

DOI: 10.19663/j.issn2095-9869.20240305001

http://www.yykxjz.cn/

高进, 刘金叶, 陈傅晓, 王永波, 符书源. 豹纹鳃棘鲈生长差异分化相关 lncRNA 与 mRNA 加权基因共表达网络分析. 渔业科学进展, 2025, 46(1): 46–58

GAO J, LIU J Y, CHEN F X, WANG Y B, FU S Y. Weighted gene co-expression network of growth differentiation-related lncRNAs and mRNAs in *Plectropomus leopardus*. Progress in Fishery Sciences, 2025, 46(1): 46–58

豹纹鳃棘鲈生长差异分化相关 lncRNA 与 mRNA 加权基因共表达网络分析*

高 进^{1,2} 刘金叶^{1,3} 陈傅晓^{1,2,3} 王永波^{1,2} 符书源^{1,2,3①}

(1. 海南省海洋与渔业科学院 海南 海口 571126; 2. 海南热带海洋学院崖州湾创新研究院 海南 三亚 572025; 3. 海南省热带海水养殖工程技术研究中心 海南 海口 571126)

摘要 为揭示豹纹鳃棘鲈(*Plectropomus leopardus*)生长差异分化的分子调控作用机制, 该研究利用豹纹鳃棘鲈中间培育过程中不同生长阶段(42、70 和 91 日龄)的 18 个幼鱼肌肉组织样本进行全转录组测序, 获得 lncRNA 和 mRNA 基因表达谱数据, 采用加权基因共表达网络分析(WGCNA)方法构建生长差异分化相关 lncRNA 和 mRNA 的共表达网络。通过特异性基因模块中差异表达基因的 GO 和 KEGG 富集对模块生物学功能进行分析, 运用 Cytoscape 软件构建基因互作网络, 挖掘豹纹鳃棘鲈生长差异分化相关的关键候选基因。结果显示, 基因差异表达分析共筛选到 6 252 个差异表达 mRNA 和 367 个差异表达 lncRNA; 共表达网络分析获得 MEmagenta、MEgreenyellow 和 MEblack 共 3 个生长差异分化特异性基因模块; GO 和 KEGG 富集分析显示, 特异性模块中差异表达基因显著富集于肌肉蛋白形成、横纹肌组织发育调控、主轴组织形成等生物学过程, 参与了蛋白酶体、肌动蛋白细胞骨架调控及肌醇磷酸代谢等与生长差异分化相关的代谢通路; 筛选 3 个特异性模块中权重和连通度高的 lncRNA 和 mRNA 构建基因互作网络, 得到包括 *psmd6*、*nmt1*、*als2*、*brcc3*、*kank1*、*ada12*、*at131* 和 *hdac3* 等在内的生长差异分化关键基因和 *MSTRG.15660*、*MSTRG.8694*、*MSTRG.1896* 等 20 个 lncRNA, 为进一步解析豹纹鳃棘鲈生长差异分化的分子机制提供了新思路。

关键词 豹纹鳃棘鲈; 生长差异分化; WGCNA; 共表达网络

中图分类号 S917.4 文献标识码 A 文章编号 2095-9869(2025)01-0046-13

豹纹鳃棘鲈(*Plectropomus leopardus*), 俗称东星斑, 是石斑鱼类中的高端养殖品种。与大多数石斑鱼类一样, 豹纹鳃棘鲈的人工育苗和中间培育需要历经鱼苗分筛和饵料驯化, 此过程中同一批次的仔鱼、稚鱼和幼鱼会多次出现生长差异分化的现象(图 1), 是导致其养殖周期拉长、影响养殖经济效益的主要因素之一。为解决上述问题, 多数石斑鱼的育苗过程中都

会进行“标粗”, 并筛选生长差异显著的“鱼头”、“鱼中”和“鱼尾”, 在分筛鱼苗后按照不同的体型规格分别培育(彭树锋等, 2012), 一定程度上提高了鱼苗成活率并缩短了整体养殖周期。然而, “标粗”过程不仅受人为因素影响较大, 且鱼苗自身差异也影响“标粗”成效。尽管这种生长差异分化现象受环境因素的影响较大(符书源等, 2009), 但其仍存在尚未明

