ZR Zoological Research

Supplementary Materials

Author Correction: Opah (*Lampris megalopsis*) genome sheds light on the evolution of aquatic endothermy

Jing Bo[1,3], Wen-Qi Lv[2,3], Ning Sun[2,3], Cheng Wang[2,3], Kun Wang[4], Pan Liu[5], Chen-Guang Feng[2,4], Shun-Ping He[1,2,6,*], Lian-Dong Yang[2,*]

1.Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences, Sanya, Hainan 572000, China
2.State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, Hubei 430072, China
3.University of Chinese Academy of Sciences, Beijing 100049, China
4.School for Ecological and Environmental Sciences, Northwestern Polytechnical University, Xi'an, Shaanxi 710000, China
5.College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China
6.Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, Yunnan 650223, China
*Corresponding authors, E-mail: clad@ihb.ac.cn; yangld@ihb.ac.cn

Corrected version

Opah (*Lampris megalopsis*) genome sheds light on the evolution of aquatic endothermy

Jing Bo[1,3], Wen-Qi Lv[2,3], Ning Sun[2,3], Cheng Wang[2,3], Kun Wang[4], Pan Liu[5], Chen-Guang Feng[2,4], Shun-Ping He[1,2,6,*], Lian-Dong Yang[2,*]

1. Institute of Deep-Sea Science and Engineering, Chinese Academy of Sciences, Sanya, Hainan 572000, China
2. State Key Laboratory of Freshwater Ecology and Biotechnology, Institute of Hydrobiology, Chinese Academy of Sciences, Wuhan, Hubei 430072, China
3. University of Chinese Academy of Sciences, Beijing 100049, China
4. School for Ecological and Environmental Sciences, Northwestern Polytechnical University, Xi'an, Shaanxi 710000, China
5. College of Marine Sciences, Shanghai Ocean University, Shanghai 201306, China
6. Center for Excellence in Animal Evolution and Genetics, Chinese Academy of Sciences, Kunming, Yunnan 650223, China

Endothermy is the ability to generate and conserve metabolic heat to maintain body temperature above that of the surrounding environment. Endothermy enhances the physiological and ecological advantages of mammals, birds, and certain fish species. The opah, *Lampris megalopsis* (Lampridiformes), is the only known fish to exhibit whole-body endothermy. Currently, however, the underlying molecular mechanism for this remains unclear. Hence, the opah offers an excellent opportunity to study the evolutionary mechanism of whole-body endothermy in aquatic animals. In this study, we assembled the *L. megalopsis* genome (1.47 Gb in size) and performed comparative genomic analysis with ectothermic fish to reveal the genetic basis of endothermy. Based on analysis of positive selection, rapid evolution, and gene family expansion, we discovered several genes that likely contributed to thermogenesis and heat preservation. As the first reported *L. megalopsis* genome, our results not only clarify the possible molecular and genetic mechanisms involved in endothermic adaptation but also increase our understanding of endothermic fish biology.

Fish face a massive challenge in maintaining body temperature due to the high thermal conductivity of water, with fewer than 0.1% of fish species able to retain internally produced heat. Endothermic fish exhibit strong swimming efficiency, enhanced cold tolerance, high visual acuity for vertical migration, and competitive advantages in predator-prey interactions and niche expansion (Wegner et al., 2015). Recently, the opah (*Lampris guttatus*) was discovered to show whole-body endothermy, which is thought to be unique among fish. Globally distributed opahs were originally described as *Lampris guttatus* (Brünnich, 1788), but have since been divided into *Lampris incognitus*, *Lampris australensis*, *Lampris megalopsis*, *Lampris guttatus*, and *Lampris lauta* based on geographical distribution and morphology (Hyde et al., 2014; Underkoffler et al., 2018). Opahs have two thermogenic strategies, i.e., generating heat

via the metabolization of pectoral muscles and thermogenesis in specialized intracranial tissues. In addition, opahs have two heat conservation strategies, i.e., extensive retia mirabilia in intracranial and gill tissue for insulation and thick adipose tissue for heat conservation (Runcie et al., 2009; Wegner et al., 2015; Franck et al., 2019).

In the current study, an opah specimen was collected from the southern Indian Ocean onboard the Chinese tuna longline vessel "PING TAI RONG 71" and was transported to Zhoushan, China, on 18 May 2022, by "PING TAI RON LENG 2" for dissection and muscle extraction (Figure 1). Mitochondrial genes (cytochrome c oxidase I (*COI*) and cytochrome b (*cyt b*)) identified the specimen as a bigeye pacific opah (*L. megalopsis*) (Figures 2 and 3), which is known to be distributed in the southern Indian Ocean (Underkoffler et al., 2018). We assembled an *L. megalopsis* genome (1 473 Mb), which consisted of 2 234 contigs, with a contig N50 of 2 457 kb and N90 of 364 kb (Table 1). Compared to 38.8% for the previously released *L. guttatus* genome (ASM90030254v1), BUSCO assessment of single-copy orthologs showed that the completeness of our genome was 92.5%, indicating a high-quality assembly. A total of 26 999 genes were annotated in the opah genome. We constructed a phylogenetic tree with all single-copy genes (Figure 4A). The estimated divergence time of the opah was ~146 million years ago (Ma).

Table 1 Opah genome assembly

Term	Number
Total length	1 472 598 135 (bases)
Number of contigs	2 234
Number>=2 000 bp	2 234
N50	2 457 283
N90	364 684
GC rate	0.4525

The opah uses its dark red aerobic pectoral muscles for continuous swimming and metabolic heat production. On average, the muscle has a temperature that is 4.8±1.2 °C higher than that of the surrounding water (7.8–10.8 °C), and muscle content accounts for 16% of total weight, the highest among reported fish (Wegner et al., 2015). Here, we identified a rapidly evolving gene (*Smn1*), a positively selected gene (*Mlc3*), and an expanded gene family (*Myoz1*), which are involved in muscle development and differentiation (Khanna et al., 2003; Hayhurst et al., 2012).

