

Xu A, Liu Y T, Huang K, *et al.* Chromosome-level genome assembly and cave adaptation mechanism analysis of *Odontamblyopus lacepedii* [J]. *Acta Hydrobiologica Sinica*, 2026, 50(4): 042606. [许安, 刘岩涛, 黄琨, 等. 拉氏狼牙虾虎鱼染色体水平基因组组装及洞穴适应机制解析 [J]. 水生生物学报, 2026, 50(4): 042606.]

拉氏狼牙虾虎鱼染色体水平基因组组装及洞穴适应机制解析

许安 刘岩涛 黄琨 俞晶 杨易静 朱登辉 李凤辉 刘晶
龚理 刘立芹 吕振明

(浙江海洋大学海洋科学与技术学院海洋种质资源开发与利用国家工程实验室, 舟山 316022)

摘要: 研究利用多平台测序数据, 对近盲虾虎鱼亚科物种拉氏狼牙虾虎鱼(*Odontamblyopus lacepedii*)开展了染色体水平基因组组装与注释, 并结合比较基因组学分析, 探讨了近盲虾虎鱼亚科的系统发育关系及其适应潮间带穴居生活的分子机制。结果表明: 拉氏狼牙虾虎鱼基因组大小为963.28 Mb, 序列被成功锚定至23条染色体上, 共注释到23558个蛋白质编码基因。系统发育分析表明, 近盲虾虎鱼亚科鱼类并未形成独立单系类群, 而是嵌套于背眼虾虎鱼亚科(Oxudercinae)内部, 从而与其共同构成并系关系, 支持了近盲虾虎鱼亚科与背眼虾虎鱼亚科合并为一个科的分类学观点。与四个背眼虾虎鱼亚科物种及其他7个外群物种相比, 在近盲虾虎鱼亚科谱系中共鉴定出了24个扩张基因家族、243个收缩基因家族及163个正选择基因。这些扩张和正选择基因主要富集在嗅觉感知、血管生成及脂质/碳水化合物代谢等相关信号通路, 可能与其适应黑暗、缺氧及营养匮乏的泥滩洞穴生活相关。此外, 还鉴定到了552个放松选择基因, 其中相当部分与视觉、色素沉着及昼夜节律相关, 提示可能与近盲虾虎鱼类视觉退化、体色弱化及昼夜节律丧失等退化性状形成密切相关。本研究为深入理解近盲虾虎鱼及其他潮间带穴居鱼类环境适应及关键性状演化机制提供了新见解。

关键词: 基因组组装; 比较基因组学; 洞穴适应; 拉氏狼牙虾虎鱼

中图分类号: Q953 **文献标识码:** A **文章编号:** 1000-3207(2026)04-042606-13

doi: 10.3724/1000-3207.2025.2025.0351 **CSTR:** 32229.14.SSSWXB.2025.0351



从水生向陆生的过渡是脊椎动物进化史上最重要的事件之一, 并最终导致地球上四足动物的产生和爆炸式辐射进化^[1]。在硬骨鱼类首次演化出陆生特征并主动迁移至陆地之后, 四足动物的祖先便成功从鱼类演化并完成登陆^[2, 3]。这一演化过程通常伴随着呼吸、感觉、运动、循环等系统的形态与生理的巨大改变^[4], 使动物能够适应陆生环境^[5, 6]。但值得注意的是, 并非所有硬骨鱼类谱系都沿着相同的演化路径完成登陆, 其中部分类群并未真正演化出稳定的陆生生活方式, 而是采取了一种较为被动的适应策略, 即通过栖息于潮间带泥滩中的洞穴, 从而规避水陆过渡所带来的负面影响。然而, 这些鱼类是如何适应泥滩洞穴环境的, 其进化机制仍不清楚。

拉氏狼牙虾虎鱼(*Odontamblyopus lacepedii*)是该类群鱼类的典型代表, 其隶属于鲈形目(Perciformes)背眼虾虎鱼科(Oxudercidae)近盲虾虎亚科(Amblyopinae)狼牙虾虎鱼属(*Odontamblyopus*)^[7], 主要分布于中国南部、韩国和日本的沿海水域^[7, 8], 为体型小、身体细长的底栖鱼类。通常栖息于潮间带泥滩深处的洞穴中, 也分布于河口泥底及邻近浅水区的沉积物中^[9]。为更好适应洞穴环境, 近盲虾虎鱼亚科物种演化出多种形态与生理适应特征, 主要包括被皮肤覆盖的退化小眼^[9, 10]、黑色素减少的白化表型^[11]、颊鳃盖腔内上皮中丰富的血管^[12, 13]及增强的饥饿耐受能力, 使其在退潮或寒冷季节长期潜伏于洞穴中时仍能存活^[13]。然而, 近盲虾虎鱼类这些形态与生理适应的遗传基础迄今仍不清

收稿日期: 2025-10-10; 修订日期: 2025-11-14

基金项目: 国家自然科学基金(42171069)资助 [Supported by the National Natural Science Foundation of China (42171069)]

作者简介: 许安(2001—), 男, 硕士研究生; 研究方向为生化与分子生物学。E-mail: 18949778106@163.com

通信作者: 吕振明(1976—), 男, 博士, 教授; 主要从事海洋生物种质资源与遗传育种研究。E-mail: nblzmn@zjou.edu.cn

楚。在系统发育层面,近盲虾虎鱼亚科与背眼虾虎鱼亚科同隶属于背眼虾虎鱼科,但二者在形态结构与生活习性上存在显著差异:前者为眼睛退化的底栖穴居鱼类,后者则为能够长时间在陆地上活动的“弹涂鱼”^[14]。传统形态学研究依据颅骨、颌骨和鳍条等特征对两者进行了区分^[14],然而,不同研究在两亚科的系统发育关系上仍存在争议。部分研究认为,近盲虾虎鱼亚科与背眼虾虎鱼亚科并不构成单系群,而是呈并系关系^[15]。因此,厘清二者之间的系统发育关系,对于理解近盲虾虎鱼亚科鱼类在水陆过渡过程中的独特演化模式具有重要意义。

近年来,基因组测序技术和比较基因组学的发展,为揭示非模式鱼类复杂环境下的适应性进化机制提供了新的契机。例如, Han等^[16]通过基因组与比较基因组学研究,揭示了深海鱼类如何通过视觉和能量代谢相关基因的趋同替换与基因家族扩张,以及DNA修复和膜结构相关基因的正选择来适应高压和低光的深海环境的;而Tong等^[17]通过相同的方法解析了高海拔淡水鱼类如何通过离子运输、凋亡、免疫和能量代谢相关基因的加速演化与正向选择,以适应高海拔地区的强碱和低氧环境的。这些发现不仅凸显了比较基因组学在解析鱼类环境适应性中的价值,也为理解泥滩洞穴栖息的近盲虾虎鱼类的特殊进化提供了新的研究思路。因此,本研究结合下一代DNBSEQ短读长测序数据(MGI Tech Co., Ltd., 中国深圳)、第三代Nanopore长读长测序数据(Oxford Nanopore Technologies (ONT))、Hi-C及RNA-Seq测序数据,完成了首个染色体水平拉氏狼牙虾虎鱼基因组的组装与功能注释,同时结合比较基因组学分析,旨在阐明近盲虾虎鱼类为适应潮间带洞穴生活所呈现的形态与生理创新的遗传机制。本研究不仅为后续近盲虾虎鱼类研究提供重要的基因组资源,同时也为揭示其从水生向陆生过渡过程中独特的演化模式及环境适应机理提供新见解。

1 材料与方法

1.1 DNA和RNA提取

本研究所使用的拉氏狼牙虾虎鱼来源于江苏赣榆海域(119°13'E, 34°52'N)。在采集活体后,取其肌肉、肝脏及肠等组织,并立即置于液氮中冷冻保存备用。随后,利用QIAGEN试剂盒(QIAGEN, Cat. No. 13343)从肌肉中提取基因组DNA;同时,使用TRIzol试剂(Invitrogen, Carlsbad, CA, USA)分别从肌肉、肝脏和肠组织中提取总RNA。在提取完成

后,首先通过1%琼脂糖凝胶电泳检测DNA和RNA的完整性及片段大小,随后借助Nanodrop 2000c紫外分光光度计测定其浓度与纯度。

1.2 文库构建、测序及原始数据质控

对于提取的拉氏狼牙虾虎鱼基因组DNA,用SQK-LSK109试剂盒(Oxford Nanopore Technologies)按照厂商说明进行Nanopore三代测序文库制备。随后,将构建好的文库加载至PromethION DNA测序平台(Oxford Nanopore Technologies)进行测序,从而获得第三代测序(Third Generation Sequencing, TGS)数据。随后采用Nanofilt v2.8.0^[18]软件根据参数“-q=7”过滤掉TGS数据中平均测序质量值(Phred quality score)低于7和长度短于1 kb的序列,用于后续的基因组组装。同时,采用MGIEasy FS DNA Library Prep Kit (BGI, Cat. No.1000006988)试剂盒构建短插入(350—700 bp)末端配对文库,并在MGIDNB (MGIDNB T7)平台上进行测序,从而获得二代NGS (Next Generation Sequencing)数据,用于后续基因组大小评估及辅助组装。为进一步开展染色体水平的基因组组装,采用Hi-C (High-throughput chromosome conformation capture)技术构建拉氏狼牙虾虎鱼的Hi-C文库。在文库构建过程中,首先对样品肌肉组织使用1%甲醛进行交联处理,然后利用Mbo I限制性内切酶对基因组DNA进行消化处理,并通过磁珠对提取的DNA进行纯化。构建完成的Hi-C测序文库随后在MGIDNB平台上进行测序,从而获得Hi-C数据。为进一步对组装的基因组进行转录组水平注释,使用VAHTS Universal V6 RNA-seq Library Prep Kit for MGI (Vazyme, NRM604)对每个RNA样本单独构建cDNA文库,并在MGIDNB平台上进行转录组测序。对获得的RNA测序数据,采用fastp v0.23.2^[19]软件根据参数“-l 50 -w 6”去除接头序列、长度小于50 bp及平均测序质量值低于Q20的序列;随后,使用FastQC v0.11.9^[20]对质控前后的数据进行质量评估。

