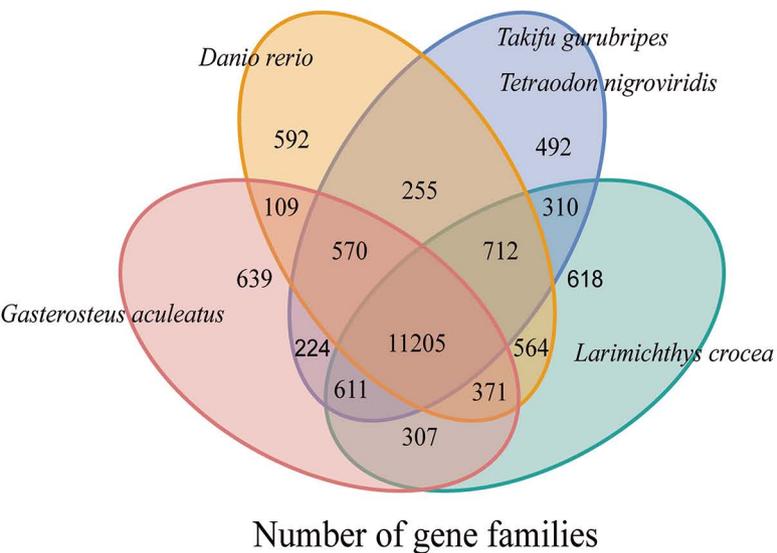
A



B



C



D

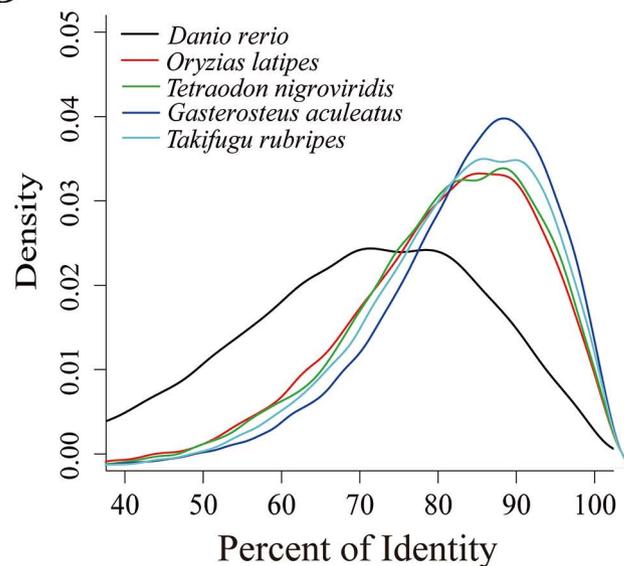
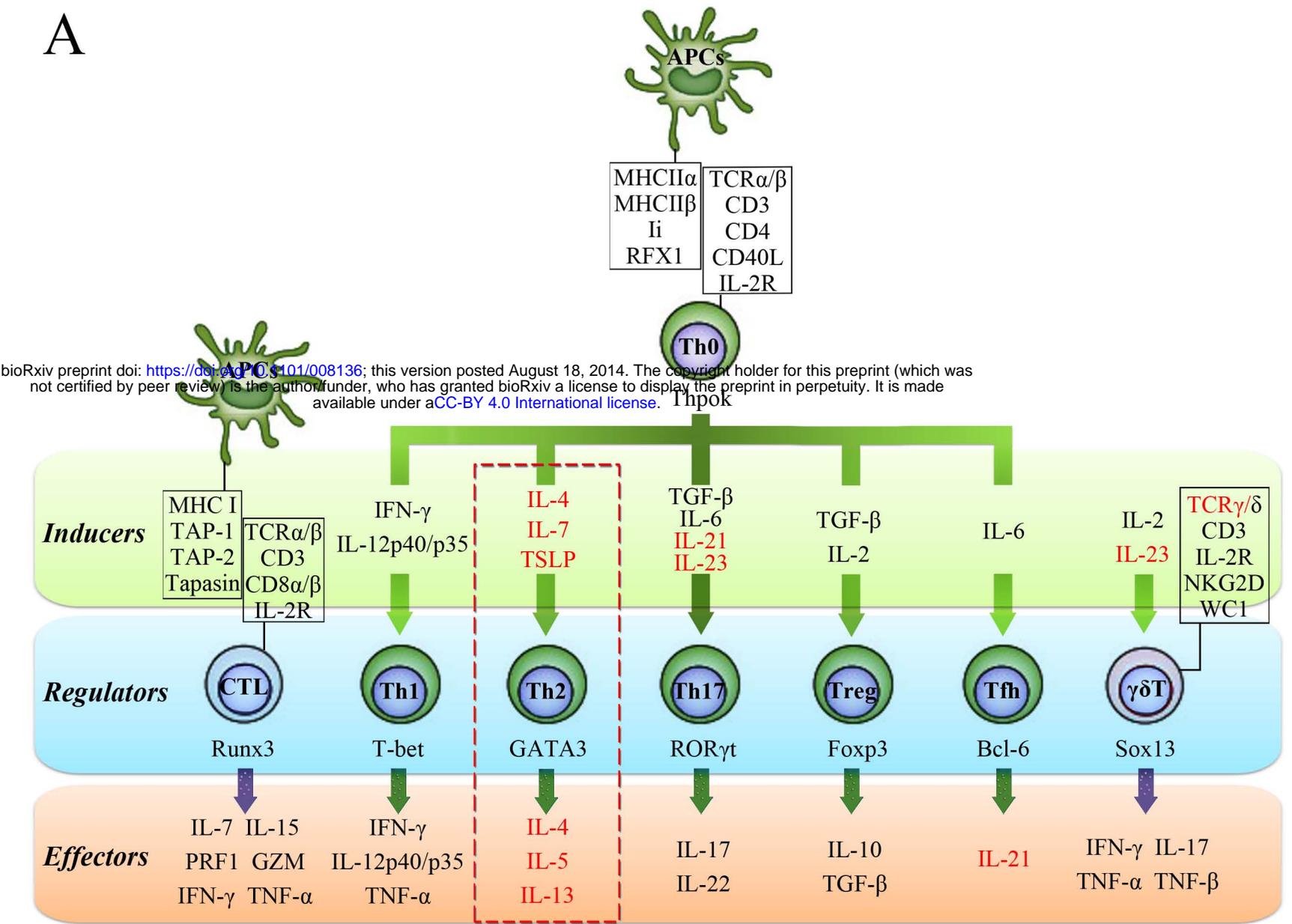