* 海南省科技计划三亚崖州湾科技城联合项目(320LH030)资助。高 进, Email: gaojin427@126.com

① 通信作者: 符书源, 副研究员, Email: fulank23@sohu.com

收稿日期: 2024-03-05, 收修改稿日期: 2024-03-14

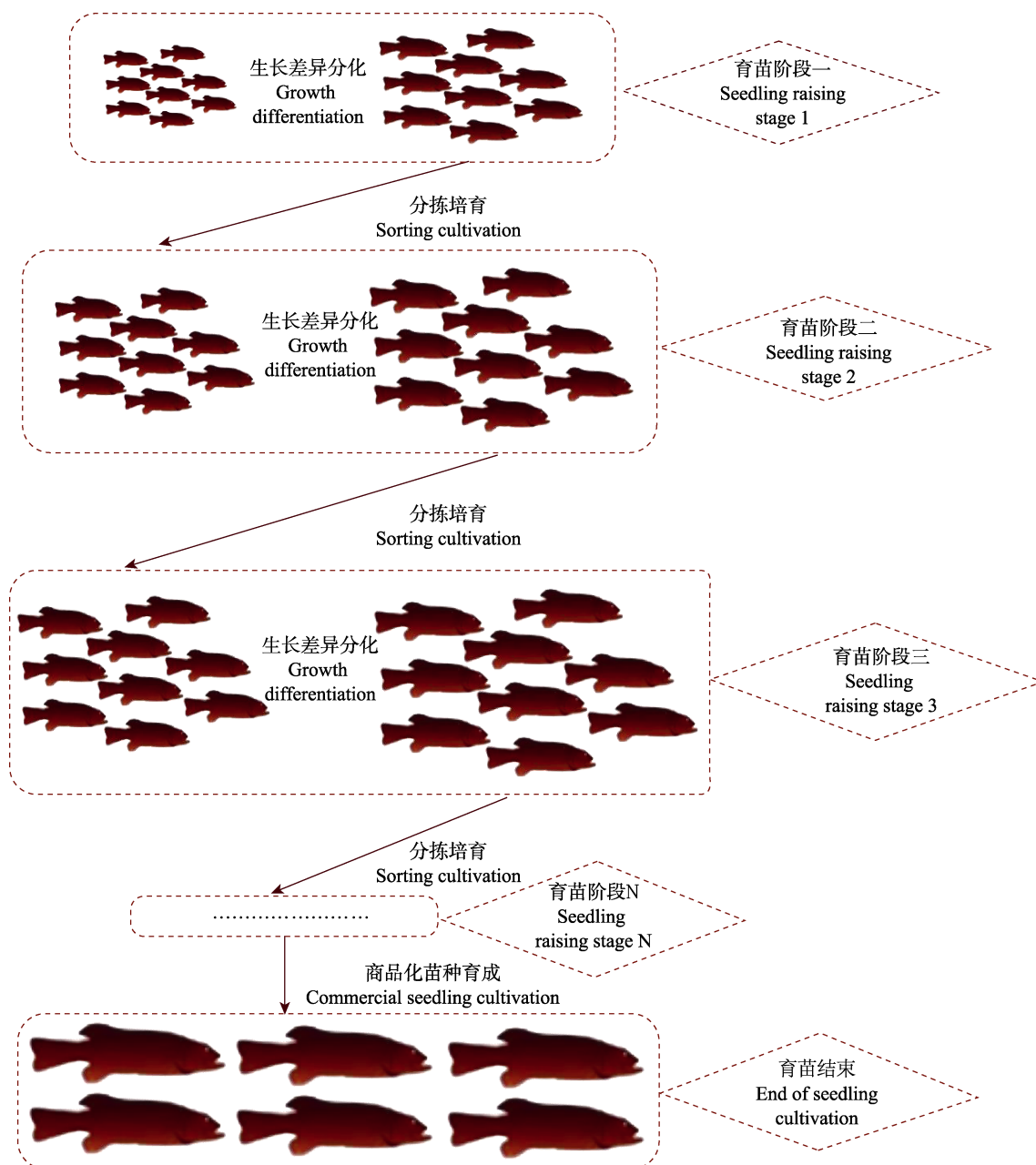


图 1 豹纹鳃棘鲈苗种中间培育过程中的生长差异分化现象

Fig.1 Growth differentiation phenomenon during intermediate cultivation of *P. leopardus* juveniles

确的潜在遗传因素。

鱼类生长性能属于复杂性状,生长差异性状分子调控和遗传作用机制解析已成为鱼类育种研究领域的热点(Enberg *et al*, 2012; Guerrero-Cozar *et al*, 2021)。生长性状的优劣是评价鱼类养殖经济效益的重要指标,同时受遗传内因与环境外因影响,因此,开展养殖鱼类生长相关性状的遗传机制及调控策略研究将为其品种改良和育种提供参考(苏军虎等, 2012)。全转录组测序技术因其具有分析辨别基因组结构功能、构建生物系统遗传网络以及可靠性高等优

点,被广泛应用于动植物遗传机制解析研究(Zhao *et al*, 2021; Niu *et al*, 2022; Li *et al*, 2022)。全转录组分析在编码 RNA 研究的基础上,进一步涵盖了多种非编码 RNA 及各种类型 RNA 间的调控网络分析(石田培等, 2019)。Ye 等(2021)对具有不同生长速度的 29 条 6 月龄草鱼(*Ctenopharyngodon idella*)的脑部和肝胰腺组织进行全转录组分析,筛选出多个与草鱼生长性状显著相关的 mRNA、miRNA 和 lncRNA 关键因子,为探索草鱼生长速率的分子遗传机制奠定了基础。Wang 等(2022)通过全转录组测序构建了半滑舌鳎

(*Cynoglossus semilaevis*) 性别大小异形性状的 lncRNA-miRNA-mRNA 网络, 揭示了其雌雄生长差异遗传机制。Wu 等(2022)基于全转录组测序的野生型和黄色突变型虹鳟(*Oncorhynchus mykiss*)皮肤中 lncRNA 和 circRNA 介导的 ceRNA 调控网络分析, 解析了二者之间先天免疫的差异, 进一步扩展了对虹鳟鱼先天免疫系统的理解, 并为虹鳟免疫机制研究和疾病抗性育种奠定了基础(Wu *et al*, 2022)。

lncRNA 是一类超 200 bp 且不具编码功能的 RNA 分子, 其通过与 RNA、DNA 或蛋白质相互作用调控靶基因的表达, 广泛参与水产动物生长、肌肉品质、生殖、胚胎发育、性别分化及免疫等多种生理功能(宋飞彪等, 2020)。由于复杂性状在遗传层面通常受多个分子共同作用, 同时利用 mRNA 和 lncRNA 开展调控机制研究, 并运用系统生物学网络的方法对基因表达数据进行深度挖掘, 构建与性状相关的基因网络已成为生物信息学分析的重要研究方向(汪涛等, 2014; 陈华峰等, 2019; Zhang *et al*, 2021)。加权共表达网络分析(WGCNA)因其在样本间基因关联的网络关系构建准确性高, 被广泛应用于转录表达方面的研究(Li *et al*, 2018; Yao *et al*, 2023)。目前, 尚未见关于豹纹鳃棘鲈生长差异分化性状共表达网络分析的报道。

本研究采用 R 语言的 WGCNA 编辑包, 对豹纹鳃棘鲈苗种“标粗”过程中 3 个生长阶段共计 18 个肌肉组织样本的转录组表达数据进行分析, 构建与生长差异分化相关的 lncRNA 与 mRNA 共表达模块。通过差异表达基因分析和目标模块内基因的 GO 和 KEGG 富集分析, 确定目标模块中的核心基因, 深入挖掘与生长差异分化相关的基因及调控网络, 以期为豹纹鳃棘鲈的生长差异分化遗传机制解析和良种选育提供依据。

1 材料与方法

1.1 实验数据来源

本研究在豹纹鳃棘鲈生长差异分化的全转录组测序数据的基础上进行。课题组于 2022 年在海南省海洋与渔业科学院海研热带海水鱼类良种场开展了豹纹鳃棘鲈苗种培育工作, 在鱼苗“标粗”过程中的 3 个生长阶段(42、70 和 91 日龄)分别采集了“鱼头”中生长差异分化显著鱼苗的肌肉组织, 每个时期大小规格各 3 个生物学重复, 共计 18 个样本送至广州基迪奥生物科技有限公司进行全转录组测序。测序获得的 mRNA 和 lncRNA 原始下机数据于 2023 年 4 月返回, 后续生物信息学分析在海南省海洋与渔业科学院