1.3 基因组大小估计

在基因组组装前,首先采用GCE v1.0.2^[21]软件分两步对拉氏狼牙虾虎鱼基因组大小进行survey评估:首先,使用GCE内置程序kmerfreq v5.016,根据参数“-k=17, -l”计算K-mer频数。随后,利用内置程序gce根据上述K-mer频数对基因组的大小、杂合度及重复序列比例进行估算。

1.4 基因组组装和染色体挂载

采用“初步组装—打磨校正”的策略对Nanopore三代测序数据进行基因组组装。首先,使用Next-Genovo v2.4.0^[22]软件对经过质控的Nanopore长读

长数据进行拆分、比对及初步组装, 参数设置为“read_type=ont, read_cutoff=1k, pa_correction=3”。随后, 基于高质量的NGS二代短读长数据, 采用NextPolish v1.4.124^[23]软件根据参数“sgs_options=-max_depth 100-bwa, lgs_options=-min_read_len 1k -max_depth 100, lgs_minimap2_options=-x map -ont”对初步组装的基因组进行打磨和校正。接着, 使用Purge_haplotigs v1.0.4^[24]对打磨后的基因组进行去冗余处理, 去除杂合contigs, 从而获得纯合的单倍体基因组。

针对经过打磨和去冗余处理的基因组, 基于顺式互作强于反式互作的原理构建染色体互作图谱, 并进行可视化分析。具体步骤: (1)使用juicer v1.6^[25]软件将经过质控的Hi-C数据比对到基因组; (2)利用3D-DNA v180922^[26]软件对比对后的数据进行打断、排序和聚类, 以生成初步染色体互作图谱; (3)使用JuiceBox v1.11.08^[27]软件对互作图谱进行可视化, 并手动校正错误, 以获得最终的染色体级别基因组。

1.5 基因组组装质量评估

为评估最终组装的基因组质量, 使用BUSCO v5.3.2^[28]将基因组与辐鳍鱼类数据库(actinopterygii_odb10)进行同源基因比对, 其中单拷贝同源基因的检出比例越高, 表明基因组组装的完整性越好。为进一步验证组装质量, 还将测序获得的第三代(TGS)数据、二代(NGS)数据及转录组数据分别比对至拉氏狼牙虾虎鱼基因组上, 并统计比对率。其中, TGS Nanopore数据使用Minimap2 v2.17^[29]按参数“-ax map-ont”进行比对, NGS MGI数据则使用BWA-MEM v0.7.12^[30]并采用默认参数进行比对。比对率越高, 表明组装的准确性和完整性越高。

1.6 基因组注释

在进行基因组结构注释之前, 需要先识别并屏蔽基因组中大量的重复序列, 以提高注释的准确性。为此, 在DNA层面和蛋白质层面分别对基因组重复序列进行预测。在DNA水平上, 首先使用TRF v4.09^[31]和RepeatModeler v1.0.11软件基于基因组序列自身特征进行重复序列的初步预测; 随后采用RepeatMasker v4.0.9^[32]软件基于RepBase(<http://www.girinst.org/repbase>)及Dfam数据库进行进一步预测, 运行参数设置为“-nolow -no_is -norna”。在蛋白质水平上, 使用RepeatProteinMask (RepeatMasker 内置程序)基于RepBase数据库进行重复序列预测。最终, 通过RepeatMasker内置脚本“merge_repeatMasker_out.pl”整合多种方法的预测结果, 并对重复序列进行统计汇总与分析。

在基因组基因结构预测中, 采用从头(de novo)预测、同源预测和基于转录组预测3种方式, 具体步骤: (1)从头(de novo)预测: 使用Augustus v3.3.2^[33]软件对基因组进行结构预测, 参数设置为“-species=zebrafish -uniqueGeneId=true -noInFrame-Stop=true -gff3=on -strand=both”; (2)同源预测: 首先从NCBI数据库下载与拉氏狼牙虾虎鱼近缘且高质量的物种蛋白编码序列, 包括青鳉*Oryzias latipes* (GCF_002234675.1)、大弹涂鱼*Boleophthalmus pectinirostris* (GCF_026225935.1)、大鳍弹涂鱼*Periophthalmus magnuspinnatus* (GCF_009829125.3)、红鳍东方鲀*Takifugu rubripes* (GCF_901000725.2)及斑马鱼*Danio rerio* (GCF_000002035.6)。随后使用Genewise v2.4.1^[34]基于上述蛋白序列进行同源预测。(3)基于转录本的预测: 首先使用StringTie v1.3.5^[35]进行转录本组装, 参数为“-f 0.1 -m 200 -a 10 -c 2.5 -g 50 -M 1.0”。然后使用Hisat2 v2.1.0^[36]按参数“-dta -summary-file -S -x -1 -2”将转录本比对到基因组上, 并通过PASA v2.4.1^[37]软件进行基于转录本的基因预测。最终, 使用Maker v2.31.1^[38]软件综合整合上述3种预测结果, 生成最终的基因结构注释。为了研究注释基因的生物学功能, 将所得蛋白序列比对至多个公共蛋白数据库, 包括InterPro、GO、KEGG、SwissProt、TrEMBL、AnimalTFDB、Pfam、NR及KOG, 使用BLAST v2.3.0^[39], 参数设为“-e-value=1e-5”。所有的原始数据、基因组组装和注释文件均已存放到NCBI数据库中(登录号: BioProject PRJNA1001292; Biosample SAMN-36805684)。

1.7 系统发育树构建与分歧时间估算

本研究选取拉氏狼牙虾虎鱼与13个硬骨鱼类的基因组进行比较基因组学分析, 这13个硬骨鱼类的基因组均来自NCBI数据库, 包括: 斑马鱼(*Danio rerio*)、眼斑雀鳢(*Lepisosteus oculatus*)、大黄鱼(*Larimichthys crocea*)、红腹食人鱼(*Pygocentrus nattereri*)、暗绿鲀(*Tetraodon nigroviridis*)、青弹涂鱼(*Scartelaos histophorus*)、尼罗罗非鱼(*Oreochromis niloticus*)、广东弹涂鱼(*Periophthalmus modestus*)、瑞贝卡狼牙虾虎鱼(*Odontamblyopus rebecca*)、须鳗虾虎鱼(*Taenioides* sp.)、大鳍弹涂鱼(*Periophthalmus magnuspinnatus*)、大弹涂鱼(*Boleophthalmus pectinirostris*)及黑口新虾虎鱼(*Neogobius melanostomus*)。首先, 采用OrthoFinder v2.3.11^[40]软件按照默认参数鉴定这些物种的单拷贝直系同源基因。随后, 利用MUSCLE 3.8.31^[41]软件对这些单拷贝直系同源基因进行比对, 并使用Gblock v0.91b^[42]

软件对比对后的序列进行保守区块的提取, 之后将每个单拷贝直系同源基因的保守区块转换成氨基酸序列, 并采用RAxML (v 8.2.12)^[43]软件基于GTR模型构建最大似然树。接着利用PAML软件包(v 4.8)中MCMCtree程序, 基于贝叶斯松弛分子钟方法估算物种间的分歧时间^[44], 分歧时间校准点参考TIMETREE 数据库(<http://www.timetree.org>)^[45]的估算结果, 选取了3个校准点: 大鳍弹涂鱼与广东弹涂鱼(0.005亿—0.169 亿年前)、大弹涂鱼与拉氏狼牙虾虎鱼(0.315亿—0.373 亿年前)、斑马鱼与红腹食人鱼(1.32亿—1.70 亿年前)。

1.8 基因家族扩张与收缩鉴定

为了揭示近盲虾虎鱼类对泥滩洞穴环境的适应性进化机制, 开展了基因家族扩张与收缩鉴定。近盲虾虎鱼亚科中基因家族的扩张收缩采用CAFE v4.2^[46]软件进行鉴定。首先, 基于已构建的系统发育树及分歧时间, 利用CAFE的随机出生-死亡概率模型推断基因家族大小的变化, 在分析中, 将近盲虾虎鱼亚科谱系设定为关注的前景支, 并将其他鱼类作为背景支进行对比, 以识别其在进化过程中发生显著扩张或收缩的基因家族, 并以 $P < 0.05$ 作为显著性阈值。随后, 利用R包clusterProfiler (函数enrichGO和enrichKEGG)对扩张和收缩的基因家族进行GO和KEGG富集分析, 对于富集结果, 采用Benjamini-Hochberg方法进行多重检验校正, 并以 $P < 0.05$ 作为显著性阈值, 以确定显著富集的GO条目和KEGG通路。

1.9 正选择分析

为进一步揭示近盲虾虎鱼类对泥滩洞穴环境的适应性进化机制, 基于上述14个物种的单拷贝直系同源基因开展了正选择分析。首先将拉氏狼牙虾虎鱼、瑞贝卡狼牙虾虎鱼和须鳗虾虎鱼3个近盲虾虎鱼亚科鱼类设为前景支, 其余11个物种作为背景支, 采用PAML v4.8^[47]软件中的CODEML程序, 基于分支位点模型估算前景分支的dN/dS比值。随后采用分支位点模型 A (model=2, NSsites=2), 并将其与相应的零模型(ω 固定为 1)进行似然比检验(LRT), 再根据卡方分布计算显著性($P < 0.05$), 以识别前景支上的正选择位点。进一步采用clusterProfiler包对3个前景物种的正向选择基因进行GO和KEGG富集分析, 并以 $P < 0.05$ 作为显著性阈值, 从而识别出显著富集的GO条目和KEGG通路。