Gene Ontology (GO) enrichment analysis identified three expanded gene families (*Myo*, *Calm*, *Act*) enriched in "myofibril assembly" (Figure 4B). Calcium ions are involved in the regulation of contractile protein activity during muscle contraction and relaxation (Cho et al., 2017). In our study, based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) calcium signaling pathway, two genes (*Phkg1*, *Htr2b*) were found to be rapidly evolving and three gene families (*Ltb4r1*, *Ighd*, *Htr7*) were expanded, which may be related to the tireless nature of opah pectoral muscle contraction.

Figure 1 Sample collection. Picture on left was taken onboard during opah collection. Right two pictures show Zhoushan dissection of opah muscle tissue.

Intracranial tissues of fish are specialized for non-shivering thermogenic activity, primarily mediated by *Sln* (Franck et al., 2019). Here, in the *Sln* pathway, the *Ryr* and *Serca* gene families were expanded in opah (Figure 4C). *Ryr1* is discretely expressed in specialized thermogenic tissues and promotes Ca^{2+} release from the sarcoplasmic reticulum (Maurya and Periasamy, 2015). *Serca* acts as a calcium pump and binds to *Sln*, causing the release of Ca^{2+} back into the cytoplasm (Franck et al., 2019). In opahs, *Ryr1* and *Serca* may lead to additional calcium supply for circulation pathways, further consumption of energy, and greater generation of heat.

In thermogenesis studies of energy supply, the *Hk* and *Mdh* gene families are expanded and participate in gluconeogenesis and the tricarboxylic acid cycle. Increased *Hk* activity in skeletal muscle is a coordinated process of long-term metabolic regulation, associated with increased aerobic carbohydrate catabolism and basal metabolic rate in muscle tissue (Kubišta et al., 1971). Research has shown that faster swimming speeds are associated with higher red muscle ratio and Mdh enzymatic activity (Pinte et al., 2021).

Opah fish use two heat preservation mechanisms, including counter-current retia mirabilia (Wegner et al., 2015). Apelin produced from arterial endothelial cells stimulates the expression of the apelin receptor (*Apj*) in venous endothelial cells to induce alignment of arteries and veins, which is involved in thermoregulation. In the current study, we identified *Apj* as a rapidly evolving gene in opah. As such, *Apj* may provide a new way in which to study the distribution of retia mirabilia in endotherms. However, the mechanism underlying the development of counter-current exchange retia mirabilia in specific tissues is not yet clear.