进行。

1.2 筛选差异表达基因

通过去除含有接头、不确定性占比大于 10% 的碱基以及低质量($Q_{\text{phred}} \leq 10$)的碱基数占整条 read 50% 以上的 reads 对原始下机数据(raw data)进行严格的过滤, 获得高质量测序数据(clean data)以提高后续转录组分析的准确性。利用 HISAT2 软件(Kim *et al*, 2015)将过滤后的 clean reads 高效比对到豹纹鳃棘鲈参考基因组(<https://db.cngb.org/search/project/CNP0000859>), 通过 StringTie 软件(Pertea *et al*, 2015)进行基因水平的定量, 获取表达谱数据。采用 CPC2 (Kang *et al*, 2017)、CNCI (Sun *et al*, 2013)和 LGC (Wang *et al*, 2019) 3 个软件预测样本中的 lncRNA, 为降低鉴定结果的假阳性, 取 3 个软件分析结果的交集 lncRNA 用于后续分析。采用 FPKM (fragments per kilobase of transcript per million fragments mapped)法(Florea *et al*, 2013)归一化 reads 数和转录本长度, 使不同样本间基因表达具有可比性。对苗种“标粗”过程中 3 个生长阶段中大规格(L)和小规格(S)样本的 lncRNA 和 mRNA 表达谱数据进行对比(L-Stage 1 vs S-Stage 1、L-Stage 2 vs S-Stage 2 以及 L-Stage 3 vs S-Stage 3), 根据各转录本的 FPKM 值, 利用 DESeq2 软件(Love *et al*, 2014)进行差异表达分析, 并以 Fold Change ≥ 2 且 FDR (false discovery rate) ≤ 0.05 为阈值标准筛选差异表达基因。

1.3 WGCNA 分析

采用 R 语言(v4.0.5)的 WGCNA 编辑包(v1.72.5) (Langfelder *et al*, 2008)构建包括全部 mRNA 和 lncRNA 的无尺度网络。以 17 个(剔除 1 个 S-Stage 2 离群样本)生长差异分化转录组中标准化 FPKM 值的表达矩阵作为输入数据, 筛选过滤表达值方差 > 0 和缺失 $\leq 10\%$ 的基因构建加权共表达网络。利用 WGCNA 中的 pickSoftThreshold 函数计算无尺度拓扑拟合指数, 筛选共表达网络的软阈值, 并利用 powerEstimate 函数估计其最优值; 选择最优软阈值后, 使用 TOMsimilarityFromExpr 函数计算拓扑重叠矩阵(topological overlapmatrix, TOM); 利用拓扑相异矩阵(1-TOM)进行基因分层聚类(hclust), 并采用动态剪切算法(cutreeDynamic)划分模块(minClusterSize=50); 最后, 使用 mergeCloseModules 合并相似度大于 75% 的基因模块(mergeCutHeight=0.25), 确定最终的共表达基因模块。对共表达基因模块中的所有基因表达值进行主成分分析, 计算出各模块的特征向量(module eigengene, ME), 统计各模块的 ME 值与生长差异分

化性状(不同生长规格)的相关性,按照显著性选取相关系数绝对值大于特定阈值的模块作为豹纹鳃棘鲈生长差异分化特异性功能模块进行后续分析。

1.4 特异性功能模块富集分析

采用 R 语言 clusterProfiler 编辑包(Yu *et al*, 2012), 基于 eggNOG 数据库对选取的特异性功能模块中差异表达的 mRNA 及 lncRNA 分别进行基因本体论(GO)和京都基因与基因组百科全书(KEGG)功能通路富集分析, 并以 $P<0.05$ 为阈值富集获得显著性的 GO term 和 KEGG pathway。

1.5 关键基因筛选和网络可视化

通过 WGCNA 的 exportNetworkToCytoscape 函数输出特异性功能基因子网络的边列表和节点列表数据,筛选基因权重(Weight)和连通度(Degree)排序靠前的 mRNA 和 lncRNA, 基于功能基因注释结果, 选择前 65 个基因和 lncRNA, 利用 Cytoscape(v3.10.1)软件(Shannon *et al*, 2003)中的 CytoHubba 插件估计关键性(hub)大小, 并进行关键基因筛选及网络可视化。

2 结果

2.1 生长差异分化表型及差异表达基因

分别从 42、70 及 91 日龄相同生长规格鱼苗中挑选的生长差异分化个体作为实验用鱼, 其全长表型信息见表 1。基于全转录组测序获取 mRNA 和 lncRNA 基因表达谱, 利用 DESeq2 软件进行不同生长阶段的差异基因表达分析, 去除各生长阶段大小规格样本比对时重复出现的差异表达 mRNA 和 lncRNA, 共获得 5 944 个差异表达已知 mRNA、302 个 novel mRNA 以及 367 个差异表达 lncRNA(表 2), 其中, 各生长阶段差异表达上调或下调的 mRNA 和 lncRNA 数统计见图 2。3 个生长阶段同时检出的差异表达 mRNA 占全部差异表达 mRNA 的 17.3%, 大于差异表达 lncRNA 的占比 5.7%(图 3)。

2.2 共表达网络模块

为了减少构建无尺度分布网络模块的计算复杂度和计算时间, 筛选最适软阈值为 22, 此时, 无尺度网络的拟合指数 $R^2>0.8$, 平均连通性为 94.23(附图 1)。基于相异度矩阵 dissTOM 进行层次聚类得到基因聚类树, 利用混合动态树剪切法划分出 37 个基因模块(见附图 2 中 dynamic tree cut), 随后以相似度大于 75% 为标准合并相关性高的模块并获得 16 个共表达基因模块(见附图 2 中 merged dynamic)。共表达网络中各模块包含的 mRNA 和 lncRNA 及其差异表达数见表 3, 其中, MEMagenta 模块中的 mRNA 和 lncRNA

表 1 生长差异分化样本的全长表型平均数和标准差

Tab.1 Mean and standard deviation of phenotype of total length for growth differentiated samples

日龄 Days post hatching	大小规格 Size specification	平均数±标准差 Mean±SD/cm
42	L	4.00±0.10
	S	3.17±0.12
70	L	5.77±0.15
	S	4.30±0.17
91	L	7.40±0.10
	S	4.00±0.10