1.10 放松选择分析

为全面评估近盲虾虎鱼类对泥滩洞穴环境的潜在适应机制, 还采用HYPHY (2.5.74)软件中的RELAX程序按照默认参数对拉氏狼牙虾虎鱼及其

所在的近盲虾虎鱼亚科物种中可能存在的选择放松进行分析^[48]。在分析过程中, 将3个近盲虾虎鱼亚科物种(拉氏狼牙虾虎鱼、瑞贝卡狼牙虾虎鱼和须鳗虾虎鱼)设为前景支, 其余11个外群物种设为背景支, 采用RELAX程序计算选择强度参数 k , 以量化前景支相对于背景支的 ω 分布变化。当 $k < 1$, ω 分布向1靠拢, 表示选择压力放松; 当 $k > 1$, ω 分布远离1, 表示选择压力强化。对选择压力放松或是强化的判定采用零假设模型($k=1$)与备择模型($k \neq 1$)进行比较评估, 并使用似然比检验(LRT)来估算其显著性: 若 $k < 1$ 且 $P < 0.05$, 则前景分支经历显著放松选择; 若 $k > 1$ 且 $P < 0.05$, 则经历显著强化选择。

2 结果

2.1 拉氏狼牙虾虎鱼基因组组装与表征

通过对测序数据的处理和过滤, 共获得253.27 Gb (68.87 Gb NGS reads, 57.07 Gb Nanopore reads, 128.33 Gb Hi-C reads)的clean reads (表 1), 测序深度约为拉氏狼牙虾虎鱼基因组大小的59.25 $\times$ 。基于K-mer分析(图 1), 预估拉氏狼牙虾虎鱼基因组大小约为976.75 Mb。随后使用NextDenovo (v2.4.0)

表 1 拉氏狼牙虾虎鱼的de novo测序数据统计

测序平台 Sequencing platform	插入片段 Insert size	平均序列长度 Mean read length (bp)	干净数据 Clean data (Gb)	测序深度 Sequence coverage ($\times$)
MGIDNB	300	150	68.87	70.46
Nanopore	—	17351	57.07	59.25
Hi-C	300	150	128.33	133.22

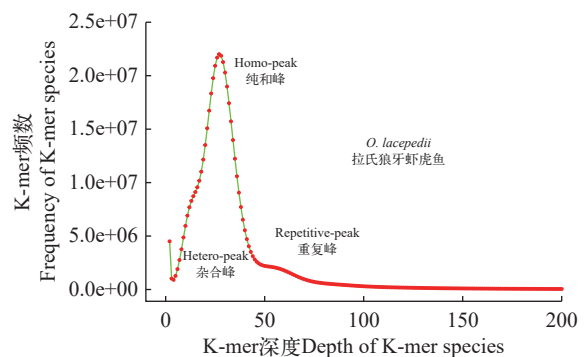


图 1 拉氏狼牙虾虎鱼基因组测序数据的17-mer频率分布图
Fig. 1 17-mer frequency distribution of the genome sequencing data of *O. lacepedii*

图中横坐标表示K-mer的测序深度分布, 纵坐标表示K-mer的数量。图中出现的峰从左到右分别为杂合峰、纯合峰及重复峰
The X-axis indicates K-mer depth, while the Y-axis shows the number of K-mers. In both graphs, the peaks from left to right correspond to the heterozygous peak, the homozygous peak, and the repetitive peak

软件对Nanopore三代测序数据进行初步组装, 获得大小为981.74 Mb的拉氏狼牙虾虎鱼基因组。随后, 使用Purge haplotigs软件对基因组进行去冗余处理, 最终获得拉氏狼牙虾虎鱼的纯合基因组, 长度为963.28 Mb (表 2)。进一步利用 Hi-C 数据将组装的基因组提升至染色体级别。结果显示, 共有960298431 bp的序列成功挂载到23条染色体上, 挂载率约为99.69%, 各染色体长度为24.18—52.56 Mb。Hi-C互作热图显示各条染色体的边界分明, 互作信号强烈且集中, 提示染色体级基因组的挂载质量较高, 结果具有较高可靠性(图 2A)。通过原始数据回帖、BUSCO评估及与近缘种的共线性分析, 对最

终获得的拉氏狼牙虾虎鱼染色体水平基因组组装质量进行综合评估。结果显示, 该基因组的contig N50长度达12.12 Mb, scaffold N50长度达42.70 Mb; 测序数据比对显示, NGS、Nanopore及转录组数据的比对成功率分别达到99.64%、99.84%和98.94%; BUSCO分析结果显示, 在辐鳍鱼类数据库(Actinopterygii_odb10)所含的3640个直系同源基因中, 有96.02%能够在拉氏狼牙虾虎鱼基因组中被成功识别(表 3); 此外, 共线性分析结果进一步表明, 组装所得的拉氏狼牙虾虎鱼基因组与近缘模式物种斑马鱼基因组之间呈现出较高的共线性(图 2C)。综上所述, 本次组装的拉氏狼牙虾虎鱼基因组具有较

表 2 拉氏狼牙虾虎鱼基因组组装结果

Tab. 2 Genome assembly results of *O. lacepedii*

组别Group	N50 (bp)	N90 (bp)	最长序列长度 Maximum sequence length (bp)	基因组大小 Genome size (bp)	序列数目 Sequence number
NextDenovo	12122453	1224113	31943383	981742888	449
NextPolish	12687803	1444685	31943383	964250794	352
Hi-C	42702054	36610850	52560117	963284094	61

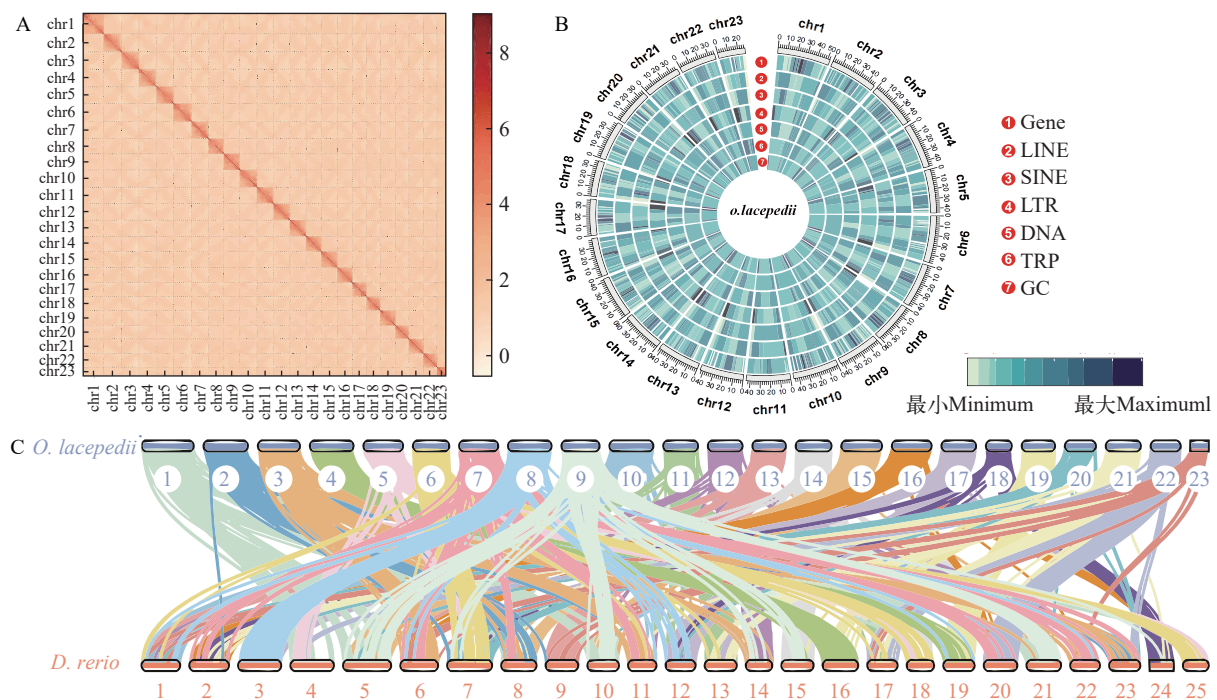


图 2 拉氏狼牙虾虎鱼的基因组组装、注释及共线性结果

Fig. 2 Genome assembly, annotation, and synteny analysis of *O. lacepedii*

A. 拉氏狼牙虾虎鱼中染色体相互作用热图。图中的方块代表23条假染色体。颜色条指示接触密度, 从蓝色(低)到红色(高); B. 拉氏狼牙虾虎鱼基因组元素分布的Circos图; 从外圈到内圈: 1. 蛋白质编码基因; 2. 长散在核元件(LINEs); 3. 短散在核元件(SINEs); 4. 长末端重复序列(LTRs); 5. DNA转座子(DNA); 6. 串联重复序列(TRPs); 7. GC含量; C. 拉氏狼牙虾虎鱼和斑马鱼基因组之间的共线性分析结果; 图中的数字代表每个物种的染色体身份

A. A heatmap of chromosomal interactions in *O. lacepedii*. The blocks represent the 23 pseudochromosomes. The colour bar illuminates the contact density from blue (low) to red (high); B. Circos plot of distribution of the genomic elements in *O. lacepedii*. From the outer circle to the inner circle: 1. protein-coding genes; 2. long interspersed nuclear elements (LINEs); 3. short interspersed nuclear elements (SINEs); 4. long terminal repeats (LTRs); 5. DNA transposons; 6. tandem repeats (TRPs); 7. GC content; C. Synteny between genomes of *O. lacepedii* and *D. rerio*; The number in the figure represents the chromosome identity for each species