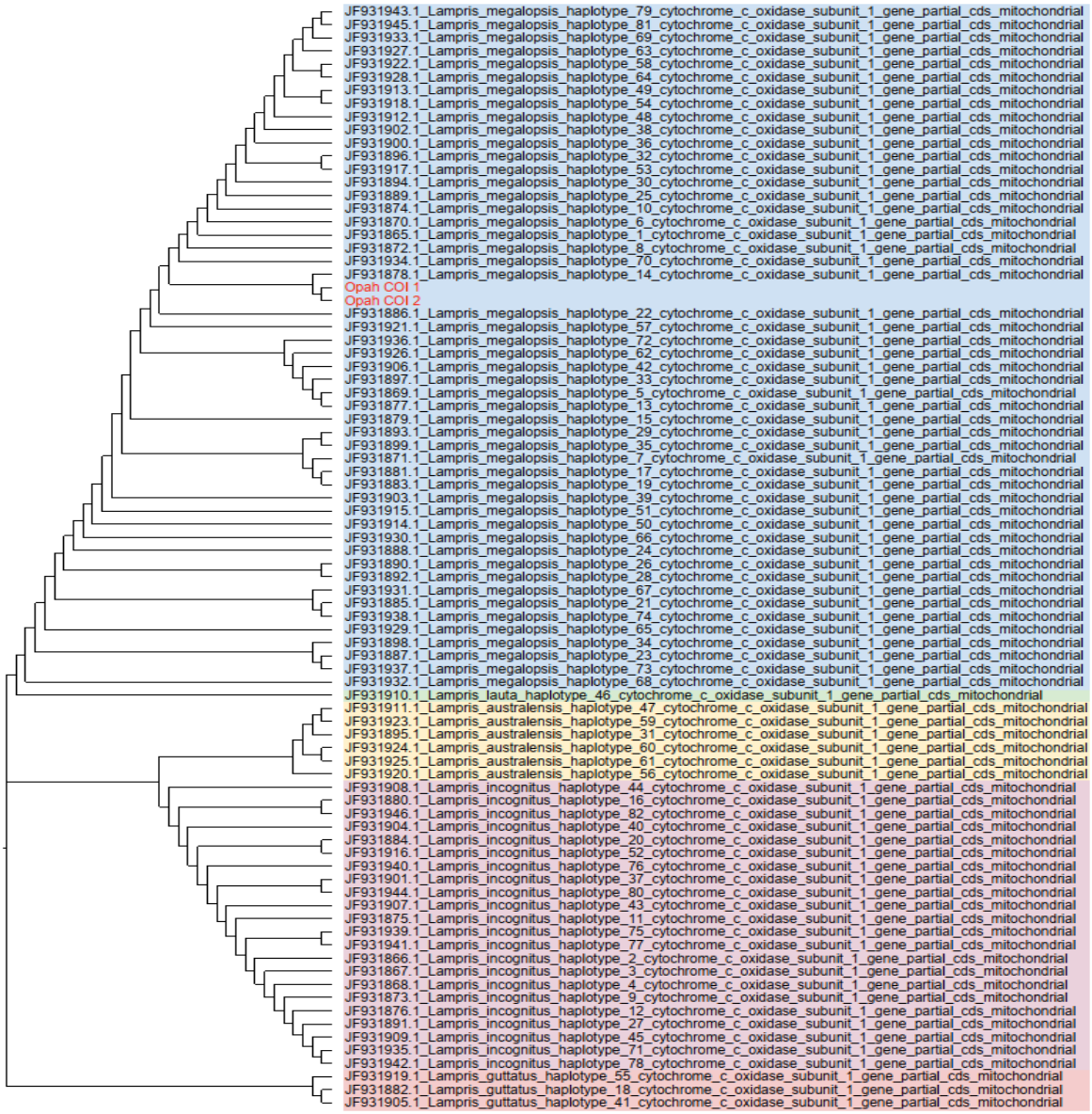

Figure 2 Previously published opah *COI* genes (GenBank accession numbers JF931865–JF931947) (Hyde et al., 2014) were aligned with opah counterparts from this study. Genes were aligned using MUSCLE and a maximum-likelihood tree was constructed using MEGA v7 (Kumar et al., 2016) with default parameters. The opah specimen from this study was clustered with *COI* of *L. megalopsis*.

Opahs also preserve heat via thick fatty tissue (Wegner et al., 2015), although little is known in relation to their adipose-related pathways. Regarding fatty acid metabolism, we found that *Acsl5*, *Mcat*, *Hadha*, and *Acat1* were rapidly evolving genes in opah. Given their participation in fatty acid metabolism, these three genes may help clarify the position of fatty acids in heat producing areas (Kao et al., 2006; Smith et al., 2012; Bowman et al., 2016; Chowdhury et al., 2016).

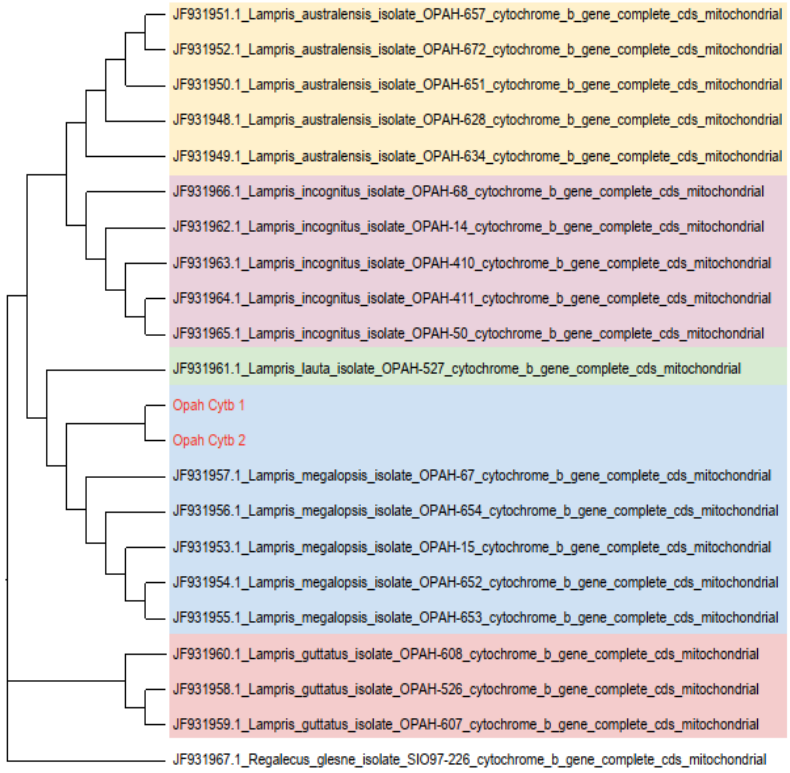

Figure 3 Previously published opah *cyt b* genes (GenBank accession numbers JF931948–JF931967) (Hyde et al., 2014) were aligned with opah counterparts from this study. Genes were aligned using MUSCLE and a maximum-likelihood tree was constructed using MEGA v7 with default parameters (Kumar et al., 2016). The opah specimen from this study was clustered with *cyt b* of *L. megalopsis*.

Fat in thermogenic regions and gills not only serves as physical insulation but may also influence the function of internal blood vessels by releasing adipokines. Adipokines released by perivascular adipose tissue affect blood vessel function via AMP-activated protein kinase (AMPK) (Wu et al., 2018). Based on KEGG pathway analysis, we identified an expanded gene family (*Igf1r*) enriched in the "AMPK signaling pathway".

We sequenced and assembled a complete opah genome, which should provide information for endothermic research. We conducted comparative genomic analysis of opah and ectothermic fish based on two significant aspects of endothermy: i.e., heat production and preservation. The genetic changes in opah are summarized in Figure 4D and Tables 2–4. Although the changes found in the opah genome require functional confirmation, this study should improve our understanding of endothermic adaptation in opahs and endothermic evolution in fish.