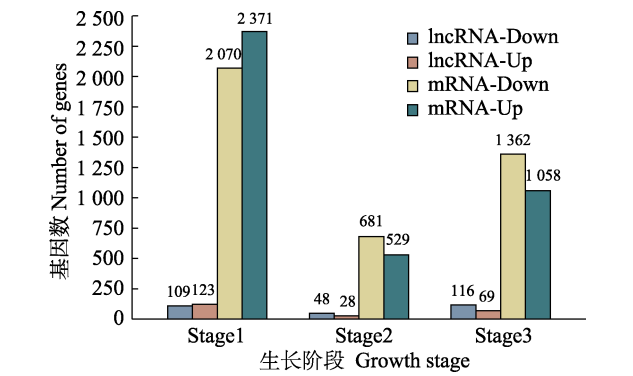


图 2 不同生长阶段差异表达上调或
下调 mRNA 和 lncRNA 数统计
Fig.2 Number of up-regulated or down-regulated mRNA
and lncRNA counts at different growth stages

表 2 mRNA 和 lncRNA 的差异表达统计

Tab.2 Statistics of differential expression of mRNA and lncRNA

差异表达类别 DE type	L-Stage 1 vs S-Stage 1	L-Stage 2 vs S-Stage 2	L-Stage 3 vs S-Stage 1	去重复 Deduplication
差异表达已知 mRNA 数 No. of DE known mRNAs	4 220	2 311	3 232	5 944
差异表达 novel mRNA 数 No. of DE novel mRNAs	221	109	198	302
差异表达 lncRNA 数 No. of DE lncRNAs	232	76	185	367

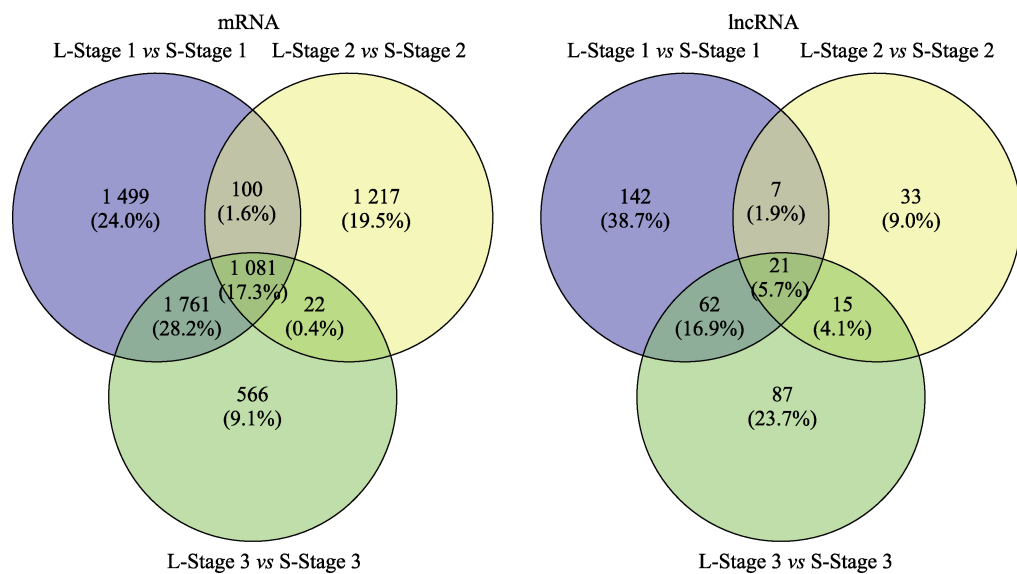


图 3 不同生长阶段差异表达 mRNA 和 lncRNA 韦恩图
Fig.3 Venn plot of DE mRNA and lncRNA at different growth stages

表 3 共表达基因模块内 mRNA 和 lncRNA 及差异表达数统计
Tab.3 mRNA, lncRNA and differential expression counts in co-expression modules

基因模块 Gene module	mRNA 数 No. of mRNA	lncRNA 数 No. of lncRNA	差异表达 mRNA 数 No. of DE mRNA	差异表达 lncRNA 数 No. of DE lncRNA	基因模块 Gene module	mRNA 数 No. of mRNA	lncRNA 数 No. of lncRNA	差异表达 mRNA 数 No. of DE mRNA	差异表达 lncRNA 数 No. of DE lncRNA
MEagenta	6 835	460	2 879	196	MElightyellow	226	20	56	3
MEturquoise	4 588	115	878	20	MEdarkolivegreen	197	7	36	1
MEgrey	3 821	252	1172	74	MEdarkred	195	6	59	3
MEgreenyellow	727	46	452	37	MEdarkorange	144	7	43	0
MEblack	547	21	104	15	MEskyblue	129	1	23	0
MEdarkturquoise	469	42	121	4	MEsaddlebrown	120	9	36	2
MEsalmon	376	13	124	5	MEdarkmagenta	92	0	30	0
MEdarkgrey	254	5	201	4	MEyellowgreen	66	1	18	1

数最多，包含 6 835 个 mRNA 和 460 个 lncRNA，该模块中 mRNA 和 lncRNA 差异表达数也是最多的，包含 2 879 个差异表达 mRNA 和 196 个差异表达 lncRNA；MEyellowgreen 模块中 mRNA 数和差异表达数最少，为 66 和 18；MEdarkmagenta 模块中不包含 lncRNA，MEdarkorange 和 MEskyblue 模块中未发现差异表达 lncRNA；MEgrey 模块中包含的是没有分配到其他模块中的 mRNA 和 lncRNA 集。

由豹纹鳃棘鲈不同生长阶段的生长差异分化性状与构建的模块相关关系(图 4)可知，本研究共鉴定到 3 个与生长差异分化特异相关的共表达模块，筛选标准为 3 个生长阶段都符合($0.65 \leq |r| \leq 1, P \leq 0.01$)。其中，MEagenta 模块在 3 个阶段的差异分化都呈正相关，而 MEgreenyellow 模块都呈负相关；MEblack 模块 42 日龄时与生长差异分化呈正相关，而 70 日龄和

91 日龄时则呈负相关。为了确定模块划分的可靠性，以上述 3 个共表达模块(MEagenta、MEblack 和 MEgreenyellow)为特异性模块进行深入分析。