好的完整性和较高质量。

在进行基因组结构注释之前,使用RepeatModeler v1.0.11与RepeatMasker v4.0.9^[32]对基因组重复序列进行了注释。结果表明,拉氏狼牙虾虎鱼基因组中重复序列总长度约为485.5 Mb,占基因组总大小的50.40%。其中主要类型为DNA转座子(占基因组的18.19%)、长散在核元件(LINE, 10.37%)、短散在核元件(SINE, 1.07%)及长末端重复序列(LTR, 5.88%)等。同时采用de novo预测、同源预测、转录本预测3种方法对拉氏狼牙虾虎鱼基因组进行基因结构预测,结果共预测到了23558个蛋白编码基因(表4)。进一步将上述蛋白编码基因比对到InterPro、GO、KEGG、SwissProt、TrEMBL、TF、Pfam、NR及KOG数据库以进行功能注释。结果显示,在拉氏狼牙虾虎鱼基因组中共鉴定到22177个蛋白编码基因,注释成功率达到94.14%。这些基因的平均长度为16059.11 bp,其中外显子平均长度为176 bp,内含子平均长度为1708 bp。

2.2 比较基因组学和系统发育分析

在拉氏狼牙虾虎鱼与13个外群物种的基因组

表3 基于BUSCO分析的拉氏狼牙虾虎鱼基因组组装质量评估
Tab. 3 Assessment of genome assembly quality for *O. lacepedii* based on BUSCO analysis

指标Index	基因数量 Gene number	百分比 Percentage (%)
完整BUSCOs	3495	96.02
完整且为单拷贝BUSCOs	3449	94.75
完整且为重复拷贝BUSCOs	46	1.26
片段化BUSCOs	21	0.58
缺BUSCOs	124	3.41
总BUSCO基因数	3640	100.0

表4 拉氏狼牙虾虎鱼基因组功能注释结果

Tab. 4 Functional annotation of predicted protein-coding genes in *O. lacepedii*

基因组Gene	数目Number	百分比Percent age (%)
Annotated	22177	94.14
InterPro	20643	87.63
GO	15706	66.67
KEGG	21585	91.62
SwissProt	19657	83.44
TrEMBL	21854	92.77
TF	3635	15.43
Pfam	19917	84.54
NR	22048	93.59
KOG	18248	77.46
Unannotated	1381	5.86
总计Total	23558	100

中,共鉴定出了3610个单拷贝直系同源基因。随后利用鉴定到的这些单拷贝直系同源基因的氨基酸序列构建了最大似然(ML)系统发育树。结果显示,拉氏狼牙虾虎鱼与同亚科的瑞贝卡狼牙虾虎鱼和须鳗虾虎鱼首先聚为一支,置信度为100,显示三者存在紧密的亲缘关系。但有意思的是,该支系并未形成独立的近盲虾虎鱼类单系群,而是进一步与背眼虾虎鱼亚科的大弹涂鱼和青弹涂鱼聚类,最终与广东弹涂鱼和大鳍弹涂鱼形成一个更大的分支。该结果表明,近盲虾虎鱼亚科Amblyopinae可能并非独立的单系群,而是镶嵌于背眼虾虎鱼亚科之中,呈现出明显的并系关系(图3A)。该结果与前期基于线粒体基因组的分析结论相吻合^[49],并与Stepan等^[50]基于线粒体及核基因所进行的分析结果一致。通过化石校准进一步估算了拉氏狼牙虾虎鱼及其代表的近盲虾虎鱼类的起源时间,结果表明:拉氏狼牙虾虎鱼、瑞贝卡狼牙虾虎鱼和须鳗虾虎鱼的共同祖先与大弹涂鱼大约在33.19 Ma(渐新世早期)产生分化(图3A)。分歧时间与之前根据线粒体基因组分析估计的近盲虾虎鱼亚科和背眼虾虎鱼亚科之间的分歧时间(34.5 Ma)基本一致^[49]。渐新世早期(~34—27 Ma)被广泛认为是全球变冷阶段的开始,该时期由于大气中二氧化碳含量的急剧下降而使全球气温急剧降低,从而导致全球海平面下降^[51, 52]。而全球变冷及海平面下降被认为与陆生虾虎鱼的快速演化密切相关,因为较低的海平面可能提供了更多泥滩环境,有利于硅藻的生长,从而促进某些弹涂鱼物种的辐射演化^[1]。这种情况是否也导致了近盲虾虎鱼亚科鱼类的起源与分化仍有待进一步研究。

2.3 与洞穴适应相关的基因组适应性进化特征

为进一步解析近盲虾虎鱼基因组适应性进化特征及其泥滩洞穴环境适应机理,分析拉氏狼牙虾虎鱼及其所在的近盲虾虎鱼类基因组中基因家族的扩张和收缩情况。与其他11个物种相比,在近盲虾虎鱼亚科谱系中共鉴定出24个扩张的基因家族和243个收缩的基因家族($P < 0.05$),并对这些基因家族进行功能富集。结果显示,这些扩张的基因家族主要富集于模式识别受体Pattern recognition receptors、嗅觉感知Sensory perception of smell、嗅觉受体活性Olfactory receptor activity、G蛋白偶联受体G protein-coupled receptors等信号通路(图4A)。而收缩的基因家族主要富集于肽抗原结合peptide antigen binding、视黄醇代谢Retinol metabolism、淀粉和蔗糖代谢Starch and sucrose metabolism等信号通路(图4B)。

为进一步理解拉氏狼牙虾虎鱼及其所在的近盲虾虎鱼类泥滩穴居适应的基因组机制, 鉴定了包括拉氏狼牙虾虎鱼 *O. lacepedii*、瑞贝卡狼牙虾虎鱼 *Odontamblyopus rebecca* 和须鳗虾虎鱼 *Taenioides* sp. 在内的近盲虾虎鱼谱系中的正选择基因, 结果表明: 相对于11个外群物种, 在近盲虾虎鱼分支中共检测到163个正选择基因(PSGs)。对这163个正选择基因进行功能注释, 发现某些基因在近盲虾虎鱼类应对洞穴环境挑战的过程中可能产生重要影响, 这些基因功能类别包括: DNA修复(*CDC45*、*MMS19*、*RECQL4*)^[53–55]、免疫应答(*IFRD1*、*IL21R*)^[56, 57]、造血或血管生成(*ANGPTL1*、*TRPC1*、*VEGFC*)^[58–60]、胰岛素抵抗(*FOXO1*、*TPCN2*、*CILP2*)^[61–63]。

此外, 部分退化性状可能也在洞穴鱼类适应其独特生活环境中扮演着重要角色。为了理解近盲虾虎鱼类在泥滩洞穴环境适应中部分退化性状的

基因组机制, 以拉氏狼牙虾虎鱼、瑞贝卡狼牙虾虎鱼和须鳗虾虎鱼在内的3个近盲虾虎鱼亚科谱系物种为前景支, 11个硬骨鱼类为外群。共鉴定出552个放松选择基因。对处于放松选择的直系同源基因进行功能注释, 发现其中某些基因的功能与近盲虾虎鱼类穴居适应下的退化性状有关, 这些基因功能类别包括视觉感知(*SAMD7*、*CIQTNF5*、*EFEMP1*、*FYCO1*、*CHD7*、*RTN4IP1*)、色素沉积(*OCA2*、*DCT*、*KITLG*、*HPS1*、*HPS4*、*BCO1*)和昼夜节律(*FBXL21*、*TIMELESS*、*PDPI1*)等。