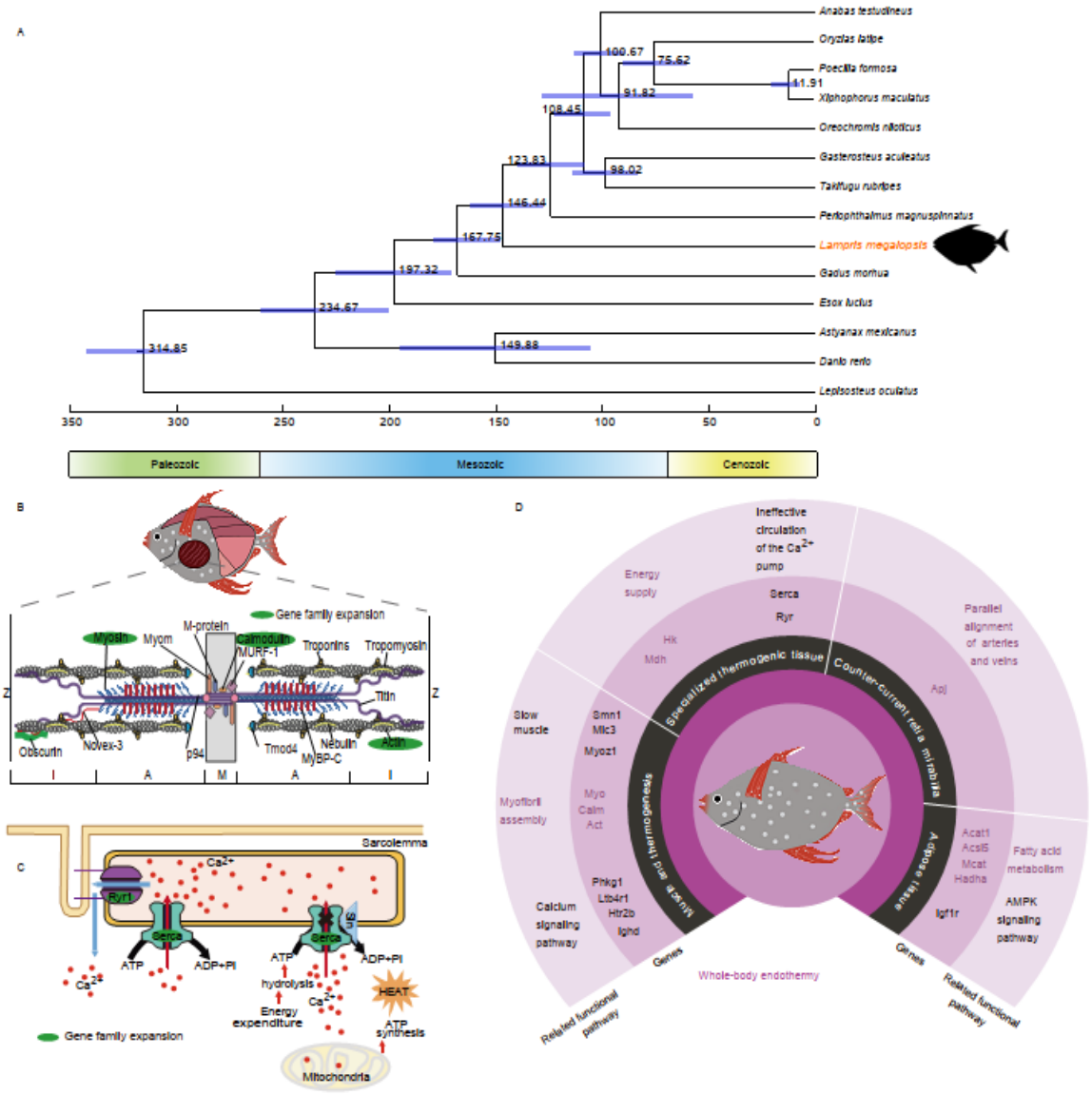

Figure 4 Genomic analyses of *L. megalopsis*. Green genes are under positive selection and blue genes exhibit rapid evolution. A: Phylogenetic relationships and divergence time of opah. Purple node bars indicate 95% high posterior density. B: Enrichment of positively selected genes in myofibril assembly. Green genes are under positive selection. C: *Sln*-mediated heat production pathways in specialized thermogenic tissue. D: Summary of genetic mechanisms of opah in whole-body endothermy, shown in inner black circle. Inner purple circle shows candidate genes based on positively selected gene sets, expanded gene families, and rapidly evolving gene sets. Outer gray violet circle shows related pathways.

Materials and Methods

Genome sequencing, assembly, and annotation

DNA was extracted from the opah muscle tissues for genome sequencing. Long reads were produced using the PacBio HiFi platform. The short reads were sequenced using the Illumina NovaSeq 6000 platform. A total of 108.96 Gb of short reads and 35.61 Gb of long reads were generated. The short reads were used to estimate genome

size with the *k*-mer method. The 17-mer depth frequency distribution was calculated using Jellyfish (Marçais and Kingsford, 2011). The predicted genome size was 826 Mb (**Figure 5**). The Hifiasm (Cheng et al., 2021) program was used to assemble the contigs. The short reads were mapped to the genome using BWA-MEM (Li, 2013) with default parameters. Two rounds of polishing were then performed using NextPolish v1.0 (Hu et al., 2020) based on aligned files. BUSCO (Simão et al., 2015) was used to assess genome quality and gene completeness with the library "actinopterygii_odb9". Gene annotation was conducted using a combination of *ab initio* and homology-based methods. The *ab initio* approach was conducted using Augustus v3.2.1 (Stanke et al., 2008), SNAP (Leskovec and Sosic, 2017), and GlimmerHMM (Majoros et al., 2004). For homology-based annotation, the protein sequences of *Danio rerio*, *Oryzias latipes*, and *Gadus morhua* were downloaded from Ensembl (http://www.ensembl.org/index.html) and aligned to the opah genome using tBLASTn (E-value≤1e-05) (Altschul et al., 1990). GeneWise v2.4.1 (Birney et al., 2004) was then used to identify accurate gene structures.