Figure 2

A



B

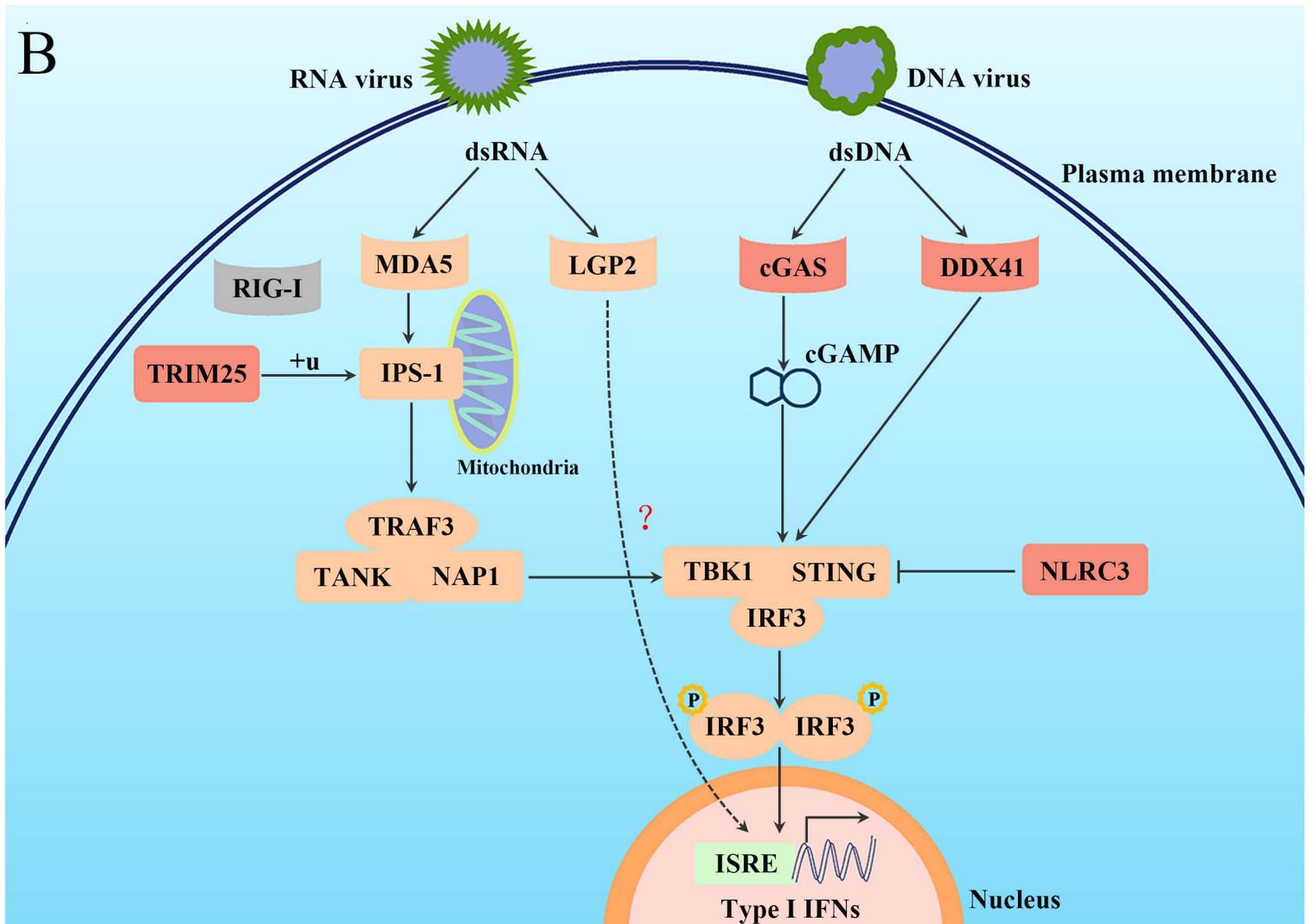