3 讨论

3.1 近盲虾虎鱼亚科的系统发育及演化历史讨论

传统观点认为, 近盲虾虎鱼亚科与背眼虾虎鱼亚科是鲈形目背眼虾虎鱼科中形态和生态截然不同的两个演化支: 前者为眼睛退化的底栖穴居鱼

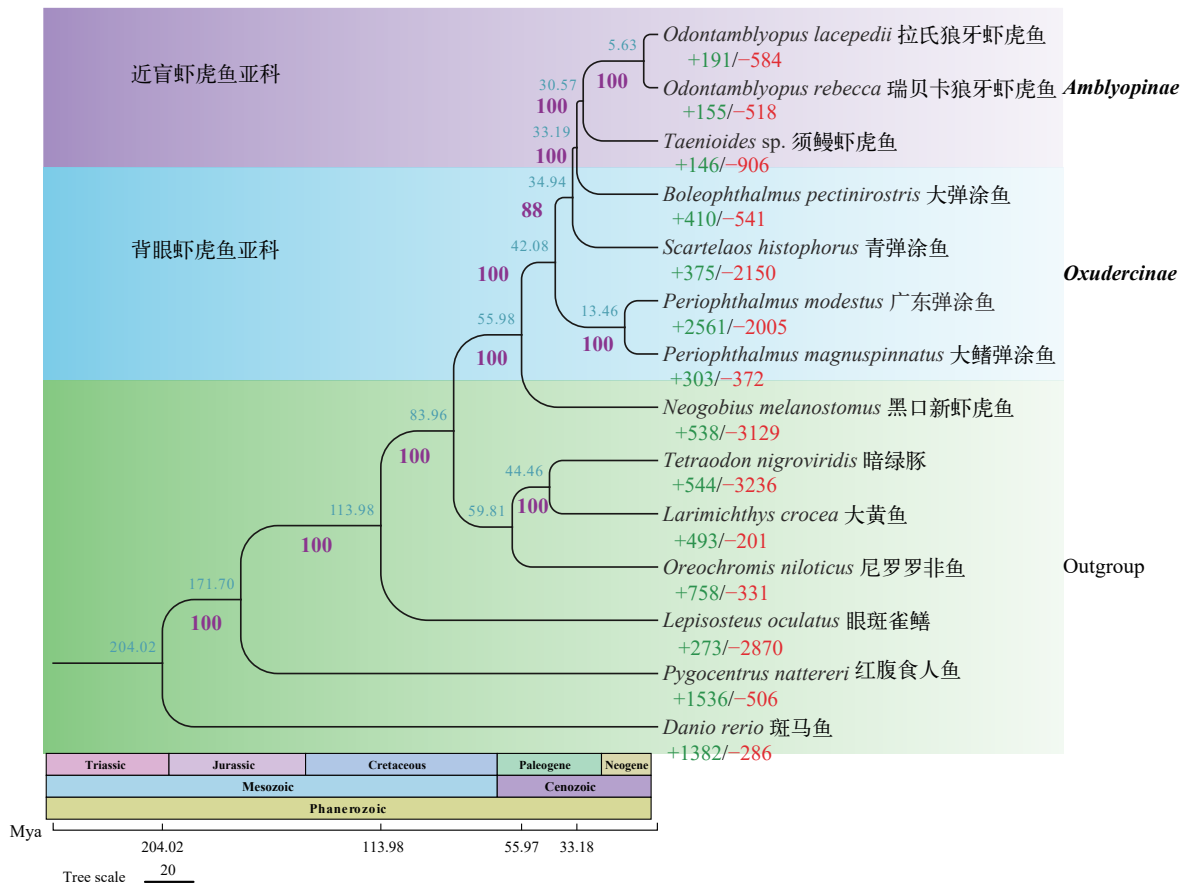


图3 拉氏狼牙虾虎鱼的系统发育分析: 拉氏狼牙虾虎鱼与13个外群物种的系统发育关系及分化时间

Fig. 3 Phylogenetic analysis of *O. lacepedii*: Phylogenetic relationships and estimated divergence times between *O. lacepedii* and 13 outgroup species

每个节点上的蓝色数字表示物种间的分化时间, 绿色和红色数字分别表示该节点处扩张和收缩的基因家族数量, 紫色数字表示置信度; 图底部显示了地质时期及其对应年代(百万年前, Mya)

Blue numbers at each node indicate estimated divergence times between species; Green and red numbers represent the number of gene families that have expanded or contracted at each node, respectively. Purple numbers indicate confidence levels; Geological time periods and corresponding ages (in millions of years, MYA) are shown at the bottom of the figure

类, 后者则为能够长时间在陆地上活动的“弹涂鱼”。Murdy 等^[14]曾提出了可以区分两者的5个形态学特征, 包括颅骨和眼睛的排列、鼻孔的形态、颞骨的形态、上颌骨的排列以及臀鳍的数目等。但Thacker等^[15]通过系统发育研究发现, 近盲虾虎鱼亚科与背眼虾虎鱼亚科并不能构成单系群, 而是呈并系关系, 并指出Murdy提出的5个形态特征在部分近盲虾虎鱼物种中同样存在, 说明这些特征并不能作为区分两个亚科的特征。本研究的系统发育结果支持了Thacker等^[15]的观点: 构建的ML系统发育树表明, 近盲虾虎鱼亚科物种(拉氏狼牙虾虎鱼、瑞贝卡狼牙虾虎鱼和须鳃虾虎鱼)与背眼虾虎鱼亚科(青弹涂鱼、大鳍弹涂鱼、大弹涂鱼、广东弹涂鱼)构成并系关系。这一结果也与前期基于线粒体基因组^[49]和Steppan等^[50]基于线粒体和核基因的研究结果相符。这种并系的系统演化关系可能暗示了近盲虾虎鱼类非常独特的系统演化路径: 即近盲虾虎鱼亚科鱼类与背眼虾虎鱼亚科鱼类首先由其共同祖先演化出陆生特征, 其后近盲虾虎鱼亚科这一独特分支则在后续的演化中陆生能力逐步退化, 而演化成现今更亲水的潮间带或潮下带穴居生活的类群。这一结论也在前期基于线粒体基因组研究的结果中得到印证^[49], 该结果可能表明背眼虾虎鱼亚科祖先在向陆生类群演化过程中并未像背眼虾虎鱼那样成功, 而可能由于古地理历史或气候变化等原因逐渐又向水生退却, 成为鱼类从水生向陆生演化上并不成功的案例。当然, 这一假说仍需结合更多分子证据和化石记录加以验证。

3.2 近盲虾虎鱼亚科鱼类适应洞穴生活的基因组特征

近盲虾虎鱼作为洞穴生活的鱼类, 其通常生活于黑暗、低氧和寡营养的洞穴环境中, 为了应对该独特生活环境的挑战, 其基因组也产生了相应的适应性改变。例如由于长期生活在黑暗的深穴环境中, 穴居鱼类普遍表现出视觉退化, 但常伴随其他感觉补偿机制的出现, 尤其是嗅觉感知的增强。已有研究表明, 洞穴鱼相比水面近缘种, 对化学信号的感知灵敏度显著提升, 能够响应浓度低 10^5 倍的刺激^[64]。在本研究中, 通过比较基因组分析发现, 近盲虾虎鱼谱系的OR (Olfactory Receptor, 嗅觉受体)基因家族发生了显著扩张。这种扩张可能是其在黑暗环境条件下, 通过增强特定化学信号的感知能力以补偿视觉功能不足的一种补偿机制, 该结果也得到了后续正选择分析结果的证实。正选择分析结果显示, 近盲虾虎鱼谱系中V2R1 (Vomerolateral type 2 receptor 1, 犁鼻器受体2型基因1)存在显著的正选择信号。在硬骨鱼中, V2R主要在鼻腔的嗅觉上皮(OE)中表达。多项研究表明, 硬骨鱼的V2R能够感知引发特定摄食行为的氨基酸^[65]。此外, 通过遗传学手段阻断V2R表达神经元的神经传递会导致鱼类对氨基酸混合物的趋向反应消失, 进一步证明了V2R在摄食行为中的关键作用^[66]。综上所述, 近盲虾虎鱼在洞穴环境中可能通过嗅觉受体基因的扩张和关键基因的适应性进化, 加强了对化学信号的感知, 从而使嗅觉在环境感知中占据更重要的位置, 并在一定程度上弥补了视觉功能的退化。

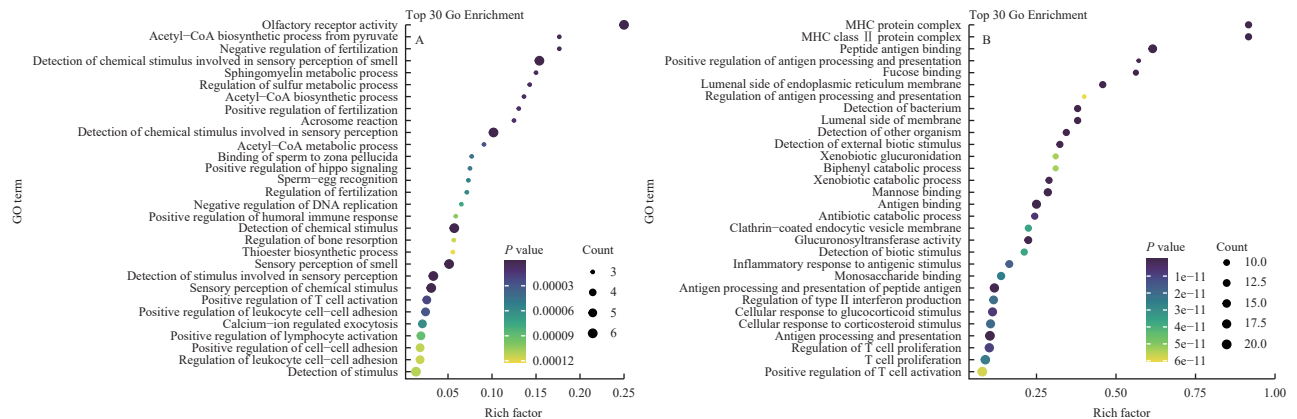


图4 扩张收缩基因的GO富集分析

Fig. 4 GO enrichment analysis of expanded and contracted selection genes

A. 近盲虾虎鱼扩张基因Top 30 GO富集结果, 横轴代表富集因子, 纵轴代表富集的GO term; B. 近盲虾虎鱼收缩基因Top 30 GO富集结果, 以拉氏狼牙虾虎鱼*O. lacepedii*、瑞贝卡狼牙虾虎鱼*Odontamblyopus rebecca*和须鳃虾虎鱼*Taenioides* sp.为前景分支

A. Top 30 GO terms significantly enriched among expanded genes in Amblyopinae; The horizontal axis depicts enrichment factors, while the vertical axis shows the enriched GO terms; B. Top 30 GO terms significantly enriched among contracted genes in Amblyopinae, with *Odontamblyopus lacepedii*, *Odontamblyopus rebecca*, and *Taenioides* sp. as the foreground branches

为了适应洞穴中频繁出现的低氧环境^[13], 拉氏狼牙虾虎鱼已进化出鳃盖腔内高度血管化的上皮组织, 从而在氧气不足时能够直接呼吸空气^[67]。野外生态学观察与实验室研究结果均表明: 拉氏狼牙虾虎鱼能够在缺氧环境下启动空气呼吸, 从而适应泥滩洞穴的低氧水体^[12]。然而, 该性状的遗传基础尚不清楚。本研究通过比较基因组分析发现, 近盲虾虎鱼谱系中有3个与造血或血管生成相关的基因表现出正选择信号(*ANGPTL1*、*TRPC1*、*VEGFC*), 这些基因编码的蛋白质对于血管的生成和发育至关重要^[58–60]。例如, *ANGPTL1*是鱼类造血与血管分化的关键基因, 研究表明, *ANGPTL1*与*ANGPTL2*的同时敲降会导致其出芽期内皮细胞凋亡增加, 从而造成严重的血管缺陷^[58]。此外, *TRPC1*基因在鱼类血管发育中也具有重要作用, 其敲降后会导致斑马鱼血管生成障碍^[59]。因此, 本研究检测到的这些基因的正选择可能有助于近盲虾虎鱼类增强鳃盖腔内部的血管生成能力及维持口腔上皮的血管化, 从而有利于其应对洞穴低氧环境的挑战。