We extracted mitochondrial genes *COI* and *cyt b* to identify species. An EZNA® Tissue DNA Kit (OMEGA, Wuhan, China) was employed to extract DNA from opah muscle. The mitochondrial *COI* gene was amplified by polymerase chain reaction (PCR) using primers FishF1-COI 5' TCA ACC AAC CAC AAA GAC ATT GGC AC and FishR1-COI 5' TAG ACT TCT GGG TGG CCA AAG AAT CA. The mitochondrial *cyt b* gene was amplified by PCR using primers L14504-ND6 5' GCC AAW GCT GCW GAA TAM GCA AAG GTG and H15149-CYB 5' GCK CCT CAG AAG GAC ATT TGK CCT CA. Previously published opah *COI* and *cyt b* genes were downloaded from GenBank (accession JF931865-JF931967) (Hyde et al., 2014) and aligned with the counterpart genes of opah from this study. Each gene set was aligned using MUSCLE and a maximum-likelihood tree was constructed in MEGA v7 (Kumar et al., 2016) with default parameters.

Table 2 Positively selected genes in opah genome

ID	Gene	MA0 lnL	MA lnL	*P*-value	Description
EVM.Contig114.00094	*Mlc3*	−1 524.149140	−1 526.612165	0.026	myosin light chain 3;

Table 3 Rapidly evolving genes in opah genome

ID	Gene	Foreground-branch w	Background-branch w	*P*-value	FDR	Description
EVM.Contig371.00001	*Smn1*	0.83359	0.13734	0.004	0.040	survival motor neuron protein 1-like isoform X1; GO:0061061 muscle structure development
EVM.Contig17.00086	*Phkg1*	0.20475	0.06187	0.005	0.047	phosphorylase b kinase gamma catalytic chain, skeletal muscle/heart isoform; ko04020, calcium signaling pathway
EVM.Contig221.00008	*Htr2b*	0.52823	0.06636	0.003	0.030	5-hydroxytryptamine receptor 2B-like; ko04020, calcium signaling pathway
EVM.Contig593.00012	*Apj*	0.36011	0.06088	0.000	0.008	apelin receptor B-like;
EVM.Contig77.00015	*Acsl5*	0.26350	0.09102	0.002	0.028	long-chain-fatty-acid-CoA ligase 5; ko01212, fatty acid metabolism
EVM.Contig41.00022	*Mcat*	0.99212	0.05907	0.000	0.005	malonyl-CoA-acyl carrier protein transacylase;

						ko01212, fatty acid metabolism
EVM.Contig248.00004	*Hadha*	0.16190	0.06070	0.003	0.032	trifunctional enzyme subunit alpha; ko01212, fatty acid metabolism
EVM.Contig186.00023	*Acat1*	0.31986	0.07230	0.000	0.008	acetyl-CoA acetyltransferase; ko01212, fatty acid metabolism

Table 4 Expanded genes in opah genome

ID	Gene	Divergence size	Species size	Description
EVM.Contig8.00162 EVM.Contig825.00001 EVM.Contig825.00002	*Myoz1*	2	3	Myozenin-1; GO:0015629, actin cytoskeleton
EVM.Contig147.00066 EVM.Contig1226.00001 EVM.Contig2156.00001 EVM.Contig16.00123 EVM.Contig97.00033	*Igf1r*	4	5	Insulin-like growth factor-1 receptor; ko04152, AMPK signaling pathway
EVM.Contig191.00049 EVM.Contig25.00070 EVM.Contig448.00028 EVM.Contig5.00214 EVM.Contig5.00215 EVM.Contig5.00227 EVM.Contig549.00009	*Myo*	6	8	Myosin; GO:0014866, skeletal myofibril assembly

EVM.Contig83.00114				
EVM.Contig6.00189 EVM.Contig43.00013	*Calm*	1	2	Calmodulin; GO:0014866, skeletal myofibril assembly
EVM.Contig11.00072 EVM.Contig139.00013 EVM.Contig19.00153 EVM.Contig26.00017 EVM.Contig297.00011 EVM.Contig345.00021 EVM.Contig345.00022 EVM.Contig495.00016 EVM.Contig5.00024 EVM.Contig853.00006 EVM.Contig853.00007 EVM.Contig90.00002 EVM.Contig90.00003 EVM.Contig987.00005	*Act*	11	14	Actin; GO:0014866, skeletal myofibril assembly
EVM.Contig112.00025 EVM.Contig112.00026 EVM.Contig112.00027 EVM.Contig349.00008	*Ltb4r1*	3	4	Leukotriene B4 receptor 1-like; ko04020, calcium signaling pathway
EVM.Contig1153.00008 EVM.Contig1167.00008	*Ltb4r1*	1	2	Leukotriene B4 receptor 1-like; ko04020, calcium signaling pathway
EVM.Contig25.00012 EVM.Contig25.00018 EVM.Contig25.00020	*Ighd*	5	13	Immunoglobulin delta heavy chain; ko04020, calcium signaling pathway

EVM.Contig25.00024 EVM.Contig25.00036 EVM.Contig25.00037 EVM.Contig132.00006 EVM.Contig132.00015 EVM.Contig132.00029 EVM.Contig132.00031 EVM.Contig132.00039 EVM.Contig310.00001 EVM.Contig310.00014				
EVM.Contig2.00116 EVM.Contig297.00005 EVM.Contig33.00035 EVM.Contig606.00002 EVM.Contig8.00191 EVM.Contig90.00009	*Ryr*	4	6	Ryanodine receptor; ko04020, calcium signaling pathway
EVM.Contig559.00005 EVM.Contig559.00006	*Htr7*	1	2	5-hydroxytryptamine receptor 7; ko04020, calcium signaling pathway
EVM.Contig151.00016 EVM.Contig151.00017	*Serca3*	1	2	Sarcoplasmic/endoplasmic reticulum calcium ATPase 3;
EVM.Contig451.00001 EVM.Contig701.00012 EVM.Contig701.00013	*Serca1*	2	3	Sarcoplasmic/endoplasmic reticulum calcium ATPase 1;
EVM.Contig12.00209 EVM.Contig19.00147 EVM.Contig19.00148	*Hk*	3	4	Hexokinase; ko00010, glycolysis/gluconeogenesis

