

# 膨腹海马生长性状全基因组关联

## 分析及候选基因的挖掘

郑磊<sup>1,2,3</sup>, 梁建韬<sup>4</sup>, 左小玲<sup>4</sup>, 韦丽云<sup>4</sup>, 闫旭<sup>4</sup>, 卢柯宇<sup>4</sup>,  
张敏琳<sup>4</sup>, 梁凯珊<sup>4</sup>, 涂传灯<sup>1,2,3</sup>, 何丽斌<sup>5</sup>, 郑乐云<sup>1,2,3</sup>, 王庆<sup>4</sup>

(1. 厦门海洋职业技术学院海洋生物学院, 福建 厦门 361102; 2. 海洋资源保护与生态治理福建省高等学校应用技术工程中心, 福建 厦门 361102; 3. 厦门市智慧渔业重点实验室, 福建 厦门 361102; 4. 华南农业大学海洋学院, 广东 广州 510642; 5. 福建省水产研究所, 福建 厦门 361012)

**[摘要]** 传统的膨腹海马 (*Hippocampus abdominalis*) 生长性状的选育方法研究进展较慢, 育种周期长, 结合分子育种技术可加快育种进程。但是目前关于膨腹海马生长性状相关的分子标记和基因还不清楚。采集相同条件下极端快速生长 (79 尾) 和极端缓慢生长 (75 尾) 的膨腹海马, 通过基因组重测序检测到 4 172 760 个单核苷酸多态性 (single nucleotide polymorphism, SNP) 位点, 并通过全基因组关联分析 (genome-wide association study, GWAS) 鉴定出 27 个与生长性状关联的 SNP 位点, 它们主要分布于 6 号染色体上, 显著关联 SNP 对性状的解释贡献率合计达 49.53%。此外, 通过挖掘 SNP 位点上下游 50 kb 区域, 找到 42 个潜在影响生长的候选基因, 基因本体 (gene ontology, GO)、京都基因与基因组百科全书 (Kyoto encyclopedia of genes and genomes, KEGG) 富集分析显示, 这些基因主要富集在碳水化合物消化吸收、矿物质吸收和细胞连接等相关通路。进一步分析基因在快速生长个体和缓慢生长个体中的差异表达情况, 发现 15 个基因可能与膨腹海马生长性状相关。

**[关键词]** 膨腹海马; 生长性状; 全基因组关联分析; 单核苷酸多态性; 候选基因

**[中图分类号]** S 917.4

## Gene-Wide Association Study to Identify Candidate Genes for Growth Traits in *Hippocampus abdominalis*

ZHENG Lei<sup>1,2,3</sup>, LIANG Jiantao<sup>4</sup>, ZUO Xiaoling<sup>4</sup>, WEI Liyun<sup>4</sup>, YAN Xu<sup>4</sup>, LU Keyu<sup>4</sup>, ZHANG Minlin<sup>4</sup>,  
LIANG Kaishan<sup>4</sup>, TU Chuandeng<sup>1,2,3</sup>, HE Libin<sup>5</sup>, ZHENG Leyun<sup>1,2,3</sup>, WANG Qing<sup>4</sup>

(1. School of Marine Biology, Xiamen Ocean Vocational College, Xiamen 361102, China; 2. Applied Technology Engineering Center of Fujian Provincial Higher Education for Marine Resource Protection and Ecological Governance, Xiamen 361102, China; 3. Xiamen Key Laboratory of Intelligent Fishery, Xiamen 361102, China; 4. College of Marine Sciences, South China Agricultural University, Guangzhou 510642, China; 5. Fisheries Research Institute of Fujian, Xiamen 361012, China)

**Abstract:** The traditional breeding methods for growth traits in *Hippocampus abdominalis* exhibit slow pro-

**[收稿日期]** 2025-02-07

**[基金项目]** 国家自然科学基金项目 (32170415); 福建省科技计划项目 (2024S01020061)

**[通信作者]** 王庆 (1988—), 博士, 副教授, 从事鱼类分子遗传育种研究, E-mail: wangqing@scau.edu.cn; 郑乐云 (1969—), 教授, 从事海水鱼类养殖及遗传育种研究, E-mail: zhengleyun@xmoc.edu.cn

gress and a prolonged breeding cycle. The integration of molecular breeding techniques can expedite the breeding process. However, the molecular markers and genes associated with growth traits in *Hippocampus abdominalis* remain unclear. Here, big-belly seahorse individuals exhibiting extremely fast-growing (79 individuals) and extremely slow-growing (75 individuals) were collected under identical rearing conditions. A total of 4 172 760 single nucleotide polymorphism (SNP) sites were identified by genome resequencing. Using genome-wide association study (GWAS), 27 growth-related SNPs were detected, and predominantly located on chromosome 6. These significantly associated SNPs collectively explained 49.53% of the total phenotypic variation. Furthermore, 42 candidate genes potentially influencing growth were identified within 50 kb upstream and downstream regions surrounding SNP loci. Gene ontology (GO) and Kyoto encyclopedia of genes and genomes (KEGG) enrichment analysis revealed that these genes were primarily enriched in pathways related to carbohydrate digestion and absorption, mineral absorption, and cell junctions. Further investigation into the differential expression of genes between fast-growing and slow-growing individuals indicated that 15 genes might be associated with growth traits in *Hippocampus abdominalis*.

**Keywords:** *Hippocampus abdominalis*; growth trait; genome-wide association study; single nucleotide polymorphism; candidate genes

## 0 引言

海马 (*Hippocampus* spp.) 属于海龙科小型硬骨鱼类, 因其外观似马所以命名为“海马”, 是独特的海洋鱼类之一。膨腹海马 (*Hippocampus abdominalis*) 又称大肚海马, 为辐鳍亚纲 (Actinopterygii) 棘背鱼目 (Gasterosteiformes) 海龙科 (Syngnathus) 海马属 (*Hippocampus*), 主要分布在南太平洋区的澳洲及新西兰的热带、亚热带海域中<sup>[1]</sup>, 2017 年被列入《世界自然保护联盟濒危物种红色名录》<sup>[2]</sup>。我国每年对于人药干品海马的需求量超过 200 t, 大部分来源于天然海域的自然捕捞, 这会造成天然海域海马资源的极度匮乏<sup>[3]</sup>。为了保护海马资源, 其人工繁育和养殖成为了唯一的选择<sup>[4]</sup>。2020 年我国膨腹海马规模化人工繁育技术取得突破<sup>[4]</sup>, 2021 年苗种产量达 300 万尾, 2024 年国内养殖产量达 3 000 万尾左右, 膨腹海马养殖产业发展迅猛, 现已受到福建、山东等省沿海养殖业者的热捧, 成为我国重要的海水养殖特色品种。在养殖过程中, 膨腹海马生长速度参差不齐, 而海马价格与体质量息息相关, 大规格的海马价格往往比小规格的海马高 1 倍以上, 因此开展膨腹海马生长性状的选育迫在眉睫。

全基因组关联分析 (genome-wide association study, GWAS) 是从全基因组范围内开发优良性状分子标记和性状相关基因挖掘的有效方法<sup>[5]</sup>, GWAS 在生物育种中已成为解析复杂性状遗传基础的关键技术。在作物和畜禽育种中, 应用 GWAS 在全基因组范围内筛选与目标性状显著关联的单核苷酸多态性 (single nucleotide polymorphism, SNP) 为分子标记辅助育种提供精准靶点。例如, 在水稻育种中, GWAS 成功定位到控制产量和抗逆性的关键基因区域, 如 *OsSPL14* 基因通过调控分蘖数显著提高水稻产量<sup>[6]</sup>; 在猪育种中, GWAS 揭示了与生长速度、瘦肉率相关的 SNP 标记, 并发现 *IGF2* 基因的 3'UTR 区域突变对肌肉发育的调控作用<sup>[7]</sup>; 在奶牛基因组选择中, 基于 GWAS 筛选的高遗传潜力个体通过 50 000 SNP 芯片评估, 使牧场后备牛选育效率提升 40% 以上<sup>[8]</sup>。相较于作物与畜禽, 基于全基因组分析的鱼类育种技术的研究和应用较为滞后, 但是 GWAS 在鱼类育种中已成为解析复杂性状遗传基础的核心工具。例如, 采用 GWAS 挖掘到了大黄鱼 (*Larimichthys crocea*)<sup>[9]</sup> 和黄河鲤 (*Cyprinus carpio*)<sup>[10]</sup> 多个经济性状相关的 SNP 标记和基因, 成功定位到斑鳢 (*Channa maculata*)<sup>[11]</sup>、鲢 (*Hypophthalmichthys molitrix*)<sup>[12]</sup> 与生长性状相关的 SNP 标记及候选基因。这些研究表明, GWAS 是挖掘鱼类重要经济性状标记和基因的有效方法。膨腹海马生长性状为微效多基因性状, 遗传结构复杂, 采用传统选育方法周期长, 选育进展缓慢, 如能结合分子标记开展全基因组预测可以提高选育准确性, 加