2.3 特异性模块差异表达 mRNA 及 lncRNA 的 GO 和 KEGG 富集分析

为了探究特异性模块中差异表达 mRNA 及 lncRNA 的主要生物学功能和参与的基因调控通路，本研究对 MEagenta、MEblack 和 MEgreenyellow 模块中的差异表达基因进行了 GO 和 KEGG 富集分析。GO 富集分析表明，MEagenta 和 MEgreenyellow 模块中的差异表达基因显著富集于蛋白修饰、折叠、去泛素化以及组装等肌肉蛋白形成等相关的生物学过程(图 5)，而 MEblack 模块中的差异表达基因则富集于横纹肌组织发育调控、主轴组织形成等生物学过程

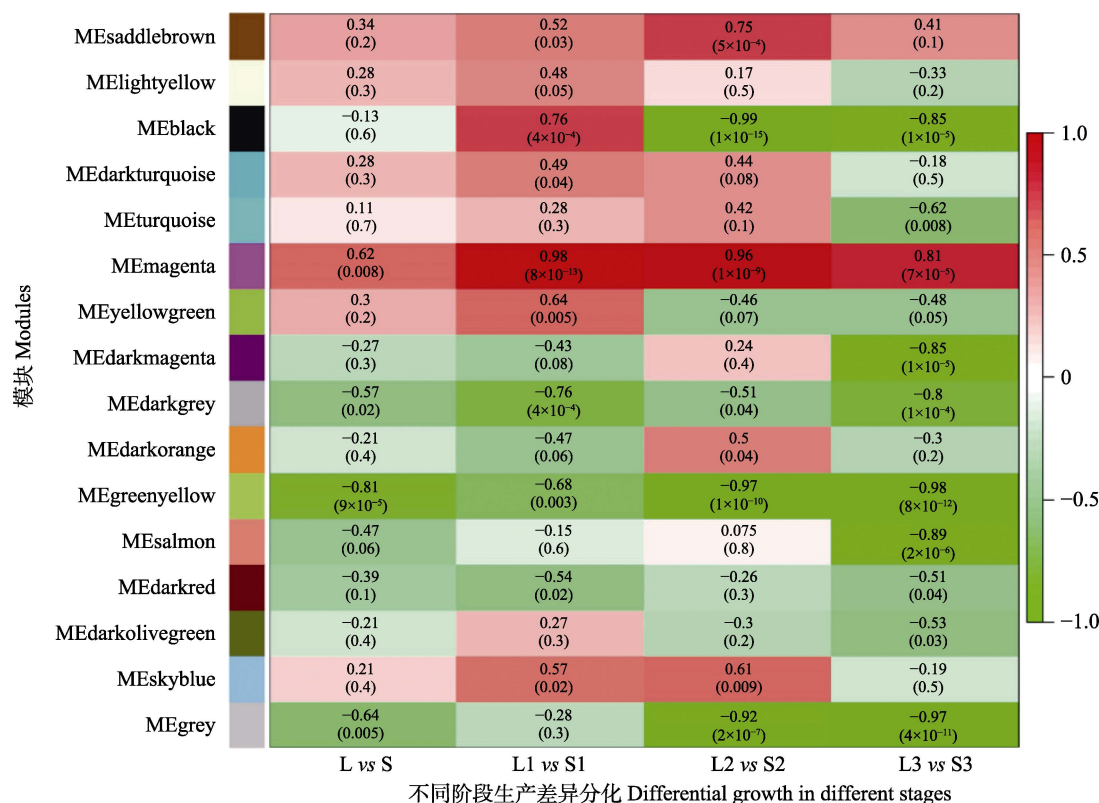


图 4 模块与性状的相关性热图

Fig.4 Heat map of the correlation between modules and traits

(图 5)。在 KEGG 富集分析结果中, MEMagenta 和 MEgreenyellow 模块中的差异表达基因显著富集在核糖体形成、蛋白酶体、肌动蛋白细胞骨架调控及 DNA 修复蛋白等与生长分化相关的通路,而 MEblack 模块中的差异表达基因则与肌细胞的肾上腺素能信号传导、肌醇磷酸代谢和泛酸盐和辅酶 a 生物合成等通路相关(图 6)。

2.4 共表达模块中核心基因的鉴定及基因互作网络构建

构建 3 个与豹纹鳃棘鲈生长差异分化特异性相关模块中的基因互作网络,将 WGCNA 分析结果中特异性模块的节点 node 和边 edge 文件导入 Cytoscape 软件,以权重和连通度排序靠前的 65 个基因(含 lncRNA)进行互作网络可视化,并计算关键性(hub)值,以 hub 值靠前的 10 个基因作为模块网络中的 hub 基因(图 7~图 9)。MEMagenta 模块中的 10 个 hub 基因分别为 *psmd6*(26S proteasome non-ATPase regulatory subunit 6)、*nmt1*(Glycylpeptide N-tetradecanoyltransferase 1)、*als2*(alsin Rho guanine nucleotide exchange factor)、*csn1*(COP9 signalosome complex subunit 1)、*nu107*(nuclear pore complex protein Nup107)、*brcc3*(Lys-63-specific deubiquitinase BRCC36)、*pyrd1*(pyridine

nucleotide-disulfide oxidoreductase domain-containing protein 1)、*rb6i2*(ELKS/Rab6- interacting/CAST family member 1)、*atg16*(autophagy-related protein 16)和 *pp2aa*(serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform),另包含 4 个 lncRNA (见图 7 中三角形节点);MEgreenyellow 模块中的 9 个 hub 基因分别为 *kank1*(KN motif and ankyrin repeat domain-containing protein 1)、*bdh2*(3-hydroxybutyrate dehydrogenase type 2)、*cra1b*(collagen alpha-1(XXVII) chain B)、*pacn3*(protein kinase C and casein kinase II substrate protein 3)、*leg*(beta-galactoside-binding lectin)、*ada12*(disintegrin and metalloproteinase domain-containing protein 12)、*ulk1*(serine/threonine-protein kinase ULK1)、*pedf*(pigment epithelium-derived factor)和 *pp2aa*(serine/threonine-protein phosphatase 2A catalytic subunit alpha isoform),1 个 hub lncRNA (MSTRG.15660)和另外 10 个 lncRNA(图 8 中三角形节点);MEblack 模块中的 8 个 hub 基因分别为 *at131*(manganese-transporting ATPase 13A1)、*atrx*(transcriptional regulator ATRX)、*xpo5*(Exportin-5)、*hnrpl*(heterogeneous nuclear ribonucleoprotein L)、*pyr1*(pyrabactin resistance 1)、*ima5*(importin subunit alpha-5)、*hdac3*(histone deacetylase 3)和 *tia1*