EVM.Contig64.00022				
EVM.Contig78.00048 EVM.Contig44.00007	*Mdh*	1	2	Malate dehydrogenase; ko00020, citrate cycle (TCA cycle)

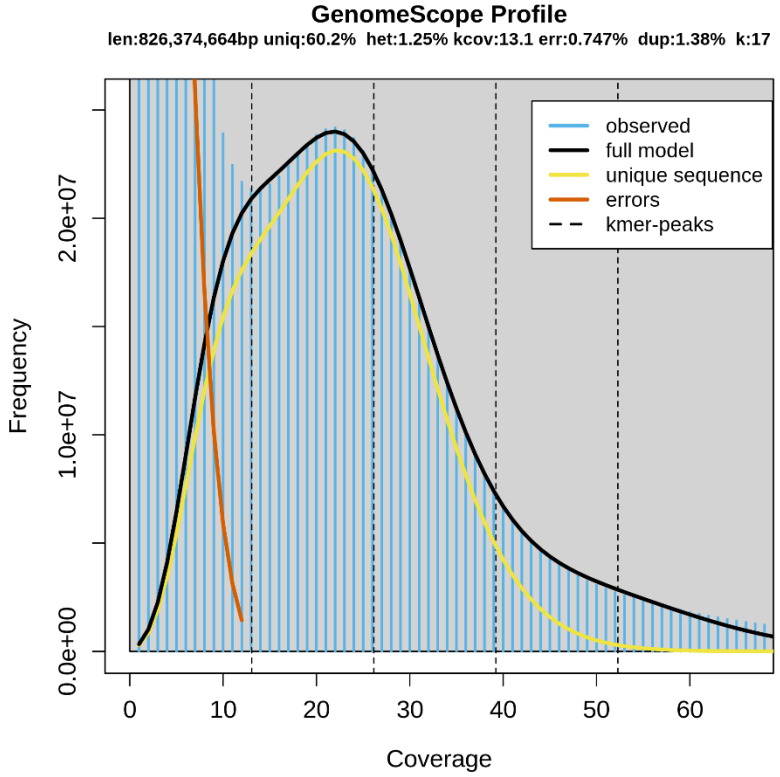

Figure 5 Illumina reads were used to estimate genome size with Jellyfish (Marçais & Kingsford, 2011). 17-mer depth frequency distribution was calculated using the *k*-mer method. Predicted genome size was 826 Mb.

Phylogenetic analysis

The protein sequences of zebrafish (*Danio rerio*; accession: GRCz11), medaka (*Oryzias latipes*; accession: ASM223469v1), Atlantic cod (*Gadus morhua*; accession: gadMor3.0), platyfish (*Xiphophorus maculatus*; accession: X_maculatus-5.0-male), stickleback (*Gasterosteus aculeatus*; accession: BROAD S1), Mexican tetra (*Astyanax mexicanus*; accession: Astyanax_mexicanus-2.0), Japanese puffer (*Takifugu rubripes*; accession: fTakRub1.2), spotted gar (*Lepisosteus oculatus*; accession: Eluc_v4), Amazon molly (*Poecilia formosa*; accession: Poecilia_formosa-5.1.2), climbing perch (*Anabas testudineus*; accession:fAnaTes1.2), Nile tilapia (*Oreochromis niloticus*; accession: O_niloticus_UMD_NMBU), big-finned mudskipper (*Periophthalmus magnuspinnatus*; accession: PM.fa), and northern pike (*Esox Lucius*; accession: Eluc_v4) were downloaded from Ensembl. The protein sequences were filtered based on the longest transcripts and low-quality sequences. Orthologous groups were constructed by OrthoFinder (Emms and Kelly, 2015) using the filtered sequences. A total of 151 one-to-one orthologous genes in opah and 13 ectothermic fish (*Danio rerio*, *Oryzias latipes*, *Gadus morhua*, *Xiphophorus maculatus*, *Gasterosteus aculeatus*, *Astyanax mexicanus*, *Takifugu rubripes*, *Lepisosteus oculatus*, *Poecilia formosa*, *Anabas testudineus*, *Oreochromis niloticus*, *Periophthalmus magnuspinnatus*, and *Esox lucius*) were identified. Multi-sequence alignments were generated at the protein level for the one-to-one orthologous genes using MAFFT v7.453 (Katoh and Standley, 2013). Gblocks v0.91b (Castresana, 2000) was used to extract the conserved sites in the multi-sequence alignment. Based on the filtered multi-sequence alignment, we constructed a phylogenetic tree using IQ-tree v2.0.3 with default parameters (Lam-Tung et al., 2015). MCMCTREE (PAML v9.1i) (Yang, 2007) was used to estimate opah divergence times. Fossil calibration information (divergence time of *Oryzias latipes* and *Xiphophorus maculatus* is 79–107 million years, *Xiphophorus maculatus* and *Oreochromis niloticus* is 88–139 million years, *Gasterosteus aculeatus* and *Takifugu rubripes* is 99–127 million years, *Danio rerio* and *Lepisosteus oculatus* is 295–334 million years, and *Danio rerio* and *Lampris megalopsis* is 206–252 million years) was obtained from http://timetree.org. The end time node used was 315 million years (http://timetree.org/). The gamma prior for the rate for genes was set to 1, 3.565 based on substitution rates inferred using BaseML. Tracer (Andrew et al., 2018) was used to analyze the MCMC.txt file and detect effective sample size (ESS). The ESS values were all higher than 200, indicating convergence.