快育种进程。但是目前关于膨腹海马生长性状相关的分子标记和基因还不清楚,难以开展分子选育工作。因此,本研究应用 GWAS 开展膨腹海马生长相关分子标记和基因的挖掘,筛选与膨腹海马生长性状相关的 SNP 位点及候选基因,为膨腹海马快速生长品种的选育提供基因资源和理论基础。

# 1 材料与方法

## 1.1 实验材料

实验材料于 2020 年采自厦门小嶝水产科技有限公司。选取海马亲本群体池中同一天出生的海马 8 100 尾,放在同一个培育池中进行养殖,培育池容量为 30 t,养殖周期从出生到 110 d,到 110 d 时测量池中每一尾海马的体质量。选择生长速度最快(fast growth, FG)的 79 个个体和生长速度最慢(slow growth, SG)的 75 个个体,共计 154 个样本,FG、SG 的初始体质量均为  $(0.01 \pm 0.002)$  g,末期平均体质量分别为  $(2.45 \pm 0.36)$ 、 $(1.70 \pm 0.31)$  g,分别取其背部肌肉适量,于  $-80\text{ }^{\circ}\text{C}$  保存备用。

## 1.2 DNA 提取、基因组测序和测序数据的初步处理

采用基因组 DNA 提取试剂盒(北京金沙生物技术有限公司)提取每个样品的 DNA,并用质量分数为 1% 的琼脂糖凝胶电泳检测 DNA 的完整性,使用 Nanodrop 2000 型超微量分光光度计测定 DNA 的浓度,通过  $A_{260}/A_{280}$  来评估 DNA 的质量。样品经过凝胶电泳检测,条带清晰未拖尾、 $A_{260}/A_{280} = 1.8 \sim 2.0$ ,并且分光光度计测定浓度大于 100 mg/L,然后进行重测序用于 SNP 分型。基因组 DNA 用酶随机切成短的 DNA 片段后进行平末端修复,然后在 DNA 片段两端连接 dA 尾,并连接测序接头。加上接头的 DNA 片段经过 AMPure XP 磁珠进行纯化,并选择 300 ~ 400 bp 范围的片段进行 PCR 扩增。建好的文库经过纯化、库检,HiSeq X10 PE150 上机测序,测序由基迪奥生物科技有限公司完成。测序原始数据上传到国家生物信息中心,编号为 CRA023810。

因为稀有等位、高缺失率、高杂合率等位点会引起群体分析和全基因组关联分析异常(线性模型对极端情况的处理能力较弱),这可能会使软件给出错误的结果,所以对原始的标记位点使用 perl 脚本结合 plink2 软件<sup>[13]</sup>按照以下条件进行过滤:1)去除非二等位位点;2)去除第二等位基因频率(minor allele frequency, MAF)小于 0.05 的位点;3)去除缺失率大于 0.5 的位点;4)去除杂合比例大于 0.8 的位点。

## 1.3 群体结构与连锁不平衡分析

基于多态性 SNP 标记,利用 MEGA-X 软件采用邻接法(neighbor-joining, NJ)对 164 份膨腹海马材料进行聚类分析。使用 plink2<sup>[13]</sup>按照 100 kb 的步长、10 nt 的窗口,去除  $r^2 > 0.2$  的标记对中的一个(物理位置靠后的位点被去除)。基于连续不平衡分析(linkage disequilibrium, LD)过滤后,剩余的 SNP 标记使用 Admixture 软件(v1.3)进行群体结构的推断,分别假设样品的分群数( $K$  值)为 1 ~ 9,然后进行聚类。根据交叉验证(cross-validation, CV)误差来确定最优分群数,交叉验证错误率最小的  $K$  值对应最优的分群数。使用 pophelper 软件(v2.2.7)分析每个样本在各个亚群的遗传组成。筛选后的标记使用 gcta 软件(v1.93.2)进行亲缘关系分析,获得两两样本间的亲缘关系矩阵。使用 PopldDecay 软件(v3.41)计算两两标记间的 LD 大小( $r^2$ ),并绘制 LD 随距离增大的变化情况图,LD 强度以  $D'$  或  $r^2$  表示。

## 1.4 全基因组关联分析

使用 gemma 软件(v0.98.1)实现 GWAS 常用模型(MLM(QK)模型)的计算。将 Admixture 最优  $K$  值对应的群体结构矩阵作为相应模型的  $Q$  矩阵,gcta 软件计算的样品间亲缘关系矩阵作为相应模型的  $K$  矩阵,分别绘制曼哈顿图和  $Q$ - $Q$ (quantile-quantile)图。

1.5 候选基因的筛选与功能注释

采用膨腹海马基因组<sup>[14]</sup>对 GWAS 获得的显著 SNP 位点进行注释,扫描范围为候选位点上下游 50 kb,鉴定潜在的候选基因。随后通过查找文献和 NCBI 数据库查询,确定获得的潜在候选基因的功能,进一步筛选候选基因。采用基因本体 (gene ontology, GO)<sup>[15]</sup>功能富集分析并确定与显著差异基因 (与生长性状显著相关 SNP 位点上下 50 kb 内找到的基因) 相关的生物功能;利用京都基因与基因组百科全书 (Kyoto encyclopedia of genes and genomes, KEGG)<sup>[16]</sup>通路富集分析显著差异基因的生物学解释;运用 DAVID(<https://david.ncifcrf.gov/>)在线工具进行 GO 和 KEGG 富集分析;使用 ggpubr、ggplot2、Hmisc 软件包绘制 GO 功能富集分析和 KEGG 通路富集分析气泡图。

1.6 RNA 的提取及实时定量 PCR

采集 FG 组和 SG 组的膨腹海马的肌肉组织于液氮中速冻后,转移到 -80 ℃ 冰箱保存样品,每组 (FG 组和 SG 组) 采集 8 条海马。采用 Trizol Reagent 说明书及操作步骤提取总 RNA,采用 cDNA 合成试剂盒 (TOYOBO, Japan) 反转录成 cDNA 模板,用于实时定量 PCR。实时定量 PCR 反应体系为 20 μL,其中包括 2 × SYBR © Green Realtime PCR Master Mix 试剂 7.5 μL、正向引物 0.3 μL、反向引物 0.3 μL、cDNA 0.6 μL 和 ddH<sub>2</sub>O 6.3 μL,采用 384 孔板的反应模块。将反应体系加入 384 孔板,低速离心后,利用 Roche Light Cycler 480 real time PCR system 进行基因表达的检测。反应条件为:95 ℃ 活化 5 min,95 ℃ 20 s,57 ℃ 20 s,72 ℃ 20 s,共 40 个循环,在 72 ℃ 延伸时进行荧光测定得到 Ct 值,实时荧光定量 PCR 进行 3 个重复。分析溶解曲线,确定 RT-PCR 的质量,用 Excel 表格初步处理数据,然后使用 2<sup>-ΔΔCt</sup> 方法计算 FG 组和 SG 组膨腹海马中的基因表达水平,相对表达水平与内参基因 *ef1α* 表达水平进行比较。实时荧光定量 PCR (real-time quantitative polymerase chain reaction, RT-qPCR) 引物如表 1 所示。