Figure 3

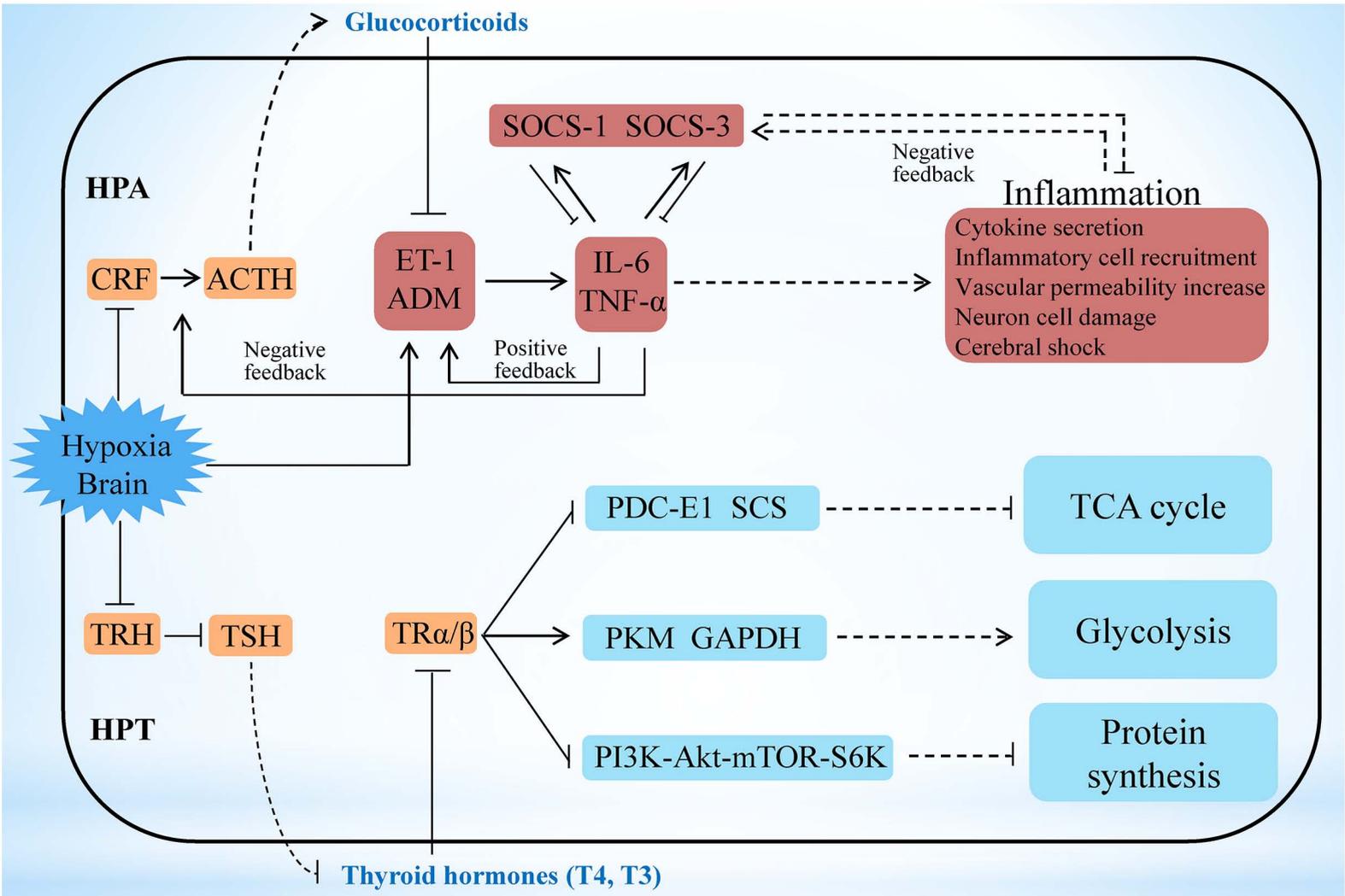