Positive selection and rapid evolution analysis

As the oarfish (*Regalecus glesne*; accession: GCA_900302585.1) is the closest species to opah (Malmstrøm et al., 2016), we downloaded genomic data of the oarfish from the NCBI database (https://www.ncbi.nlm.nih.gov/). The corresponding oarfish genome sequence was searched against the single-copy orthologous proteins of opah using aTRAM v2.0 (Allen et al., 2018). Orthologous groups were constructed using OrthoFinder v.2.2.7 (Emms and Kelly, 2015) with filtered sequences of six fish species (i.e., opah, oarfish, stickleback, medaka, Atlantic cod, and northern pike). In total, 16 877 one-to-one orthologous genes from five species (oarfish, stickleback, medaka,

Atlantic cod, and northern pike) and opah were identified and compared. All orthogroups of the six species were aligned based on the coding sequences using Prank v.170427 (Löytynoja, 2014). Alignments gaps and ambiguous positions were removed using Gblocks under parameters -t=c -b5=h (Castresana, 2000). The base-substitution mutation rate of non-synonymous mutations is *Ka* and the base-substitution mutation rate of synonymous mutations is *Ks*. *Ka*/*Ks*>1 (i.e., positive selection) indicates that most non-synonymous mutations are favorable and are evolutionarily retained, and thus positive selection is the main driving force of biological evolution. To detect positively selected genes, the branch-site model applied in CODEML (PAML v9.1i) (Yang, 2007) was used, with opah as the foreground branch. *P*-values were calculated by likelihood ratio tests using chi-square distribution. Furthermore, the false discovery rate (FDR) function in the R package was applied to correct *P*-values. Positive sites with an FDR-corrected *P*>0.05 and no significant marker (higher than 0.95) for Bayes Empirical Bayes (BEB) were filtered out. For the remaining positively selected genes, MEGA (Kumar et al., 2016) was used to visually check whether the positive selection sites were located in poorly aligned regions. Thus, 108 genes were identified as under positive selection.

To identify genes under rapid evolution, we calculated the *Ka/Ks* values for single-copy genes using the branch model in CODEML, with opah as the foreground branch. *P*-values were calculated as described above. A higher *Ka/Ks* ratio on the foreground branch is classified as under rapid evolution, i.e., genetic change occurs rapidly enough to have a measurable impact on simultaneous ecological change. In total, 272 genes were identified as under rapid evolution.

Gene family expansion and contraction analysis

A gene family is a set of several similar genes, formed by duplication of a single original gene and generally with similar biochemical functions. Because of the low quality of the oarfish genome, gene family expansion and contraction analyses were performed for five species (opah, stickleback, medaka, Atlantic cod and northern pike) using Café v3.1 (Han et al., 2013). Analysis indicated that 1 157 gene families were expanded and 202 gene families were contracted.

Identification of candidate genes involved in endothermy

Positively selected genes, rapidly evolving genes, and expanded gene families were annotated using KEGG (https://www.genome.jp/kaas-bin/kaas_main). Eggnog (Huerta-Cepas et al., 2018) annotation and InterPro (Hunter et al., 2012) structural annotation results were integrated into GO annotation, with GO enrichment analysis then performed using Ontologizer (Bauer et al., 2008). According to thermogenesis and heat preservation, the enriched candidate genes were divided into muscle differentiation and development, enhancement of muscle contractile function, metabolic energy supply, reduction of lactic acid accumulation, *Sln*-mediated thermogenesis, formation of retia mirabilia, and fat formation and metabolism. SMART (Letunic et al., 2011) was used to view gene domain information. We also applied PolyPhen-2 (Adzhubei et al., 2010) and PROVEAN (Choi and Chan, 2015) to analyze the effects of the mutation sites on

proteins, and Phyre2 (Kelley et al., 2015) to predict the 3D structure.

DATA AVAILABILITY

The genomic, raw, and whole-genome sequence data reported in this paper were deposited in NCBI under BioProjectID PRJNA776990 and GSA at the National Genomics Data Center under accession CRA008922, which is publicly accessible at https://ngdc.cncb.ac.cn.

REFERENCES

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. 2010. A method and server for predicting damaging missense mutations. *Nature Methods*, **7**(4): 248-249.

Allen JM, LaFrance R, Folk RA, Johnson KP, Guralnick RP. 2018. aTRAM 2.0: An improved, flexible locus assembler for NGS data. *Evolutionary Bioinformatics*, **14**: 1176934318774546.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol*, **215**(3): 403-410.

Andrew R, Drummond AJ, Xie D, Guy B, Suchard MA. 2018. Posterior summarisation in Bayesian phylogenetics using Tracer 1.7. *Systematic Biology*(5): 5.

Bauer S, Grossmann S, Vingron M, Robinson PN. 2008. Ontologizer 2.0—a multifunctional tool for GO term enrichment analysis and data exploration. *Bioinformatics*, **24**(14): 1650-1651.

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

Bowman TA, O'Keeffe KR, D'Aquila T, Yan QW, Griffin JD, Killion EA, et al. 2016. Acyl CoA synthetase 5 (ACSL5) ablation in mice increases energy expenditure and insulin sensitivity and delays fat absorption. *Molecular metabolism*, **5**(3): 210-220.

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

Cheng H, Concepcion GT, Feng X, Zhang H, Li H. 2021. Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. *Nature methods*, **18**(2): 170-175.

Cho CH, Woo JS, Perez CF, Lee EH. 2017. A focus on extracellular Ca2+ entry into skeletal muscle. *Experimental and Molecular Medicine*, **49**.