2 实验结果及其分析

2.1 重测序结果

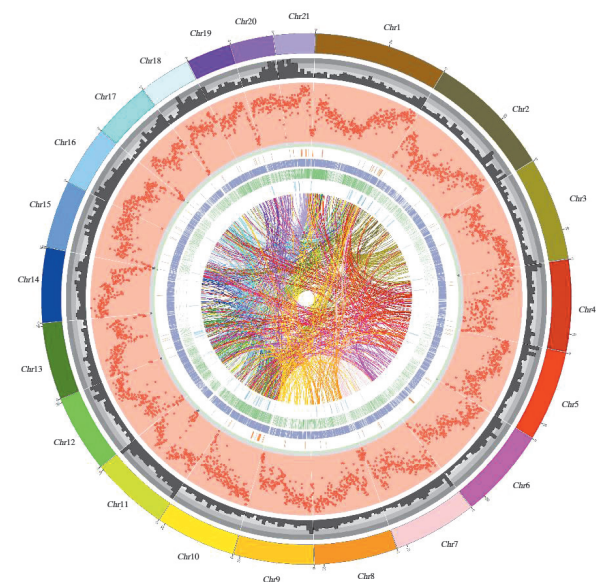
对 154 份膨腹海马基因组 DNA 样本进行了重测序,总体分析结果如图 1 所示。对序列、碱基组成和 reads 进行过滤后,共获得 1 798 036 057 800 个 reads,平均长度为 300 bp,高质量 reads 占比为 99.32%。用过滤后的测序数据与参考基因组进行比对,总共筛选到 5 899 507 个变异位点,包括 5 053 216 个 SNP 位点和 846 291 个 InDel 位点。其中外显子区域 SNP 和 InDel 为 278 998 个,占比为 4.73%,非表达区域 SNP 和 InDel 为 5 619 770 个,占比为 95.27% (见表 2)。

表 1 RT-qPCR 引物

Tab. 1 Sequences of the primers used for RT-qPCR

引物名称	序列(5'-3')
<i>adcy8</i> -F	ATCTACGTCAAAGGCATCAG
<i>adcy8</i> -R	GTTGTGCTTGAGGATCTGCT
<i>arnt</i> -F	GAATGACATCCAACCATCAAG
<i>arnt</i> -R	GTCAGTCCTGAGTAGGCATT
<i>clk2</i> -F	ATCCTGTTTGAGTACTATCTG
<i>clk2</i> -R	CAGCATGCTTTTCGATGAGGTC
<i>cyc</i> -F	ACGAGATTCTGGAATACGAC
<i>cyc</i> -R	TGTCGTTTCATGTAGTACAGG
<i>phf1</i> -F	CATCCTGCTACACGACTTGC
<i>phf1</i> -R	CACGTTGTCCCACCTCCACGA
<i>pkm</i> -F	CAACGACTGCATCATGCTGA
<i>pkm</i> -R	GATAGCATGTGAGCAGACCT
<i>ptpn6</i> -F	CTCAGCTTCCTCACACAAGTC
<i>ptpn6</i> -R	TGGTGATCTGGATGTACTCG
<i>rbp1</i> -F	ACTTCAACATGGACTTTGTGTC
<i>rbp1</i> -R	CTCATCTCCAGATGCAGCTTGT
<i>rgl2</i> -F	GACATGCGCATCATCAGGATC
<i>rgl2</i> -R	TGAATCAGCTCGTAGGATGAC
<i>scmh1</i> -F	CTGCATTACGACAGACGCAC
<i>scmh1</i> -R	GTCGATGTCTTTGATGTAGC
<i>setdb1</i> -F	GAAGAGTCATGTTACATCATCG
<i>setdb1</i> -R	CCACTTCGTAGTTGTAGTCC
<i>foxo3</i> -F	TCTGGCAACGTCCTGCTCAC
<i>foxo3</i> -R	GATAGAGTCCATGTCACACTC
<i>cited3</i> -F	CAGCAGCTCATGGCTAGCAT
<i>cited3</i> -R	CCATGTGCTGCATGTTAGCCA
<i>cyp4b1</i> -F	CTCACAGTCATACGGAGCAA
<i>cyp4b1</i> -R	GCACTTGTTCTGCTGTTCC
<i>cers2</i> -F	CCAAGATGTTCAACTATGCTG
<i>cers2</i> -R	CAATCCGTATGATGAGAATG
<i>ef1α</i> -F	TCGCTTTTGCTGTGCGTGACAT
<i>ef1α</i> -R	GGCAGCCTTCTGTGCAGACTTT





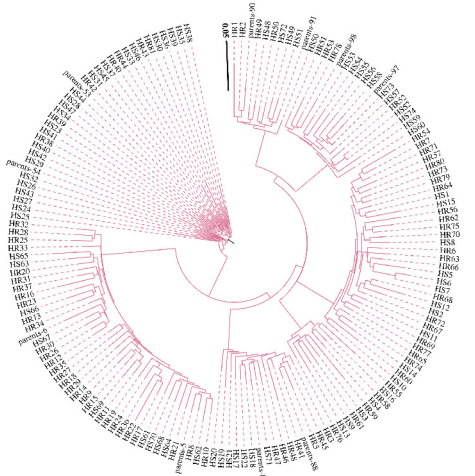
说明: 橙色线表示串联重复(DUP)在染色体上的位置; 紫色线表示大片段缺失(DEL)在染色体上的位置; 绿色线表示插入类型的结构性变异(INS)在染色体上的位置; 蓝色线表示倒置类型的结构性变异(INV)在染色体上的位置; 内圈的连线表示染色体间易位类型的结构性变异(BND)在两两染色体上的位置。

图 1 测序总体分析的结果

Fig.1 Overall analysis results of sequencing

2.2 群体结构及连锁不平衡分析

基于 SNP 标记, 采用 NJ 法对膨腹海马群体进行聚类分析 (见图 2a), 结果表明, 154 份海马材料被分为 7 个类群。同时, 使用 gcta 软件 (v1.93.2) 进行亲缘关系分析, 亲缘关系热图显示, 大部分亲缘值分布在 0~0.5 范围内, 但是出现 7 个连锁群 (见图 2b)。此外, 对 154 份海马材料进行群体结构分析, 发现样品的最优分群数  $K=7$ , 交叉验证误差最小 (见图 2c)。最后开展了连锁不平衡分析, 结果表明, LD 随着物理位置的增加而衰减, 当  $r^2$  衰减至 0.1 时, 整个基因组的 LD 平均衰减距离约为 100 kb (见图 2d)。



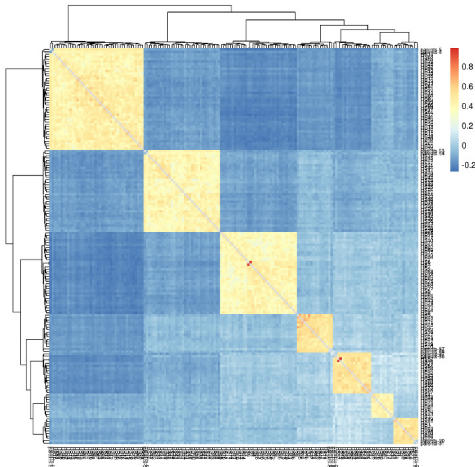
a) 聚类分析图

表 2 SNP-InDel 位置信息统计

Tab. 2 SNP-InDel location information statistics

类别	数量
intronic	2 826 249
intergenic	2 091 074
UTR3	216 878
exonic	278 998
upstream	184 098
downstream	171 527
UTR5	77 628
upstream, downstream	49 558
splicing	1 234
UTR5, UTR3	1 492
Exonic, splicing	32

说明: intronic 为内含子区; intergenic 为基因间区; UTR3 为 3' UTR; exonic 为外显子区; upstream 为转录起始位点上游 1 kb 区域内; downstream 为转录终止位点下游 1 kb 区域内; UTR5 为 5'UTR; splicing 为可变剪切位点 2 bp 以内。



b) 样本间亲缘关系热图

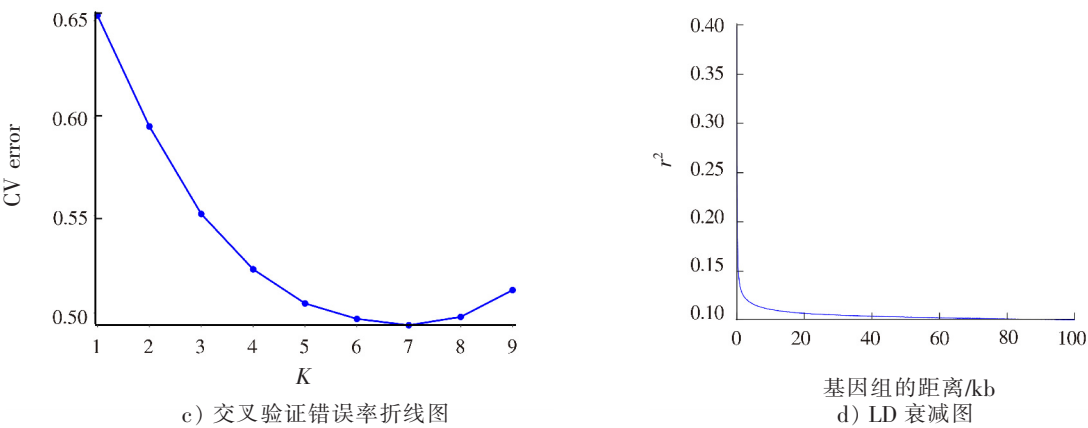


图 2 膨腹海马的群体结构及连锁不平衡分析  
Fig.2 Population structure and LD analysis of *Hippocampus abdominalis*