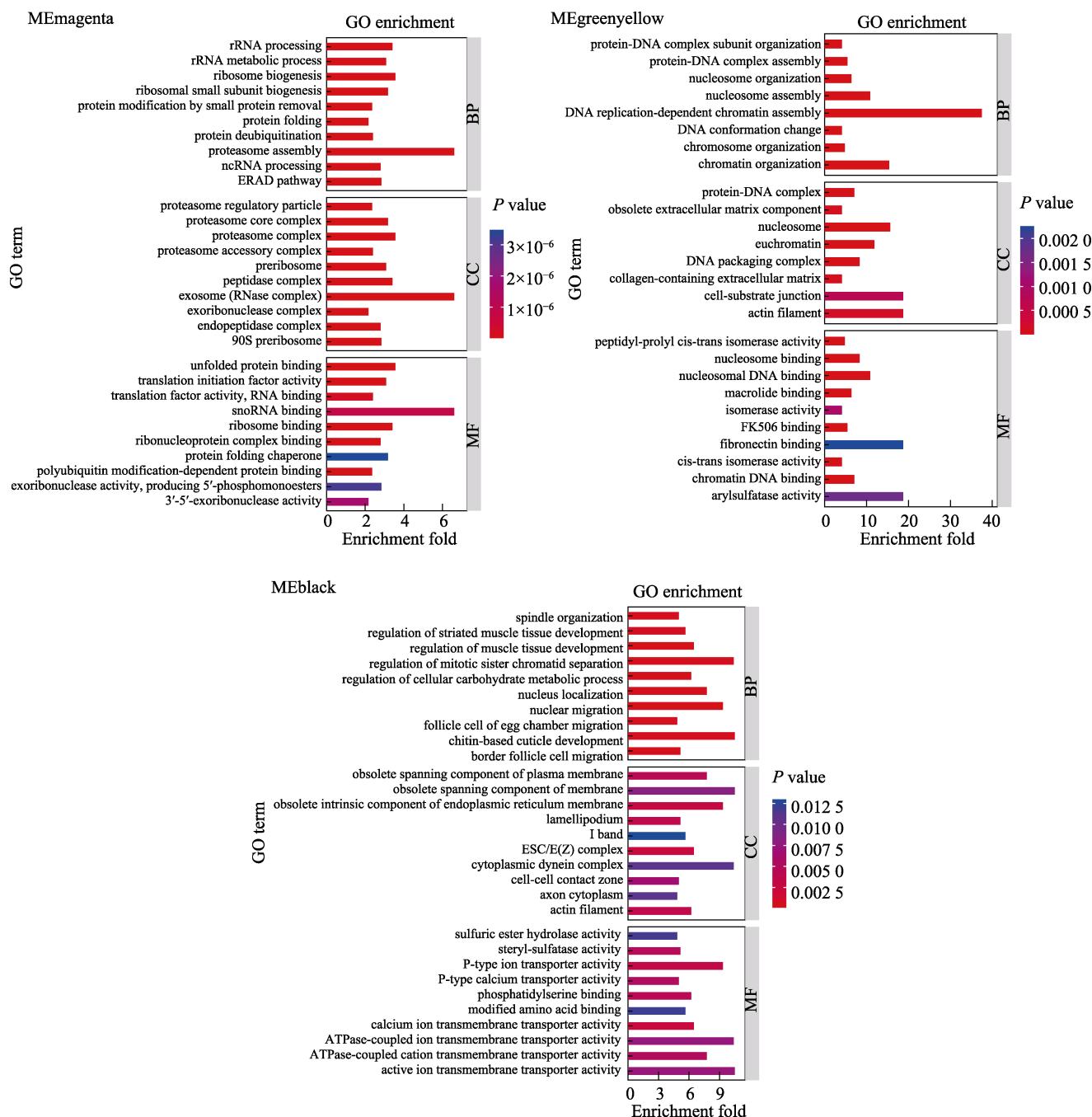


图 5 MEMagenta、MEgreenyellow 和 MEblack 模块差异表达基因 GO 富集图

Fig.5 GO enrichment diagram of differential expression genes in MEMagenta, MEGreenyellow and MEblack module

(nucleolysin TIA-1 isoform p40), 2 个 hub lncRNA (MSTRG.8694 和 MSTRG.1896) 和另外 4 个 lncRNA (图 9 中三角形节点)。

3 讨论

生长差异分化是鱼类发育过程中广泛存在的一种现象, 这种群体中相同生长发育阶段内的生长差异也是鱼类主要的个体差异之一, 受遗传因素和环境因子共同影响。豹纹鳃棘鲈的人工育苗和中间培育过程

中的个体间生长差异极其显著, 决定其差异分化的遗传机制尚不清楚。目前, Zhou 等(2024)开发了豹纹鳃棘鲈的 20K SNP array 芯片, 不仅为其重要经济性状的遗传作用机制解析提供了一种高通量的检测工具, 而且对体长和体质量指数性状进行了全基因组关联分析, 筛选到多个与生长性状显著关联的遗传位点; Wang 等(2023)利用全基因组重测序对 307 尾豹纹鳃棘鲈的体质量、全长、体长、体高和体厚等 5 个生长相关性状进行全基因组关联分析, 共定位到 12 个与