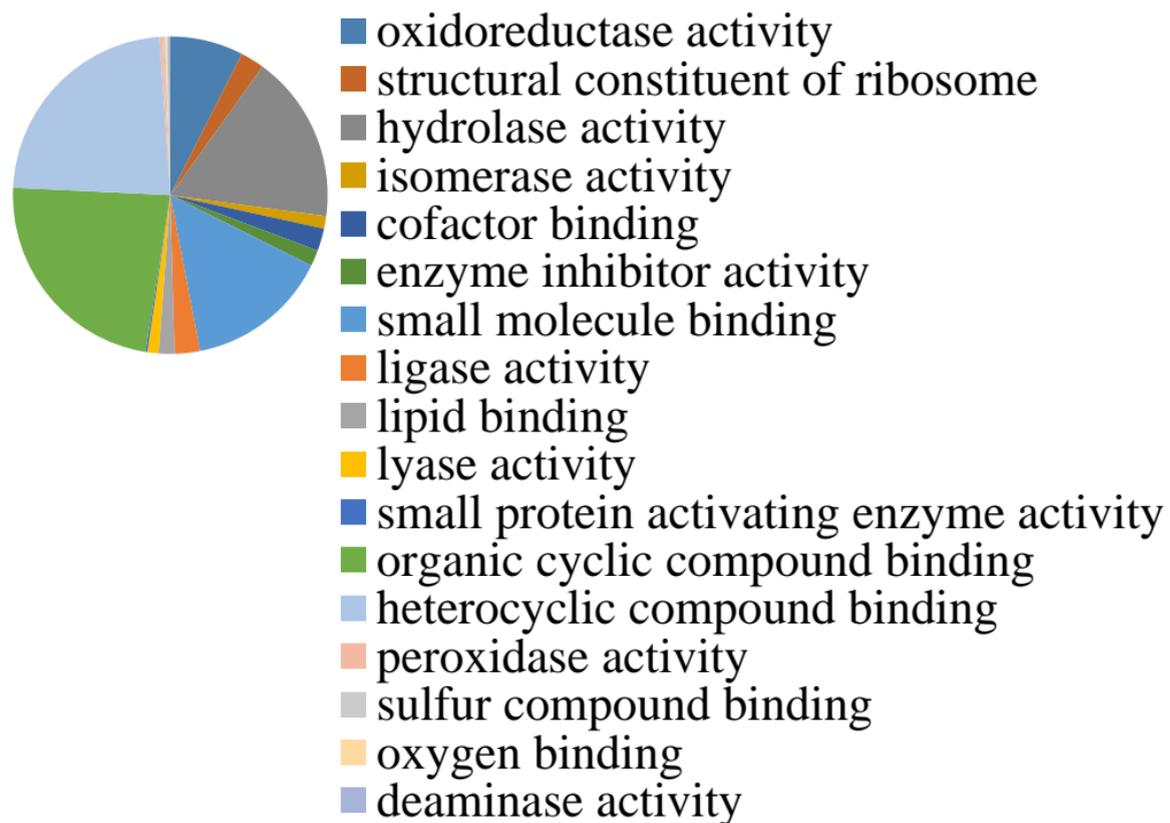


Figure 4

A



B