2.3 GWAS 分析

前面聚类分析和群体结构分析结果显示，154 个样本可能来自 7 个亚群，因此本文采用混合线性模型（mixed linear model，MLM）进行 GWAS 分析。使用 TASSEL 软件对 154 份样本中 SNP 位点进行过滤，最终采用 4 834 844 个 SNP 进行 GWAS 分析，显著关联位点的筛选阈值为  $-\lg P = 6$ 。分析结果显示，有 27 个 SNP 位于阈值线上，显著关联的 SNP 聚集在 6 号染色体上，个别位于 2、3 和 13 号染色体上（见图 3a）。此外， $Q$ - $Q$  图表明 GWAS 的结果是可信的， $r^2$  为 0.129 333 9（见图 3b）。此外，为检测显著关联 SNP 与邻近位点间的连锁不平衡，对位于 3 号染色体和 6 号染色体显著关联区域的 SNP 进行了 LD 区块分析，结果显示，显著关联 SNP 位点及其周边 SNP 位点均未形成单倍型块（见图 3c），表明这两个区域 SNP 位点与其周边位点间不存在连锁不平衡。

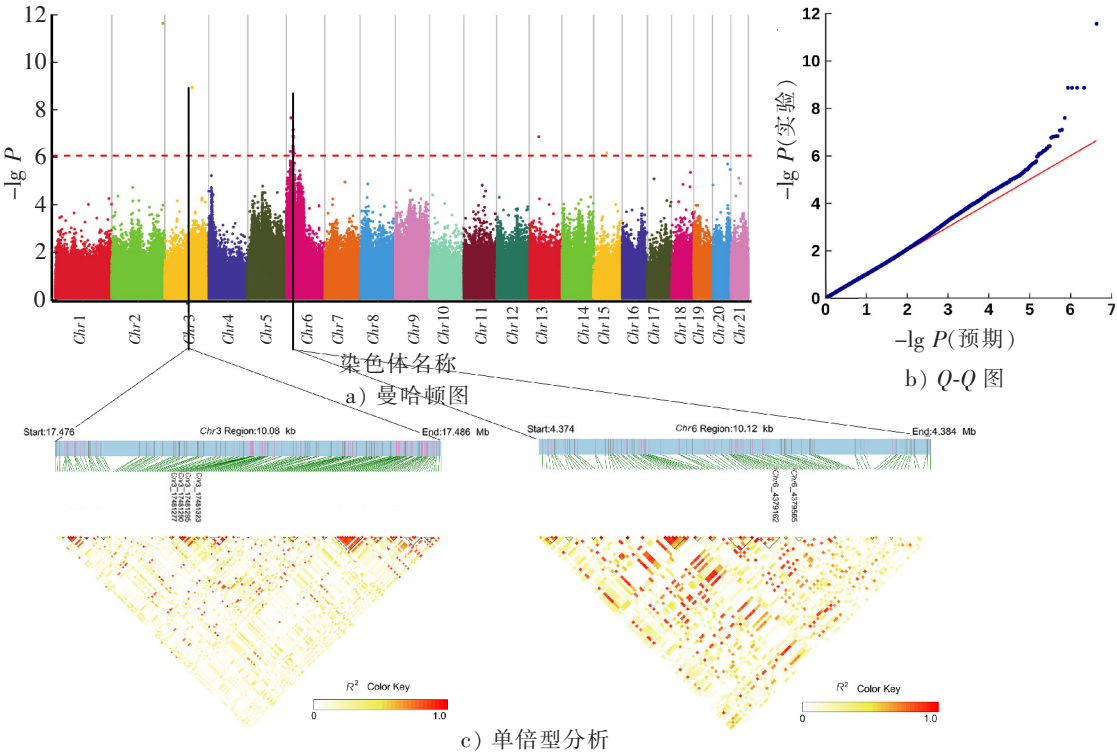


图 3 用 TASSEL 分析的 GWAS 数据曼哈顿图、Q-Q 图和单倍型分析  
Fig.3 Manhattan plot, Q-Q plot of GWAS data generated with TASSEL and haplotype analysis

2.4 显著性 SNP 位点的功能注释和候选基因的筛选

对上述 GWAS 分析获得的 27 个 SNP 进行注释, 筛选显著性关联 SNP 上下游 50 kb 的基因, 初步鉴定到了 42 个潜在的候选基因 (见表 3)。