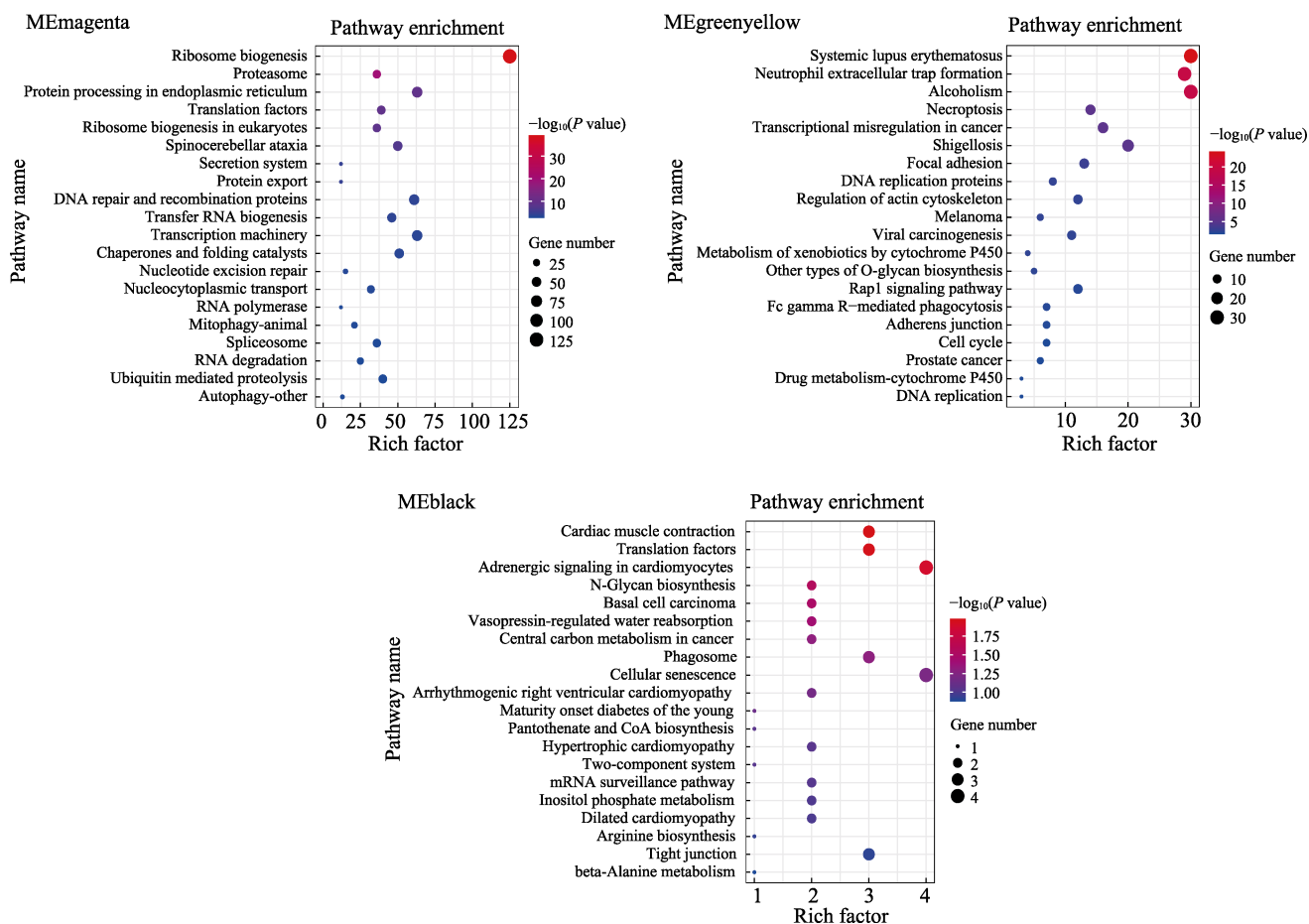


图 6 MEMagenta、MEgreenyellow 和模块差异表达基因 KEGG 富集图
Fig.6 KEGG enrichment diagram of differential expression genes in MEMagenta, MEgreenyellow and MEblack module

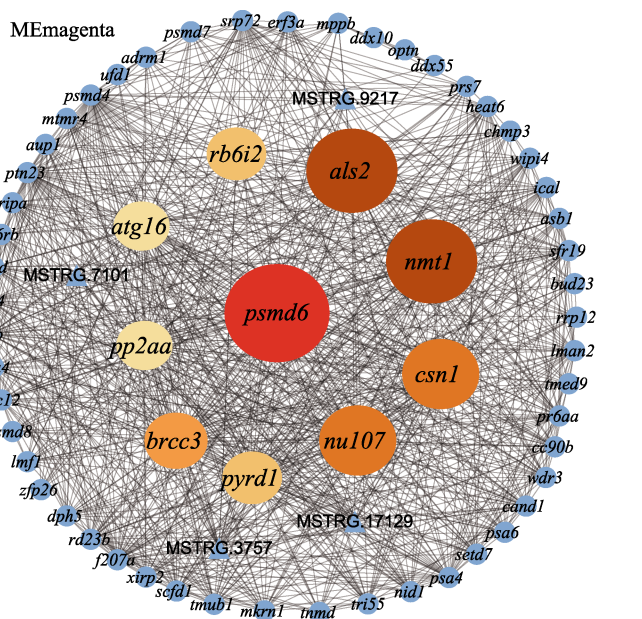


图 7 MEmagenta 模块基因的互作网络
Fig.7 Interaction network of genes in MEmagenta module

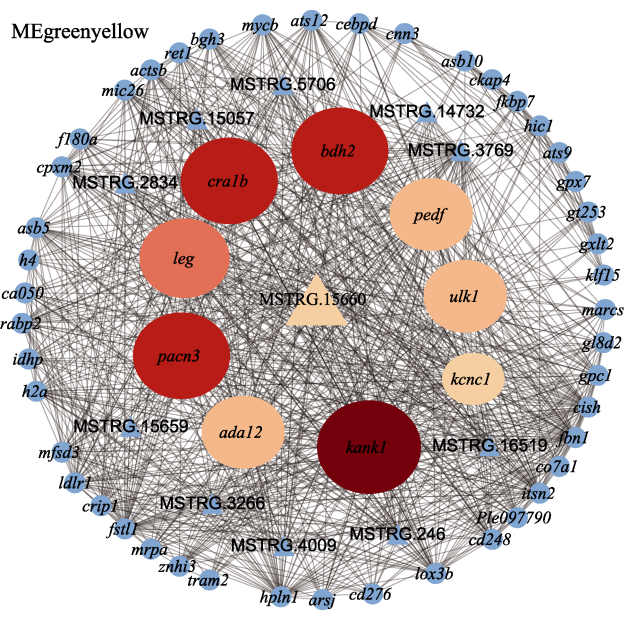


图 8 MEgreenyellow 模块基因的互作网络
Fig.8 Interaction network of genes in MEgreenyellow module