此外, 穴居动物通常面临周期性食物短缺。为了适应这种营养匮乏环境, 它们往往进化出胰岛素抵抗的生理特征, 从而维持较高血糖水平并增强能量储存能力, 以应对饥饿期间能量供应不足, 这一特征在墨西哥脂鲤(*Astyanax mexicanus*)中亦有报道^[68]。通过比较基因组分析进一步印证了洞穴鱼类的这一适应策略。本研究发现, 近盲虾虎鱼谱系中3个与胰岛素代谢相关的基因呈显著正选择信号(*FOXO1*、*TPCN2*和*CILP2*)。在哺乳动物中, *FOXO1*是胰岛素信号通路的核心转录因子, 其过表达通常会导致胰岛素抵抗^[61]。转基因小鼠实验显示, *FOXO1*持续激活可显著升高空腹血糖水平并降低胰岛素敏感性^[69]。而对*FOXO1*进行敲除则可增强胰岛素敏感性, 从而有效减轻胰岛素抵抗^[70]。已有研究表明, 拉氏狼牙虾虎鱼的进食活动主要发生在涨潮期^[13]。结合野外观测结果发现, 该物种通常在半月潮期间才大量出洞觅食, 且在寒冷季节活动明显减少, 长期潜伏于洞内。这种有限的觅食时机意味着个体需要在长时间的食物缺乏期维持能量平衡。因此, 这些基因的正选择可能赋予近盲虾虎鱼类一定程度的胰岛素抵抗能力, 使其能够在营养匮乏的洞穴环境中, 或在低潮及寒冷季节长期埋栖时, 维持较高的血糖水平并促进能量储存, 从而在营养输入受限的条件下仍能维持机体功能, 以适应其寡营养生态位。

3.3 近盲虾虎鱼类洞穴适应下的退化性进化

在穴居鱼类进化过程中, 退化性进化有时也是

其环境适应不可忽视的一部分^[71], 它往往是不适应性的, 导致结构、功能和相应基因的丧失^[71, 72]。穴居鱼类是研究生物退化性进化的理想模型, 因为它们进化出一系列独特的穴居适应特征, 如视觉功能退化、色素沉着减少及昼夜节律丧失等^[73]。研究这些特征的形成机制, 有助于深入理解穴居鱼类的退化性适应。诚然, 本研究的近盲虾虎鱼亚科物种也进化出了退化性小眼特征, 这通常被认为是适应浑浊水域穴居环境的退化性状^[15]。在本研究中, 通过比较基因组分析发现, 近盲虾虎鱼谱系中的 γ -晶状体蛋白家族发生显著收缩, 具体表现为*CRYGM2*基因数量减少。这种收缩可能与近盲虾虎鱼类长期生活在黑暗洞穴环境中, 视觉系统逐渐退化相关。此外, 还发现多种与视觉相关的基因在近盲虾虎鱼谱系中发生了选择放松(*SAMD7*、*EFEMP1*和*CHD7*)。这些基因在视觉系统的发育及维持中发挥重要功能。其中, *SAMD7*是细胞类型特异性的多梳抑制复合物1 (PRC1)组分, 对视杆感光细胞身份的建立至关重要^[74]。在人类中, *SAMD7*基因突变可导致常染色体隐性黄斑营养不良症, 进而造成严重视功能损伤^[75]。而在斑马鱼中, *SAMD7*缺失则会导致视杆细胞分化障碍和视锥细胞谱系错乱, 从而破坏视觉功能^[76]。因此, 这些基因的收缩及选择放松可能导致近盲虾虎鱼类视觉系统在发育及功能上的改变, 从而在分子水平上反映其对洞穴黑暗环境的退化性适应。

同样, 本研究还检测到一些与色素沉着相关的基因(*OCA2*、*DCT*、*KITLG*、*HPS1*、*HPS4*、*BCO1*)呈现选择放松, 其中最为显著的是*OCA2*基因。在剑尾鱼(*Xiphophorus hellerii*)、马拉维金慈鲷(*Melanochromis auratus*)、墨西哥盲眼鱼(*Astyanax mexicanus*)等鱼类中, *OCA2*基因都被证实与色素沉着密切相关, 其突变往往会导致白化表型^[77–79]。此外, Klaassen等^[80]的遗传学研究阐明了*OCA2*在鱼类色素沉积中的作用机制, 并指出该基因的缺失是多个洞穴鱼白化表型的直接遗传原因。本研究中的大多数近盲虾虎鱼类亦表现出黑色素减少所导致的白化特征, 其体色通常呈粉色、紫色或红色^[11], 这一现象与洞穴鱼类中已报道的白化与色素减退机制相一致^[81]。因此, *OCA2*等基因的选择放松可能在近盲虾虎鱼类的色素退化过程中发挥了关键作用, 为理解穴居鱼类色素沉着的演化提供了分子证据。

此外, 本研究还发现了一些与昼夜节律相关的基因(*FBXL21*、*TIMELESS*、*PDPI*)在近盲虾虎鱼类中发生了选择放松。其中, *TIMELESS*是果蝇昼夜节律钟的核心组成部分, 它通过多层调控机制精

确控制PERIOD蛋白的周期性表达及活性,从而维持果蝇24h昼夜节律^[82]。在哺乳动物中,TIMELESS同样与昼夜节律密切相关。例如,在大鼠中条件性敲低TIMELESS蛋白的表达会扰乱视交叉上核神经元的活动节律,并改变已知核心生物钟元件的表达水平^[83]。此外,在人类中,TIMELESS的突变会导致昼夜节律的失调^[84]。大量前人的研究表明,许多洞穴鱼类的昼夜节律经历了独特进化,例如,墨西哥盲眼鱼缺乏新陈代谢的昼夜节律^[85]。据推测,生活在永久黑暗环境中的动物可能通过减弱或消除新陈代谢的昼夜节律以节约能量,从而适应营养匮乏的生态环境^[85, 86]。本研究检测到,近盲虾虎鱼类中许多与昼夜节律相关的基因同样呈现选择性放松。这可能与它们长期适应黑暗洞穴环境、光周期对昼夜节律相关基因选择压力减弱有关。然而,这些基因是否能够通过昼夜节律的消除来节约能量以适应寡营养的洞穴环境,还有待更多实验进一步验证。需要指出的是,选择放松可能为视觉退化、色素退化及昼夜节律丧失等退化性状的形成提供了一种潜在的分子机制;然而,由于缺乏功能层面的实验证据,该关联仍属于推论性结论。因此,这些基因的选择放松对退化性状的实际影响仍需结合更多实验和功能验证加以证明。

4 结论

本研究首次组装和注释了近盲虾虎鱼亚科代表物种拉氏狼牙虾虎鱼染色体级别的基因组,并通过比较基因组学研究探讨了其所在的近盲虾虎鱼亚科谱系物种的系统发育和进化,同时阐明其适应潮间带洞穴生活的可能分子机制。结果揭示了近盲虾虎鱼从水生到陆生进化过程中鲜为人知的特殊起源与进化历程。为了应对潮间带穴居生活过程中黑暗、缺氧及寡营养等环境挑战,近盲虾虎鱼类在嗅觉感知、血管生成及糖类胰岛素代谢等方面的基因发生明显的重构。同时,一些与视觉、色素沉着和昼夜节律等退化性状相关的基因也表现出选择放松,这可能在近盲虾虎鱼亚科类群的视觉退化、色素沉积减弱、昼夜节律丧失等方面发挥了重要作用。本研究不仅丰富了对近盲虾虎鱼亚科鱼类独特起源与系统发育的认知,也为理解潮间带穴居硬骨鱼类适应其独特生活环境的基因组策略与分子机制提供了新见解。

(作者声明本文符合出版伦理要求)

参考文献:

- [1] You X, Bian C, Zan Q, *et al.* Mudskipper genomes provide insights into the terrestrial adaptation of amphibious fishes [J]. *Nature Communications*, 2014(5): 5594.
- [2] Meyer A, Schloissnig S, Franchini P, *et al.* Giant lungfish genome elucidates the conquest of land by vertebrates [J]. *Nature*, 2021, **590**(7845): 284-289.
- [3] Wang K, Wang J, Zhu C, *et al.* African lungfish genome sheds light on the vertebrate water-to-land transition [J]. *Cell*, 2021, **184**(5): 1362-1376.
- [4] Long J A, Gordon M S. The greatest step in vertebrate history: a paleobiological review of the fish-tetrapod transition [J]. *Physiological and Biochemical Zoology*, 2004, **77**(5): 700-719.
- [5] Martin K L M. Time and tide wait for no fish: intertidal fishes out of water [J]. *Environmental Biology of Fishes*, 1995, **44**(1): 165-181.
- [6] Ord T J, Cooke G M. Repeated evolution of amphibious behavior in fish and its implications for the colonization of novel environments [J]. *Evolution; International Journal of Organic Evolution*, 2016, **70**(8): 1747-1759.
- [7] Murdy E O, Shibukawa K. A revision of the gobiid fish genus *Odontamblyopus* (Gobiidae: Amblyopinae) [J]. *Ichthyological Research*, 2001, **48**(1): 31-43.
- [8] Fang J Q. The phylogeny of amblyopinae fishes in coastal waters of China and a preliminary study on the molecular phylogeography of three amblyopinae species [D]. Zhoushan: Zhejiang Ocean University, 2021: 1-210. [方嘉琪. 中国沿海近盲虾虎鱼亚科鱼类的系统发育及三个种类的分子系统地理学初步研究 [D]. 舟山: 浙江海洋大学, 2021: 1-210.]
- [9] Maeda K, Hanahara N, Uehara M, *et al.* Larval study revealed diversity and life-history traits of crypto-benthic eel gobies [J]. *Journal of Fish Biology*, 2022, **101**(6): 1411-1427.
- [10] Yao C, Zhao S, Chen J, *et al.* Molecular and morphological analyses suggest cryptic diversity of eel gobies, genus *Taenioides* (Gobiidae), in coastal waters of China [J]. *Journal of Ichthyology*, 2022, **62**(6): 1025-1033.
- [11] Murdy E. Systematics of Amblyopinae [M]. *The Biology of Gobies*, Boca Raton: CRC Press, 2011: 107-118.
- [12] Gonzales T T, Katoh M, Ishimatsu A. Air breathing of aquatic burrow-dwelling eel goby, *Odontamblyopus lacepedii* (Gobiidae: Amblyopinae) [J]. *The Journal of Experimental Biology*, 2006, **209**(Pt 6): 1085-1092.
- [13] Gonzales T T, Katoh M, Ishimatsu A. Intertidal burrows of the air-breathing eel goby, *Odontamblyopus lacepedii* (Gobiidae: Amblyopinae) [J]. *Ichthyological Research*, 2008(55): 303-306.
- [14] Murdy E O. A taxonomic revision and cladistic analysis of the oxudercine gobies (Gobiidae: Oxudercinae) [J]. *Records of the Australian Museum, Supplement*, 1989(11): 1-93.
- [15] Thacker C E. Molecular phylogeny of the gobioid fishes (Teleostei: Perciformes: Gobioidi) [J]. *Molecular Phylogenetics and Evolution*, 2003, **26**(3): 354-368.
- [16] Xu H, Fang C, Xu W, *et al.* Evolution and genetic adaptation of fishes to the deep sea [J]. *Cell*, 2025, **188**(5):

- 1393-1408.
- [17] Tong C, Li M, Tang Y, *et al.* Genomic signature of shifts in selection and alkaline adaptation in highland fish [J]. *Genome Biology and Evolution*, 2021, **13**(5): evab086.
 - [18] De Coster W, D'hert S, Schultz D T, *et al.* NanoPack: visualizing and processing long-read sequencing data [J]. *Bioinformatics*, 2018, **34**(15): 2666-2669.
 - [19] Chen S, Zhou Y, Chen Y, *et al.* Fastp: an ultra-fast all-in-one FASTQ preprocessor [J]. *Bioinformatics*, 2018, **34**(17): i884-i890.
 - [20] Andrews S. *FastQC: A Quality Control Tool for High Throughput Sequence Data* [Z]. Babraham Bioinformatics, Babraham Institute, Cambridge, United Kingdom, 2010.
 - [21] Liu B, Shi Y, Yuan J, *et al.* Estimation of genomic characteristics by analyzing k-mer frequency in de novo genome projects [J]. *ArXiv: Genomics*, 2013.
 - [22] Hu J, Wang Z, Sun Z, *et al.* An efficient error correction and accurate assembly tool for noisy long reads [J]. *bioRxiv*, 2023: 2023.03. 09.531669.
 - [23] Hu J, Fan J, Sun Z, *et al.* NextPolish: a fast and efficient genome polishing tool for long-read assembly [J]. *Bioinformatics*, 2020, **36**(7): 2253-2255.
 - [24] Roach M J, Schmidt S A, Borneman A R. Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies [J]. *BMC Bioinformatics*, 2018, **19**(1): 460.
 - [25] Durand N C, Shamim M S, Machol I, *et al.* Juicer provides a one-click system for analyzing loop-resolution hi-C experiments [J]. *Cell Systems*, 2016, **3**(1): 95-98.
 - [26] Dudchenko O, Batra S S, Omer A D, *et al.* De novo assembly of the *Aedes aegypti* genome using Hi-C yields chromosome-length scaffolds [J]. *Science*, 2017, **356**(6333): 92-95.
 - [27] Durand N C, Robinson J T, Shamim M S, *et al.* Juicebox provides a visualization system for hi-C contact maps with unlimited zoom [J]. *Cell Systems*, 2016, **3**(1): 99-101.
 - [28] Simão F A, Waterhouse R M, Ioannidis P, *et al.* BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs [J]. *Bioinformatics*, 2015, **31**(19): 3210-3212.
 - [29] Li H. Minimap and minimap: fast mapping and de novo assembly for noisy long sequences [J]. *Bioinformatics*, 2016, **32**(14): 2103-2110.
 - [30] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform [J]. *Bioinformatics*, 2009, **25**(14): 1754-1760.
 - [31] Benson G. Tandem repeats finder: a program to analyze DNA sequences [J]. *Nucleic Acids Research*, 1999, **27**(2): 573-580.
 - [32] Chen N. Using RepeatMasker to identify repetitive elements in genomic sequences [J]. *Current Protocols in Bioinformatics*, 2004(4): Unit4.10.
 - [33] Stanke M, Keller O, Gunduz I, *et al.* AUGUSTUS: Ab initio prediction of alternative transcripts [J]. *Nucleic Acids Research*, 2006, **34**(Web Server issue): W435-W439.
 - [34] Birney E, Clamp M, Durbin R. GeneWise and genome-wise [J]. *Genome Research*, 2004, **14**(5): 988-995.
 - [35] Pertea M, Pertea G M, Antonescu C M, *et al.* StringTie enables improved reconstruction of a transcriptome from RNA-seq reads [J]. *Nature Biotechnology*, 2015, **33**(3): 290-295.
 - [36] Sirén J, Välimäki N, Mäkinen V. HISAT2: Fast and sensitive alignment against general human population [J]. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2014, **11**(2): 375-388.
 - [37] Haas B J, Delcher A L, Mount S M, *et al.* Improving the Arabidopsis genome annotation using maximal transcript alignment assemblies [J]. *Nucleic Acids Research*, 2003, **31**(19): 5654-5666.
 - [38] Holt C, Yandell M. MAKER2: an annotation pipeline and genome-database management tool for second-generation genome projects [J]. *BMC Bioinformatics*, 2011(12): 491.
 - [39] Madden T. The BLAST Sequence Analysis Tool [M]. The NCBI Handbook. 2nd Edition. Bethesda: National Center for Biotechnology Information, 2013: 425-436.
 - [40] Emms D M, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics [J]. *Genome Biology*, 2019, **20**(1): 238.
 - [41] Katoh K, Standley D M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability [J]. *Molecular Biology and Evolution*, 2013, **30**(4): 772-780.
 - [42] Castresana J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis [J]. *Molecular Biology and Evolution*, 2000, **17**(4): 540-552.
 - [43] Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies [J]. *Bioinformatics*, 2014, **30**(9): 1312-1313.
 - [44] Yang Z, Rannala B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds [J]. *Molecular Biology and Evolution*, 2006, **23**(1): 212-226.
 - [45] Kumar S, Stecher G, Suleski M, *et al.* TimeTree: a resource for timelines, timetrees, and divergence times [J]. *Molecular Biology and Evolution*, 2017(34): 1812-1819.
 - [46] De Bie T, Cristianini N, Demuth J P, *et al.* CAFE a computational tool for the study of gene family evolution [J]. *Bioinformatics*, 2006, **22**(10): 1269-1271.
 - [47] Yang Z. PAML 4: phylogenetic analysis by maximum likelihood [J]. *Molecular Biology and Evolution*, 2007, **24**(8): 1586-1591.
 - [48] Wertheim J O, Murrell B, Smith M D, *et al.* RELAX: detecting relaxed selection in a phylogenetic framework [J]. *Molecular Biology and Evolution*, 2015, **32**(3): 820-832.
 - [49] Lü Z, Liu Y, Zhao S, *et al.* Amblyopinae mitogenomes