表 3 膨腹海马染色体区域与 MLM 相关的候选基因

Tab. 3 Candidate genes associated with MLM in the chromosomal region of <i>Hippocampus abdominalis</i>											
基因名称	起始位点	终止位点	参考碱基	等位基因	质量值	位置类型	位置注释	FPR 校正后的 P 值	解释贡献率	最小等位基因频率	
mdgal1	17 481 277	17 481 277	T	C	40 212.57	intergenic	ncbi_109528712	0.000 000 001 345 68	0.436 218 191 093 976	0.393 292 682 926 829	
	17 481 290	17 481 290	C	T	40 629.12	intergenic	ncbi_109528712	0.000 000 001 345 60	0.436 218 191 093 976	0.387 195 121 951 220	
	17 481 295	17 481 295	C	T	41 836.99	intergenic	ncbi_109528712	0.000 000 001 345 68	0.436 218 191 093 976	0.390 243 902 439 024	
	17 481 323	17 481323	T	C	40 523.25	intergenic	ncbi_109528712	0.000 000 001 345 68	0.436 218 191 093 976	0.387 195 121 951 220	
tnfrsf1a	2 403 131	2 403 131	A	C	25 864.38	exonic	ncbi_109515323	0.000 000 664 554 30	0.428 708 457 867 267	0.487 654 320 987 654	
plekhg5	2 403 131	2 403 131	A	C	25 864.38	exonic	ncbi_109515320	0.000 000 664 554 30	0.428 708 457 867 267	0.487 654 320 987 654	
lag-3	2,403 131	2,403 131	A	C	25 864.38	exonic	ncbi_109515321	0.000 000 664 554 30	0.428 708 457 867 267	0.487 654 320 987 654	
mrpl51	2 403 131	2 403 131	A	C	25 864.38	exonic	ncbi_109515325	0.000 000 664 554 30	0.428 708 457 867 267	0.487 654 320 987 654	
vamp1	2 403 131	2 403 131	A	C	25 864.38	exonic	ncbi_110163128	0.000 000 664 554 30	0.428 708 457 867 267	0.487 654 320 987 654	
taphpl	2 403 131	2 403 131	A	C	25 864.38	exonic	ncbi_109515326	0.000 000 664 554 30	0.428 708 457 867 267	0.487 654 320 987 654	
masp2	2 403 131	2 403 131	A	C	25 864.38	exonic	ncbi_109515279	0.000 000 664 554 30	0.428 708 457 867 267	0.487 654 320 987 654	
tnfrsf14	2 403 131	2 403 131	A	C	25 864.38	exonic	ncbi_109515327	0.000 000 664 554 30	0.428 708 457 867 267	0.487 654 320 987 654	
rbp1	2 403 131	2 403 131	A	C	25 864.38	exonic	ncbi_109515328	0.000 000 664 554 30	0.428 708 457 867 267	0.487 654 320 987 654	
ptpn6	2 403 131	2 403 131	A	C	25 864.38	exonic	ncbi_109515280	0.000 000 664 554 30	0.428 708 457 867 267	0.487 654 320 987 654	
chd4	2 403 131	2 403 131	A	C	25 864.38	exonic	ncbi_109515384	0.000 000 6645 5430	0.428 708 457 867 267	0.487 654 320 987 654	
nectin-2	2 989 006	2 989 006	C	G	15 715.13	intronic	ncbi_109531669	0.000 000 025 237 33	0.450 127 311 499 123	0.483 221 476 510 067	
atp1a3	2 989 006	2 989 006	C	G	15 715.13	intronic	ncbi_109531678	0.000 000 025 237 33	0.450 127 311 499 123	0.483 221 476 510 067	
rabac1	2 989 006	2 989 006	C	G	15 715.13	intronic	ncbi_109531690	0.000 000 025 237 33	0.450 127 311 499 123	0.483 221 476 510 067	
cited3	4 293 443	4 293 443	G	T	32 870.00	intronic	ncbi_109522770	0.000 000 610 162 90	0.597 290 462 400 491	0.450 000 000 000 000	
	4 316 410	4 316 410	A	T	34 791.78	intergenic	ncbi_109522770	0.000 000 476 210 90	0.615 829 944 763 220	0.375 000 000 000 000	
	4 316 527	4 316 527	T	C	30 210.31	intergenic	ncbi_109522770	0.000 000 154 641 40	0.581 049 943 665 930	0.388 198 757 763 975	
	4 293 443	4 293 443	G	T	32 870.00	intronic	ncbi_109522793	0.000 000 610 162 90	0.597 290 462 400 491	0.450 000 000 000 000	
scmh1	4 316 410	4 316 410	A	T	34 791.78	intergenic	ncbi_109522793	0.000 000 476 210 90	0.615 829 944 763 220	0.375 000 000 000 000	
	4 316 527	4 316 527	T	C	30 210.31	intergenic	ncbi_109522793	0.000 000 154 641 40	0.581 049 943 665 930	0.388 198 757 763 975	
	4 379 162	4 379 162	T	G	33 198.47	intronic	ncbi_109522803	0.000 000 610 162 90	0.597 290 462 400 491	0.466 049 382 716 049	
	4 379 565	4 379 565	A	G	26 607.00	intronic	ncbi_109522803	0.000 000 149 191 80	0.513 841 026 812 303	0.419 753 086 419 753	
foxo3	4 388 155	4 388 155	G	A	27 187.61	intergenic	ncbi_109522803	0.000 000 084 903 66	0.531 983 845 141 562	0.421 875 000 000 000	
	4 414 086	4 414 086	G	T	22 534.18	intergenic	ncbi_109522803	0.000 000 389 060 80	0.621 821 055 479 312	0.376 543 209 876 543	
	4 414 091	4 414 091	T	G	38 399.28	intergenic	ncbi_109522803	0.000 000 389 060 80	0.621 821 055 479 312	0.380 368 098 159 509	
	4 427 431	4 427 431	G	C	19 258.43	UTR3	ncbi_109522803	0.000 000 145 046 30	0.637 328 749 526 406	0.334 355 828 220 859	
med18	4 379 162	4 379 162	T	G	33 198.47	intronic	ncbi_109522791	0.000 000 610 162 90	0.597 290 462 400 491	0.466 049 382 716 049	
	4 379 565	4 379 565	A	G	26 607.00	intronic	ncbi_109522791	0.000 000 149 191 80	0.513 841 026 812 303	0.419 753 086 419 753	
	4 388 155	4 388 155	G	A	27 187.61	intergenic	ncbi_109522791	0.000 000 084 903 66	0.531 983 845 141 562	0.421 875 000 000 000	
	4 414 086	4 414 086	G	T	22 534.18	intergenic	ncbi_109522791	0.000 000 389 060 80	0.621 821 055 479 312	0.376 543 209 876 543	
cyp4b1	4 414 091	4 414 091	T	G	38 399.28	intergenic	ncbi_109522791	0.000 000 389 060 80	0.621 821 055 479 312	0.380 368 098 159 509	
	4 427 431	4 427 431	G	C	19 258.43	UTR3	ncbi_109522791	0.000 000 145 046 30	0.637 328 749 526 406	0.334 355 828 220 859	
	4 455 704	4 455 704	G	T	29 762.48	intronic	ncbi_109522791	0.000 000 723 924 60	0.472 095 002 974 714	0.453 703 703 703 704	
	4 455 767	4 455 767	G	T	29 372.86	intronic	ncbi_109522791	0.000 000 487 613 20	0.595 896 294 762 575	0.462 732 919 254 658	
zbtb7b	4 457 744	4 457 744	T	C	33 265.98	intronic	ncbi_109522791	0.000 000 825 188 60	0.589 567 117 569 121	0.337 423 312 883 436	
	4 388 155	4 388 155	G	A	27 187.61	intergenic	ncbi_109522790	0.000 000 084 903 66	0.531 983 845 141 562	0.421 875 000 000 000	
	4 414 086	4 414 086	G	T	22 534.18	intergenic	ncbi_109522790	0.000 000 389 060 80	0.621 821 055 479 312	0.376 543 209 876 543	
	4 414 091	4 414 091	T	G	38 399.28	intergenic	ncbi_109522790	0.000 000 389 060 80	0.621 821 055 479 312	0.380 368 098 159 509	
zbtb7b	4 427 431	4 427 431	G	C	19 258.43	UTR3	ncbi_109522790	0.000 000 145 046 30	0.637 328 749 526 406	0.334 355 828 220 859	
	4 455 704	4 455 704	G	T	29 762.48	intronic	ncbi_109522790	0.000 000 723 924 60	0.472 095 002 974 714	0.453 703 703 703 704	
	4 455 767	4 455 767	G	T	29 372.86	intronic	ncbi_109522790	0.000 000 487 613 20	0.595 896 294 762 575	0.462 732 919 254 658	
	4 457 744	4 457 744	T	C	33 265.98	intronic	ncbi_109522790	0.000 000 825 188 60	0.589 567 117 569 121	0.337 423 312 883 436	