Choi Y, Chan AP. 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics*, **31**(16): 2745-2747.

Chowdhury SS, Lecomte V, Erlich JH, Maloney CA, Morris MJ. 2016. Paternal high fat diet in rats leads to renal accumulation of lipid and tubular changes in adult offspring. *Nutrients*, **8**(9): 521.

Emms DM, Kelly S. 2015. OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome biology*, **16**(1): 157.

Franck JP, Slight-Simcoe E, Wegner NC. 2019. Endothermy in the smalleye opah (Lampris incognitus): a potential role for the uncoupling protein sarcolipin.

Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology, **233**: 48-52.

Hayhurst M, Wagner AK, Cerletti M, Wagers AJ, Rubin LL. 2012. A cell-autonomous defect in skeletal muscle satellite cells expressing low levels of survival of motor neuron protein. *Developmental Biology*, **368**(2): 323-334.

Hu J, Fan J, Sun Z, Liu S. 2020. NextPolish: a fast and efficient genome polishing tool for long-read assembly. *Bioinformatics*.

Huerta-Cepas J, Szklarczyk D, Heller D, Hernández-Plaza A, Forslund SK, Cook H, et al. 2018. eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Research*, **47**(D1): D309-D314.

Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, et al. 2012. InterPro in 2011: new developments in the family and domain prediction database. *Nucleic Acids Research*, **40**(D1): D306-D312.

Hyde JR, Underkoffler KE, Sundberg MA. 2014. DNA barcoding provides support for a cryptic species complex within the globally distributed and fishery important opah (Lampris guttatus). *Molecular Ecology Resources*, **14**(6): 1239-1247.

Kao H-J, Cheng C-F, Chen Y-H, Hung S-I, Huang C-C, Millington D, et al. 2006. ENU mutagenesis identifies mice with cardiac fibrosis and hepatic steatosis caused by a mutation in the mitochondrial trifunctional protein β-subunit. *Human Molecular Genetics*, **15**(24): 3569-3577.

Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, **30**(4): 772-780.

Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJ. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. *Nature protocols*, **10**(6): 845-858.

Khanna S, Merriam AP, Gong B, Leahy P, Porter JD. 2003. Comprehensive expression profiling by muscle tissue class and identification of the molecular niche of extraocular muscle. *The FASEB journal*, **17**(10): 1370-1372.

Kubišta V, Kubištová J, Pette D. 1971. Thyroid Hormone Induced Changes in the Enzyme Activity Pattern of Energy‐Supplying Metabolism of Fast (Whit) Slow (Red), and Heart Muscle of the Rat. *European Journal of Biochemistry*, **18**(4): 553-560.

Maurya SK, Periasamy M. 2015. Sarcolipin is a novel regulator of muscle metabolism and obesity. *Pharmacological research*, **102**: 270-275.

Kumar S, Stecher G, Tamura K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution*, **33**(7): 1870-1874.

Lam-Tung N, Schmidt HA, Arndt VH, Quang MB. 2015. IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Molecular Biology & Evolution*(1): 268-274.

Leskovec J, Sosic R. 2017. SNAP: A General-Purpose Network Analysis and Graph-Mining Library. *ACM transactions on intelligent systems*, **8**(1): 1.1-1.20.

Letunic I, Doerks T, Bork P. 2011. SMART 7: recent updates to the protein domain annotation resource. *Nucleic Acids Research*, **40**(D1): D302-D305.

Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv preprint arXiv:1303.3997*.

Löytynoja A. 2014. Phylogeny-aware alignment with PRANK. In. Multiple sequence alignment methods: Springer. p. 155-170.

Majoros W, Pertea M, Salzberg S. 2004. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics*, **20**(16): 2878-2879.

Malmstrøm M, Matschiner M, Tørresen OK, Star B, Snipen LG, Hansen TF, et al. 2016. Evolution of the immune system influences speciation rates in teleost fishes. *Nature Genetics*, **48**(10): 1204-1210.

Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics*, **27**(6): 764-770.

Pinte N, Coubris C, Jones E, Mallefet J. 2021. Red and white muscle proportions and enzyme activities in mesopelagic sharks. Comparative Biochemistry and Physiology Part B: *Biochemistry and Molecular Biology*, **256**: 110649.

Runcie RM, Dewar H, Hawn DR, Frank LR, Dickson KA. 2009. Evidence for cranial endothermy in the opah (Lampris guttatus). *Journal of Experimental Biology*, **212**(4): 461-470.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**(19): 3210-3212.

Smith S, Witkowski A, Moghul A, Yoshinaga Y, Nefedov M, de Jong P, et al. 2012. Compromised mitochondrial fatty acid synthesis in transgenic mice results in defective protein lipoylation and energy disequilibrium.

Stanke M, Diekhans M, Baertsch R, Haussler D. 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. *Bioinformatics*, **24**(5): 637-644.

Underkoffler KE, Luers MA, Hyde JR, Craig MT. 2018. A taxonomic review of Lampris guttatus (Brünnich 1788) Lampridiformes; Lampridae) with descriptions of three new species. *Zootaxa*, **4413**(3): 551-566.

Wegner NC, Snodgrass OE, Dewar H, Hyde JR. 2015. Whole-body endothermy in a mesopelagic fish, the opah, Lampris guttatus. *Science*, **348**(6236): 786-789.

Wu L, Zhang L, Li B, Jiang H, Duan Y, Xie Z, et al. 2018. AMP-activated protein kinase (AMPK) regulates energy metabolism through modulating thermogenesis in adipose tissue. *Frontiers in physiology*, **9**: 122.

Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Molecular biology and evolution*, **24**(8): 1586-1591.