- provide novel insights into the paraphyletic origin of their adaptation to mudflat habitats [J]. *International Journal of Molecular Sciences*, 2023, **24**(5): 4362.
- [50] Steppan S J, Meyer A A, Barrow L N, *et al.* Phylogenetics and the evolution of terrestriality in mudskippers (Gobiidae: Oxudercinae) [J]. *Molecular Phylogenetics and Evolution*, 2022(169): 107416.
- [51] Haq B U, Hardenbol J, Vail P R. Chronology of fluctuating sea levels since the Triassic [J]. *Science*, 1987, **235**(4793): 1156-1167.
- [52] Hochmuth K, Whittaker J M, Sauermilch I, *et al.* Southern Ocean biogenic blooms freezing-in Oligocene colder climates [J]. *Nature Communications*, 2022(13): 6785.
- [53] Broderick R, Nasheuer H P. Regulation of Cdc45 in the cell cycle and after DNA damage [J]. *Biochemical Society Transactions*, 2009, **37**(Pt 4): 926-930.
- [54] Wu R, Tan Q, Niu K, *et al.* MMS19 localizes to mitochondria and protects the mitochondrial genome from oxidative damage [J]. *Biochemistry and Cell Biology*, 2018, **96**(1): 44-49.
- [55] Tan Q, Niu K, Zhu Y, *et al.* RNF8 ubiquitinates RecQL4 and promotes its dissociation from DNA double strand breaks [J]. *Oncogenesis*, 2021(10): 24.
- [56] Hanchapola H A C R, Kim G, Liyanage D S, *et al.* Molecular features, antiviral activity, and immunological expression assessment of interferon-related developmental regulator 1 (IFRD1) in red-spotted grouper (*Epinephelus akaara*) [J]. *Fish & Shellfish Immunology*, 2024(153): 109859.
- [57] Li Z P, Jiang X X, Tang Z R, *et al.* Molecular characterization and expression analysis of IL-21 and its four receptor genes in yellow catfish (*Pelteobagrus fulvidraco*) in response to *Edwardsiella ictaluri* infection [J]. *Fish & Shellfish Immunology*, 2025(163): 110371.
- [58] Kubota Y, Oike Y, Satoh S, *et al.* Cooperative interaction of Angiopoietin-like proteins 1 and 2 in zebrafish vascular development [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2005, **102**(38): 13502-13507.
- [59] Yu P C, Gu S Y, Bu J W, *et al.* TRPC1 is essential for in vivo angiogenesis in zebrafish [J]. *Circulation Research*, 2010, **106**(7): 1221-1232.
- [60] Ober E A, Olofsson B, Mäkinen T, *et al.* Vegfc is required for vascular development and endoderm morphogenesis in zebrafish [J]. *EMBO Reports*, 2004, **5**(1): 78-84.
- [61] Chen S, Villalta S A, Agrawal D K. FOXO1 mediates vitamin D deficiency-induced insulin resistance in skeletal muscle [J]. *Journal of Bone and Mineral Research*, 2016, **31**(3): 585-595.
- [62] He H, Holl K, DeBehnke S, *et al.* Tpcn2 knockout mice have improved insulin sensitivity and are protected against high-fat diet-induced weight gain [J]. *Physiological Genomics*, 2018, **50**(8): 605-614.
- [63] Deng Z, Song C, Chen L, *et al.* Inhibition of CILP2 improves glucose metabolism and mitochondrial dysfunction in sarcopenia via the Wnt signalling pathway [J]. *Journal of Cachexia, Sarcopenia and Muscle*, 2024, **15**(6): 2544-2558.
- [64] Hinaux H, Devos L, Blin M, *et al.* Sensory evolution in blind cavefish is driven by early embryonic events during gastrulation and neurulation [J]. *Development*, 2016, **143**(23): 4521-4532.
- [65] Specia D J, Lin D M, Sorensen P W, *et al.* Functional identification of a goldfish odorant receptor [J]. *Neuron*, 1999, **23**(3): 487-498.
- [66] Koide T, Miyasaka N, Morimoto K, *et al.* Olfactory neural circuitry for attraction to amino acids revealed by transposon-mediated gene trap approach in zebrafish [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2009, **106**(24): 9884-9889.
- [67] Gonzales T T, Katoh M, Ishimatsu A. Respiratory vasculatures of the intertidal air-breathing eel goby, *Odontamblyopus lacepedii* (Gobiidae: Amblyopinae) [J]. *Environmental Biology of Fishes*, 2008, **82**(4): 341-351.
- [68] Riddle M R, Aspiras A C, Gaudenz K, *et al.* Insulin resistance in cavefish as an adaptation to a nutrient-limited environment [J]. *Nature*, 2018, **555**(7698): 647-651.
- [69] Zhang W, Patil S, Chauhan B, *et al.* FoxO1 regulates multiple metabolic pathways in the liver: effects on gluconeogenic, glycolytic, and lipogenic gene expression [J]. *The Journal of Biological Chemistry*, 2006, **281**(15): 10105-10117.
- [70] Park S, Cha H N, Shin M G, *et al.* Inhibitory regulation of FOXO1 in PPAR δ expression drives mitochondrial dysfunction and insulin resistance [J]. *Diabetes*, 2024, **73**(7): 1084-1098.
- [71] Lahti D C, Johnson N A, Ajie B C, *et al.* Relaxed selection in the wild [J]. *Trends in Ecology & Evolution*, 2009, **24**(9): 487-496.
- [72] Albalat R, Cañestro C. Evolution by gene loss [J]. *Nature Reviews Genetics*, 2016, **17**(7): 379-391.
- [73] Policarpo M, Fumey J, Lafargeas P, *et al.* Contrasting gene decay in subterranean vertebrates: insights from cavefishes and fossorial mammals [J]. *Molecular Biology and Evolution*, 2021, **38**(2): 589-605.
- [74] Omori Y, Shun K, Kon T, *et al.* Samd7 is a cell type-specific PRC1 component essential for establishing retinal rod photoreceptor identity [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2017, **114**(39): E8264-E8273.
- [75] Bauwens M, Celik E, Zur D, *et al.* Mutations in SAMD7 cause autosomal-recessive macular dystrophy with or without cone dysfunction [J]. *American Journal of Human Genetics*, 2024, **111**(2): 393-402.
- [76] Volkov L I, Ogawa Y, Somjee R, *et al.* Samd7 represses short-wavelength cone genes to preserve long-wavelength cone and rod photoreceptor identity [J]. *Proceedings of the National Academy of Sciences of the United*

- States of America*, 2024, **121**(47): e2402121121.
- [77] Xing Y, Boswell W, Parker J, *et al.* A recessive *oca2* mutation underlies albinism in *Xiphophorus* fish [EB/OL]. *bioRxiv*, 2025: 2025.01.20.633999.
- [78] Kratochwil C F, Urban S, Meyer A. Genome of the Malawi golden cichlid fish (*Melanochromis auratus*) reveals exon loss of *oca2* in an amelanistic morph [J]. *Pigment Cell & Melanoma Research*, 2019, **32**(5): 719-723.
- [79] O'Gorman M, Thakur S, Imrie G, *et al.* Pleiotropic function of the *oca2* gene underlies the evolution of sleep loss and albinism in cavefish [J]. *Current Biology*, 2021, **31**(16): 3694-3701.
- [80] Klaassen H, Wang Y, Adamski K, *et al.* CRISPR mutagenesis confirms the role of *oca2* in melanin pigmentation in *Astyanax mexicanus* [J]. *Developmental Biology*, 2018, **441**(2): 313-318.
- [81] Jeffery W R. Regressive evolution in *Astyanax* cavefish [J]. *Annual Review of Genetics*, 2009(43): 25-47.
- [82] Cai Y D, Chiu J C. Timeless in animal circadian clocks and beyond [J]. *The FEBS Journal*, 2022, **289**(21): 6559-6575.
- [83] Kotwica-Rolinska J, Chodáková L, Smýkal V, *et al.* Loss of Timeless underlies an evolutionary transition within the circadian clock [J]. *Molecular Biology and Evolution*, 2022, **39**(1): msab346.
- [84] Kurien P, Hsu P K, Leon J, *et al.* TIMELESS mutation alters phase responsiveness and causes advanced sleep phase [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2019, **116**(24): 12045-12053.
- [85] Moran D, Softley R, Warrant E J. Eyeless Mexican cavefish save energy by eliminating the circadian rhythm in metabolism [J]. *PLoS One*, 2014, **9**(9): e107877.
- [86] Olsen L, Krishnan J, Banks C, *et al.* Circadian rhythm disruption linked to skeletal muscle dysfunction in the Mexican Cavefish [J]. *Current Biology*, 2023, **33**(7): R255-R256.

CHROMOSOME-LEVEL GENOME ASSEMBLY AND CAVE ADAPTATION MECHANISM ANALYSIS OF *ODONTAMBLYOPUS LACEPEDII*

XU An, LIU Yan-Tao, HUANG Kun, YU Jing, YANG Yi-Jing, ZHU Deng-Hui, LI Feng-Hui, LIU Jing, GONG Li, LIU Li-Qin and LÜ Zhen-Ming

(National Engineering Laboratory of Marine Germplasm Resources Exploration and Utilization, College of Marine Science and Technology College, Zhejiang Ocean University, Zhoushan 316022, China)

Abstract: The subfamily Amblyopinae of fish is one of the few marine bony fish that possess amphibious characteristics, primarily inhabiting intertidal mudflats. Compared with other amphibious fish with strong terrestrial capabilities, this group adopts a more passive terrestrial adaptation strategy, avoiding common terrestrial challenges by residing in mudflat burrows. However, the genetic mechanisms underlying its adaptation to mudflat burrowing remain unclear at present. This study conducted chromosome-level genome assembly and annotation of the *Odontamblyopus lacepedii*, a species of the subfamily Amblyopinae, using multi-platform sequencing data. Combined with comparative genomics analysis, it explored the phylogenetic relationships of the subfamily Amblyopinae and the molecular mechanisms underlying their adaptation to intertidal burrowing life. The results indicated that the genome size of *O. lacepedii* was 963.28 Mb, anchored to 23 chromosomes, with a total of 23558 protein-coding genes annotated. Phylogenetic analysis indicates that the subfamily Amblyopinae does not form an independent monophyletic group but is nested within the subfamily Oxudercinae, supporting a paraphyletic relationship and justifying the taxonomic viewpoint to merge the two groups. In comparison with four Oxudercinae species and seven additional outgroup identified 24 expanded and 243 contracted gene families, along with 163 positively selected genes in the Amblyopinae lineage. These genes were mainly enriched in signaling pathways related to olfactory perception, angiogenesis, and lipid/carbohydrate metabolism, which associated with their adaptation to the dark, hypoxic, and nutrient-poor cave environment in the mudflat. In addition, we identified 552 relaxation selection genes, a considerable portion of which are related to vision, pigmentation, and circadian rhythms, suggesting their role in the degeneration of traits such as visual degeneration, weakened body coloration, and loss of circadian rhythms in Amblyopinae. This study provides new insights into the environmental adaptation and key trait evolution mechanisms of the Amblyopinae and other intertidal burrowing fish.

Key words: Genome assembly; Comparative genomics; Cave adaptation; *Odontamblyopus lacepedii*