续表 3  
Continued table 3

基因名称	起始位点	终止位点	参考碱基	等位基因	质量值	位置类型	位置注释	FPR 校正后的 <i>P</i> 值	解释贡献率	最小等位基因频率
<i>hcn4</i>	4 455 704	4 455 704	G	T	29 762.48	intronic	ncbi_109522800	0.000 000 723 924 60	0.472 095 002 974 714	0.453 703 703 703 704
	4 455 767	4 455 767	G	T	29 372.86	intronic	ncbi_109522800	0.000 000 487 613 20	0.595 896 294 762 575	0.462 732 919 254 658
	4 457 744	4 457 744	T	C	33 265.98	intronic	ncbi_109522800	0.000 000 825 188 60	0.589 567 117 569 121	0.337 423 312 883 436
<i>pkm</i>	4 455 704	4 455 704	G	T	29 762.48	intronic	ncbi_109522774	0.000 000 723 924 60	0.472 095 002 974 714	0.453 703 703 703 704
	4 455 767	4 455 767	G	T	29 372.86	intronic	ncbi_109522774	0.000 000 487 613 20	0.595 896 294 762 575	0.462 732 919 254 658
	4 457 744	4 457 744	T	C	33 265.98	intronic	ncbi_109522774	0.000 000 825 188 60	0.589 567 117 569 121	0.337 423 312 883 436
<i>clk2</i>	4 585 460	4 585 460	A	T	27 244.57	intronic	ncbi_109522776	0.000 000 569 141 60	0.424 100 049 091 749	0.461 290 322 580 645
	4 585 674	4 585 674	G	C	32 327.79	intronic	ncbi_109522776	0.000 000 970 891 10	0.558 579 975 151 278	0.401 234 567 901 235
<i>rec8</i>	4 585 460	4 585 460	A	T	27 244.57	intronic	ncbi_109522780	0.000 000 569 141 60	0.424 100 049 091 749	0.461 290 322 580 645
	4 585 674	4 585 674	G	C	32 327.79	intronic	ncbi_109522780	0.000 000 970 891 10	0.558 579 975 151 278	0.401 234 567 901 235
<i>setdb1</i>	4 585 460	4 585 460	A	T	27 244.57	intronic	ncbi_109522778	0.000 000 569 141 60	0.424 100 049 091 749	0.461 290 322 580 645
	4 585 674	4 585 674	G	C	32 327.79	intronic	ncbi_109522778	0.000 000 970 891 10	0.558 579 975 151 278	0.401 234 567 901 235
	4 602 508	4 602 508	G	A	17 779.35	intergenic	ncbi_109522778	0.000 000 172 130 00	0.459 986 875 680 455	0.365 853 658 536 585
	4 603 666	4 603 666	T	G	26 910.08	intergenic	ncbi_109522778	0.000 000 079 646 78	0.608 453 104 309 160	0.420 245 398 773 006
<i>arnt</i>	4 585 460	4 585 460	A	T	27 244.57	intronic	ncbi_109522788	0.000 000 569 141 60	0.424 100 049 091 749	0.461 290 322 580 645
	4 585 674	4 585 674	G	C	32 327.79	intronic	ncbi_109522788	0.000 000 970 891 10	0.558 579 975 151 278	0.401 234 567 901 235
	4 602 508	4 602 508	G	A	17 779.35	intergenic	ncbi_109522788	0.000 000 172 130 00	0.459 986 875 680 455	0.365 853 658 536 585
	4 603 666	4 603 666	T	G	26 910.08	intergenic	ncbi_109522788	0.000 000 079 646 78	0.608 453 104 309 160	0.420 245 398 773 006
<i>cers2b</i>	4 585 460	4 585 460	A	T	27 244.57	intronic	ncbi_109522795	0.000 000 569 141 60	0.424 100 049 091 749	0.461 290 322 580 645
	4 585 674	4 585 674	G	C	32 327.79	intronic	ncbi_109522795	0.000 000 970 891 10	0.558 579 975 151 278	0.401 234 567 901 235
	4 602 508	4 602 508	G	A	17 779.35	intergenic	ncbi_109522795	0.000 000 172 130 00	0.459 986 875 680 455	0.365 853 658 536 585
	4 603 666	4 603 666	T	G	26 910.08	intergenic	ncbi_109522795	0.000 000 079 646 78	0.608 453 104 309 160	0.420 245 398 773 006
<i>celf3</i>	4 585 460	4 585 460	A	T	27 244.57	intronic	ncbi_122843442	0.000 000 569 141 60	0.424 100 049 091 749	0.461 290 322 580 645
	4 585 674	4 585 674	G	C	32 327.79	intronic	ncbi_122843442	0.000 000 970 891 10	0.558 579 975 151 278	0.401 234 567 901 235
	4 602 508	4 602 508	G	A	17 779.35	intergenic	ncbi_122843442	0.000 000 172 130 00	0.459 986 875 680 455	0.365 853 658 536 585
	4 603 666	4 603 666	T	G	26 910.08	intergenic	ncbi_122843442	0.000 000 079 646 78	0.608 453 104 309 160	0.420 245 398 773 006
<i>adamtsl4</i>	4 602 508	4 602 508	G	A	17 779.35	intergenic	ncbi_109522783	0.000 000 172 130 00	0.459 986 875 680 455	0.365 853 658 536 585
	4 603 666	4 603 666	T	G	26 910.08	intergenic	ncbi_109522783	0.000 000 079 646 78	0.608 453 104 309 160	0.420 245 398 773 006
<i>syngap1</i>	5 207 935	5 207 948	GCAG- CAAAT- GTAC- CG	—	21 034.68	intronic	ncbi_109512188	0.000 000 815 544 40	0.297 958 384 595 123	0.376 760 563 380 282
			GCAG- CAAAT- GTAC- CG	—						
<i>phf1</i>	5 207 935	5 207 948	GCAG- CAAAT- GTAC- CG	—	21 034.68	intronic	ncbi_109512277	0.000 000 815 544 40	0.297 958 384 595 123	0.376 760 563 380 282
			GCAG- CAAAT- GTAC- CG	—						
<i>rgl2</i>	5 207 935	5 207 948	GCAG- CAAAT- GTAC- CG	—	21 034.68	intronic	ncbi_109512261	0.000 000 815 544 40	0.297 958 384 595 123	0.376 760 563 380 282
			GCAG- CAAAT- GTAC- CG	—						
<i>agps</i>	5 983 061	5 983 061	G	A	15 433.28	intronic	ncbi_109507896	0.000 000 157 376 60	0.381 739 227 798 152	0.410 852 713 178 295
<i>pde11a</i>	5 983 061	5 983 061	G	A	15 433.28	intronic	ncbi_111218419	0.000 000 157 376 60	0.381 739 227 798 152	0.410 852 713 178 295
<i>eye</i>	5 983 061	5 983 061	G	A	15 433.28	intronic	ncbi_109531259	0.000 000 157 376 60	0.381 739 227 798 152	0.410 852 713 178 295
<i>osbp16</i>	5 983 061	5 983 061	G	A	15 433.28	intronic	ncbi_109531257	0.000 000 157 376 60	0.381 739 227 798 152	0.410 852 713 178 295
<i>tilb</i>	8 448 842	8 448 841	A	—	31 050.06	UTR5	ncbi_109518817	0.000 000 768 553 30	0.175 190 228 537 048	0.173 913 043 478 261
<i>kcnq3</i>	8 448 842	8 448 841	A	—	31 050.06	UTR5	ncbi_109518882	0.000 000 768 553 30	0.175 190 228 537 048	0.173 913 043 478 261
<i>hhla1</i>	8 448 842	8 448 841	A	—	31 050.06	UTR5	ncbi_113010263	0.000 000 768 553 30	0.175 190 228 537 048	0.17 391 304 347 8261
<i>efr3a</i>	8 448 842	8 448 841	A	—	31 050.06	UTR5	ncbi_107087009	0.000 000 768 553 30	0.175 190 228 537 048	0.173 913 043 478 261
<i>adcys8</i>	8 448 842	8 448 841	A	—	31 050.06	UTR5	ncbi_109518881	0.000 000 768 553 30	0.175 190 228 537 048	0.173 913 043 478 261

将候选基因使用 GO 和 KEGG 数据库进行功能富集分析，得到聚类富集分析图（见图 4）。GO 功能富集分析显示，差异基因主要富集在 smooth septate junction（GO:0005920）、septate junction（GO:005918）、ESC/（Z）complex（GO:0035098）等通路（见图 4a）；KEGG 显著富集的重要通路与 Proximal tubule bicarbonate reclamation（ko04964）、Carbohydrate digestion and absorption（ko04973）、Aldosterone-regulated sodium reabsorption（ko04960）、Mineral absorption（ko04978）等碳水化合物和离子吸收



有关 (见图 4b)。进一步对这些基因在其他物种中的可能作用和功能进行分析, 最终筛选了 15 个与生长相关的基因, 分别是 *rbp1* (retinol binding protein 1)、*ptpn6* (protein tyrosine phosphatase SHP-1)、*cited3* (cAMP-responsive element-binding protein (CBP)/p300-interacting transactivator with glutamic acid/aspartic acid-rich tail 3)、*scmh1* (scm polycomb group protein homolog 1)、*foxo3* (forkhead box O3)、*cyp4b1* (cytochrome P450 4B1)、*pkm* (pyruvate kinase M1/2)、*clk2* (CDC-like kinase 2)、*setdb1* (SET domain bifurcated histone lysine methyltransferase 1)、*arnt* (aryl hydrocarbon receptor nuclear translocator)、*cers2* (ceramide synthase 2)、*phf1* (PHD finger protein-1)、*rgl2* (repressor of ga-like2)、*cyc* (cystatin C)、*adcy8* (adenylyl cyclase 8)。同时检测了筛选的 15 个候选基因在 FG 组和 SG 组膨腹海马肌肉中的表达量, 发现 *ptpn6*、*cited3*、*scmh1*、*foxo3*、*pkm*、*clk2*、*setdb1*、*arnt*、*cers2*、*phf1*、*rgl2*、*cyc* 和 *adcy8* 的表达量在 FG 组膨腹海马中显著高于 SG 组, 而 *rbp1* 和 *cyp4b1* 的表达量在 FG 组膨腹海马中显著低于 SG 组 (见表 4)。

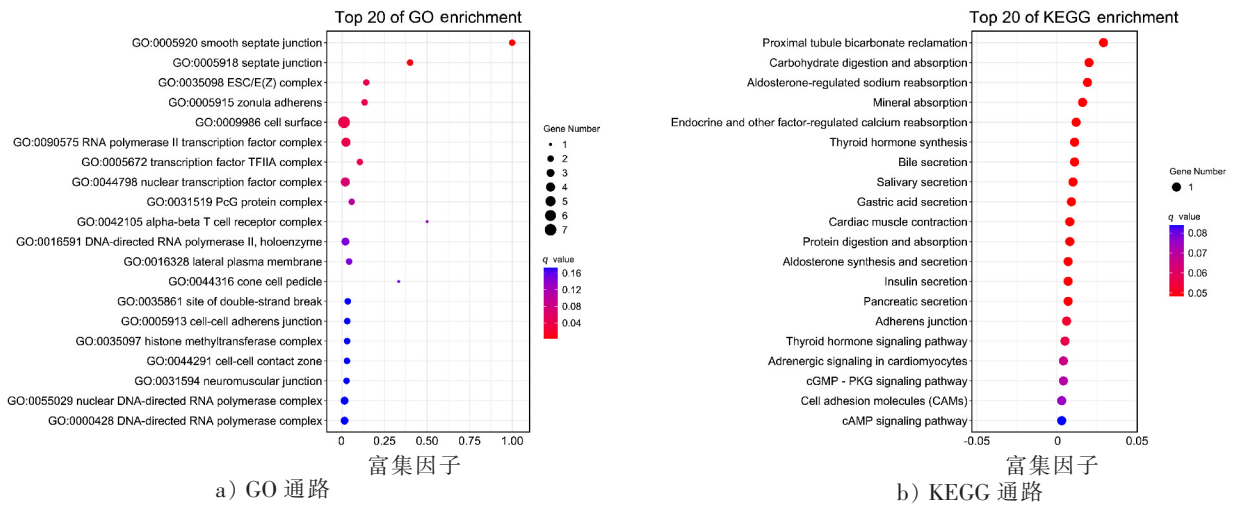


图 4 差异表达基因 GO 和 KEGG 通路富集分析

Fig.4 GO and KEGG enrichment results of the differentially expressed genes

表 4 15 个与生长相关的基因在 FG 组和 SG 组膨腹海马肌肉中的表达量比较

Tab. 4 The expression patterns of 15 growth-related genes in FG and SG in *Hippocampus abdominalis*

基因名称	基因表达量	
	FG 组	SG 组
<i>adcy8</i>	0.004 471 0 ± 0.001 173 00	0.002 665 0 ± 0.000 741 80 *
<i>arnt</i>	0.000 707 4 ± 0.000 186 80	0.000 347 4 ± 0.000 106 80 *
<i>cers2</i>	0.057 180 0 ± 0.006 534 00	0.044 010 0 ± 0.012 780 00 *
<i>citdb1</i>	0.036 100 0 ± 0.005 265 00	0.022 680 0 ± 0.008 476 00 *
<i>clk2</i>	0.000 620 1 ± 0.000 143 00	0.000 398 9 ± 0.000 092 00 *
<i>cyc</i>	0.000 611 8 ± 0.000 104 10	0.000 442 0 ± 0.000 078 40 *
<i>cyp4b1</i>	0.029 510 0 ± 0.007 255 00	0.041 350 0 ± 0.006 453 00 *
<i>foxo3</i>	0.008 556 0 ± 0.000 695 30	0.006 339 0 ± 0.001 269 00 *
<i>phf1</i>	0.000 332 0 ± 0.000 116 90	0.000 179 6 ± 0.000 354 30 *
<i>pkm</i>	0.002 078 0 ± 0.000 480 60	0.001 642 0 ± 0.000 309 60 *
<i>ptpn6</i>	0.002 445 0 ± 0.000 658 00	0.001 664 0 ± 0.000 464 70 *
<i>rbp1</i>	0.000 259 8 ± 0.000 107 70	0.000 526 8 ± 0.000 133 60 *
<i>rgl2</i>	0.000 320 2 ± 0.000 080 75	0.000 231 1 ± 0.000 066 35 *
<i>scmh1</i>	0.003 072 0 ± 0.000 860 20	0.002 183 0 ± 0.000 513 10 *
<i>setdb1</i>	0.116 100 0 ± 0.021 820 00	0.085 720 0 ± 0.010 050 00 *

说明: 基因相对表达水平用平均值 ± 标准差表示 (n = 8); \* 表示基因表达存在显著差异。

### 3 讨论

随着高通量测序技术的发展,越来越多的鱼类开展了基因组测序,这为鱼类生长性状的 GWAS 分析提供了更多的数据资源<sup>[17-18]</sup>。GWAS 已成为鉴定生长性状关联的基因标记的重要工具,它可以深入了解鱼类生长发育的分子机制<sup>[19-20]</sup>。重测序技术可以为 GWAS 分析提供准确的基因组数据,从而提高检测基因变异的准确性,促进基因座的精细定位和因果变异的识别,增强对复杂性状遗传基础的理解。GWAS 作为一种高效的高通量遗传分析技术,可以在全基因组范围内检测数以百万计的 SNP 标记,无需预先假设候选基因即可系统性筛选与目标性状显著关联的遗传位点,突破了传统数量性状基因座 (quantitative trait loci, QTL) 定位对遗传群体构建的依赖。重测序技术在 GWAS 中的应用可以提供一个全面的基因组视图,使罕见变异的检测和功能注释的增强成为可能,提高了分辨率和准确性,促进了基因座的精细定位和因果变异的识别<sup>[21]</sup>。目前,在太平洋牡蛎<sup>[22]</sup>、条石鲷<sup>[23]</sup>、大西洋鲑鱼<sup>[24]</sup>和尼罗罗非鱼<sup>[25]</sup>中均采用 GWAS 挖掘了生长性状相关的 SNP 位点和基因。本研究开展了膨腹海马高质量全基因组重测序数据研究,更好地挖掘罕见和低频变异位点,为 GWAS 分析提供高质量的遗传标记。采用来自 154 个个体的 4 834 844 个高质量 SNP 进行 GWAS 分析,首次对膨腹海马的生长性状进行分析,筛选到 27 个与膨腹海马生长性状显著相关的 SNP 位点和 42 个潜在影响生长的候选基因,这表明,膨腹海马生长性状的复杂遗传结构是由多基因控制的。

本研究共鉴定出 15 个潜在候选基因,并对潜在基因进行 GO 和 KEGG 分析,发现这些基因主要集中在营养物质吸收、转运和代谢、细胞增殖分裂相关信号通路上,尤其是 *pkm*、*arnt* 和 *adcy8* 基因在调控糖代谢过程中发挥了重要作用。丙酮酸激酶 M (Pkm) 是糖酵解中的关键酶,以多聚腺苷二磷酸核糖基化依赖方式结合核糖体,诱导翻译停滞,从而将细胞代谢与 mRNA 翻译联系起来,是糖酵解中的限速基因<sup>[26]</sup>。并且, Pkm 与磷酸烯醇丙酮酸羧基激酶 (Pepck) 的共同乙酰化可以调节糖酵解和糖异生,从而稳定糖代谢的过程<sup>[27]</sup>。此外, Arnt 是一种芳香烃受体核转运蛋白,在葡萄糖摄取和线粒体氧化磷酸化过程中发挥重要作用<sup>[28]</sup>。腺苷酸环化酶 8 (Adcy8) 是下丘脑通过调节胰岛素分泌去适应高脂肪饮食所必需的, Adcy8 在葡萄糖代谢的稳态调控中发挥了多方面的作用<sup>[29]</sup>。本文的研究发现,编码这些蛋白质的基因在 FG 组的海马中表达量高,表明膨腹海马可能通过调节糖代谢间接调节了其生长速度,但是这些基因在膨腹海马中的具体作用机制还有待进一步研究。

此外,本研究还筛选到多个与细胞增殖、分裂相关的基因,主要包括 *rbp1*、*clk2*、*cited3* 和 *arnt* 等。其中, Rbp1 是视黄酸结合蛋白中的一员, *rbp1* 基因对上皮细胞的增殖和微环境的稳定发挥重要作用,更有研究发现, *rbp1* 可以通过抑制 *e2f* 的转录导致细胞的生长停滞<sup>[30-31]</sup>。本文的研究发现, *rbp1* 在 SG 组的海马中表达量比较高,推测 *rbp1* 高表达可能抑制了膨腹海马表皮细胞生长的停滞。 Clk2 是一种在进化上具有保守性的蛋白质,它能促进非洲爪蟾胚胎的早期发育,在哺乳类动物中具有调节细胞周期进程和端粒长度的功能<sup>[32-33]</sup>。本文的研究发现, *clk2* 在 FG 组的海马中高表达,推测该基因可能促进了膨腹海马的生长和发育。 *cited3* 在鱼类肌肉和骨骼的形成中有重要作用,在斑马鱼中敲低 *cited3* 基因会损害鱼体肌肉细胞的分化和生长,导致肌肉细胞死亡,敲低补偿后可挽救肌肉细胞死亡,表明 *cited3* 是调控鱼类肌肉生长的关键基因<sup>[34]</sup>。 *arnt* 在肌肉细胞的分裂中发挥关键作用,离体培养的肌肉细胞感染 *arnt* 基因后,其细胞的分裂能力会降低,并且该基因会严重阻碍肌肉的再生能力<sup>[35]</sup>。FG 组膨腹海马的 *cited3* 和 *arnt* 基因表达显著高于 SG 组的,表明膨腹海马可能通过这两个基因调控肌肉生长和再生,从而影响体质量。

### 4 结论

本研究对 154 份膨腹海马个体进行基因组重测序,得到 4 170 485 个高质量 SNP 位点,并对生长性状开展 GWAS 分析,最终在 2、3、6 和 13 号染色体上定位到了 27 个显著相关的 SNP 位点。同时,在这些 SNP 位点上下游 50 kb 范围内筛选到 42 个与生长性状关联的基因,经过基因表达分析及功能

挖掘,初步鉴定到了15个与生长性状相关的候选基因,它们主要与糖代谢、营养物质吸收和转运、肌肉骨骼生长、细胞增殖等生理活动相关。

## [ 参 考 文 献 ]

- [1] LOURIE S A, STANLEY H F, VINCENT A C, et al. Seahorses: an identification guide to the world's species and their conservation[M]. London: VLIJ, 1999.
- [2] NEPAL T K, MANITA M. An updated checklist of globally threatened species in Bhutan as listed in IUCN red list of threatened species[J]. International Journal of Science and Research, 2021, 10(2): 1640-1646.
- [3] LU J Y, WU J Y, YANG D W, et al. Growth rates of the great horse, *Hippocampus magna*, under captive breeding conditions[J]. Journal of Fishery Sciences of China, 2001, 8(1): 59-63.
- [4] HE L B, QI J F, LIN J B, et al. Key techniques for artificial propagation and nursery of *Hippocampus hippocampus*[J]. Journal of Applied Oceanography, 2022, 41(4): 701-707.
- [5] HE Y M, XING G F, LIU M T, et al. Genome-wide association analysis: opportunities and challenges for genomics research[J]. Biotechnology Bulletin, 2013(6): 1-6.
- [6] 唐富福, 徐非非, 包劲松. 全基因组关联分析在水稻遗传育种中的应用[J]. 核农学报, 2013, 27(5): 598-606.
- [7] 牛安然, 张兴, 杨雨婷, 等. 全基因组关联分析在猪育种中的研究进展[J]. 畜牧与兽医, 2023, 55(5): 139-147.
- [8] 阿仑, 张金文, 王学峰, 等. 全基因组选择育种技术在规模化奶牛养殖场的应用实例[J]. 中国畜禽种业, 2023, 19(5): 41-44.
- [9] 方家璐, 海佳薇, 周林燕, 等. 基于黄河鲤鱼质量性状的全基因组选择模型评估[J]. 大连海洋大学学报, 2024, 39(3): 437-444.
- [10] DONG L S, XIAO S J, WANG Q R, et al. Comparative analysis of the GBLUP, emBayesB, and GWAS algorithms to predict genetic values in large yellow croaker (*Larimichthys crocea*) [J]. BMC Genomics, 2016, 17: 460.
- [11] CUI T X, LIU H Y, ZHANG J, et al. Development and evaluation of a 50 kb SNP array for blotched snakehead (*Channa maculata*) [J]. Aquaculture, 2025, 598: 742020.
- [12] LIU H Y, XIA W W, OU M, et al. A genome-wide association study to identify growth-related SNPs and genes in blotched snakehead (*Channa maculata*) [J]. Aquaculture Reports, 2024, 35: 101932.
- [13] CHANG C C, CHOW C C, CAM T L, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets [J]. GigaScience, 2015, 4: 7.
- [14] HE L B, LONG X, QI J, et al. Genome and gene evolution of seahorse species revealed by the chromosome-level genome of *Hippocampus abdominalis* [J]. Mol Ecol Resour, 2022, 22(4): 1465-1477.
- [15] ASHBURNER M, BALL C A, BLAKE J A, et al. Gene ontology: tool for the unification of biology [J]. Nat Genet, 2000, 25(1): 25-29.
- [16] OGATA H, GOTO S, SATO K, et al. KEGG: kyoto encyclopedia of genes and genomes [J]. Nucleic Acids Research, 1999, 27(1): 29-34.
- [17] WANG L, SUN F, WAN Z Y, et al. Transposon-induced epigenetic silencing in the X chromosome as a novel form of *dmrt1* expression regulation during sex determination in the fighting fish [J]. BMC Biology, 2022, 20: 1-16.
- [18] ZHANG W C, WANG H R, BRANDT D Y C, et al. The genetic architecture of phenotypic diversity in the betta fish (*Betta splendens*) [J]. Science Advances, 2022, 8: 4955.
- [19] YUE G H, WANG L, YANG Z T, et al. Genomic resources and their applications in aquaculture of Asian seabass (*Lateolabrax niloticus*) [J]. Reviews in Aquaculture, 2023, 15: 853-871.
- [20] ZHANG C, WEN H, ZHANG Y, et al. First genome-wide association study and genomic prediction for growth traits in spotted sea bass (*Lateolabrax maculatus*) using whole-genome resequencing [J]. Aquaculture, 2023, 566: 739194.
- [21] RIVAS M A, BEAUDOIN M, GARDET A, et al. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease [J]. Nature Genetics, 2011, 43: 1066-1073.
- [22] GUTIERREZ A P, MATIKA O, BEAN T P, et al. Genomic selection for growth traits in Pacific oyster (*Crassostrea gigas*): potential of low-density marker panels for breeding value prediction [J]. Frontiers in Genetics, 2018(9): 391-399.

- [23] GONG J A O,ZHAO J,KE Q,et al. First genomic prediction and genome-wide association for complex growth-related traits in Rock Bream (*Oplegnathus fasciatus*) [J]. Evolutionary Applications,2021,15(4):523-536.
- [24] TSAI H Y,HAMILTON A,TINCH A E,et al. Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array[J]. BMC Genomics,2015,16:969.
- [25] YOSHIDA G M,LHORENTE J P,CORREA K,et al. Genomewide association study and cost-efficient genomic predictions for growth and fillet yield in Nile Tilapia (*Oreochromis niloticus*) [J]. Genes,Genom,Genet,2019,9(8):2597-2607.
- [26] KEJIOU N S,ILAN L,ALEXANDER F,et al. Pyruvate kinase M (PKM) binds ribosomes in a poly-ADP ribosylation dependent manner to induce translational stalling[J]. Nucleic Acids Research,2023,51(12):6461-6478.
- [27] XIONG Y,LEI Q Y,ZHAO S,et al. Regulation of glycolysis and gluconeogenesis by acetylation of PKM and PEPCK[J]. Cold Spring Harbor Symposia on Quantitative Biology,2011,76:285-289.
- [28] HUANG C R,CHANG T W,LEE C T,et al. ARNT deficiency represses pyruvate dehydrogenase kinase 1 to trigger ROS production and melanoma metastasis[J]. Oncogenesis,2021,10(1):11.
- [29] RAOUX M,VACHER P,PAPIN J,et al. Multilevel control of glucose homeostasis by adenylyl cyclase 8[J]. Diabetologia,2015,58(4):749-57.
- [30] YU J,PERRI M,JONES J W,et al. Altered RBP1 gene expression impacts epithelial cell retinoic acid,proliferation,and microenvironment[J]. Cells,2022,11(5):792.
- [31] LAI A,MARCELLUS R C,CORBEIL H B,et al. RBP1 induces growth arrest by repression of E2F-dependent transcription [J]. Oncogene,1999,18(12):2091.
- [32] VIRGIRINIA R P,JAHAN N,OKADA M,et al. Cdc2-like kinase 2 (Clk2) promotes early neural development in *Xenopus embryos* [J]. Development Growth and Differentiation,2019,61(6):365-377.
- [33] JIANG N,BENARD C Y,KEBIR H,et al. Human CLK2 links cell cycle progression,apoptosis,and telomere length regulation[J]. Journal of Biological Chemistry,2003,278(24):21678.
- [34] DEVAKANMALAI G S,ZUMRUT H E,ERTURUL M O,et al. *Cited3* activates *Mef2c* to control muscle cell differentiation and survival[J]. Biology Open,2013,2(5):505-514.
- [35] ENDO Y,BALDINO K,LI B,et al. Loss of ARNT in skeletal muscle limits muscle regeneration in aging[J]. The FASEB Journal,2020,34(12):16086-16104.

(责任编辑 马建华 英文审校 刘静雯)