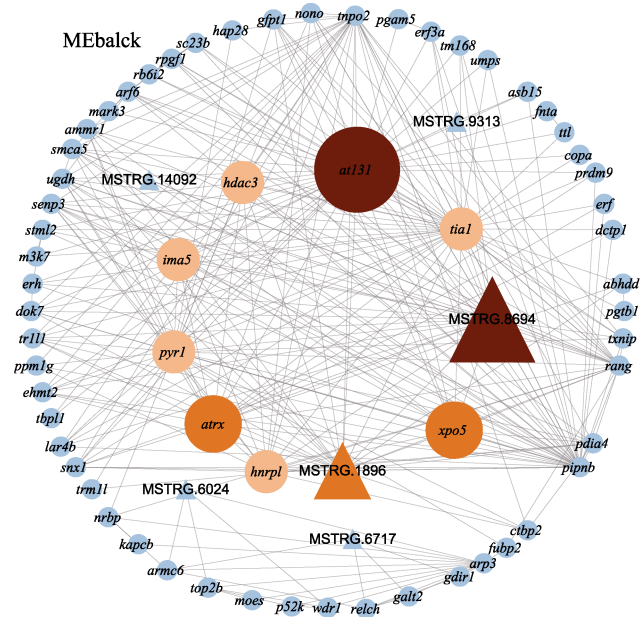


图 9 MEblack 模块基因的互作网络

Fig.9 Interaction network of genes in MEblack module

生长性状显著关联的 SNP 位点,为豹纹鳃棘鲈的分子辅助育种提供了理论基础;Yang 等(2022)基于转录组测序揭示了豹纹鳃棘鲈的不同流速下分子应答机制;Hu 等(2023)通过豹纹鳃棘鲈不同体色进行比较转录组分析解析其体色的遗传机制。上述研究中,豹纹鳃棘鲈生长性状遗传机制解析主要基于 DNA 层面的全基因组关联分析,而在 RNA 层面则主要是对抗逆和品质性状的研究。

豹纹鳃棘鲈人工育苗和中间培育过程中早期鱼苗受非遗传效应(水温、水质变化等)影响较小,且生长显著差异分化时间间隔短,适合进行遗传机制解析研究。因此,本研究选择早期鱼苗 3 个连续进行大小规格分筛的时间点(42、70 和 91 日龄)内差异分化个体的肌肉组织进行全转录组测序,并以 WGCNA 构建了豹纹鳃棘鲈生长差异分化相关 lncRNA-mRNA 加权基因共表达网络,通过网络模块与生长差异分化性状的显著关联以及模块内关键基因的深入挖掘,筛选出多个与生长差异分化相关的基因和 lncRNA。其中,MEmagenta 模块中包含差异基因和 lncRNA 数最多,模块内的核心基因 *psmd6* 在斑马鱼(*Danio rerio*)中受 *volvox* 突变体编码,通过降解多泛素化蛋白调节细胞功能影响斑马鱼晶状体上皮细胞增殖和晶状体纤维细胞分化(Imai *et al*, 2010);*nmt1* 是哺乳动物发育必需的基因,为早期胚胎发生的主要 n-肉豆蔻酰基转移酶(Yang *et al*, 2005);*als2* 基因在肌肉发育过程中发挥重要作用,该基因的突变可导致常染色体隐性肌萎缩性侧索硬化症及相关疾病(Devon *et al*,

2005);*brcc3* 作为一种细胞周期调节基因,参与磷酸化丝裂原活化蛋白激酶的产生及细胞增殖等过程(Boudreau *et al*, 2007)。值得注意的是,MEmagenta 模块基因互作网络中有多个与 *psmd6* 同一基因家族的基因,包括 *psmd4*、*psmd7* 和 *psmd8* 等,表明该基因家族可能广泛参与了豹纹鳃棘鲈的生长差异分化过程。Ahongo 等(2005)对虹鳟(*Oncorhynchus mykiss*)产卵后肉质恢复期间的肌肉进行了转录组研究,发现多个涉及蛋白酶体的 *psmd* 基因家族基因在虹鳟产卵时选择表达并于产卵后下调表达。上述研究表明,MEmagenta 模块内基因主要是通过参与肌肉细胞增殖、分化、凋亡、生长以及正常条件下应激反应等生物学过程,进而影响豹纹鳃棘鲈个体间的生长差异分化。

MEgreenyellow 模块内的核心基因 *kank1* 最初被认定为肾癌患者染色体 9p 上的一个候选肿瘤抑制基因,在肾细胞系中过表达时以抑制细胞生长,此外,*kank1* 还通过 RhoA 信号传导调控肌动蛋白聚合、肌动蛋白应激纤维形成和细胞迁移(Hensley *et al*, 2016);*ada12(adam12)* 在小鼠(*Rattus norvegicus*)肌肉和脂肪生成中起重要作用,其在斑马鱼中合成序列和氨基酸序列上与哺乳动物保持一致,并在心血管系统、红系祖细胞、脑和颌软骨中强烈表达,且 *adam12* 基因敲除的斑马鱼在幼年期表现出体型缩小,但没有明显的形态缺陷(Tokumasu *et al*, 2016)。MEblack 模块内的核心基因 *at131(atp13a1)* 在小鼠中对抗病毒蛋白 MAVS 的稳定性至关重要,*at131* 基因敲除的小鼠表现出生长迟缓和胚胎致死等问题(Yang *et al*, 2024);*hdac3* 基因为斑马鱼肝脏形成所必需,其主要作用是抑制生长分化因子 11(*gdf11*),*gdf11* 是细胞增殖的负调节因子,也是 *hdac3* 的特异性转录靶点(Farooq *et al*, 2008)。2 个特异性模块中核心基因 *ada12* 及 *at131* 的敲除分别导致斑马鱼和小鼠的生长缺陷,直接影响体型规格和生长速度,因此,可以将 *ada12* 与 *at131* 作为重要候选基因进行豹纹鳃棘鲈生长差异分化分子调控机制的深入研究。3 个特异性共表达模块中,MEgreenyellow 模块内参与了基因互作网络构建且 hub 值排序靠前的 lncRNA 最多,共有 11 个,表明该模块中的 lncRNA 相较于 MEmagenta 模块和 MEblack 模块发挥了更重要的作用。

本研究侧重分析了与豹纹鳃棘鲈生长差异分化显著相关的 3 个共表达模块,其他模块中也可能存在重要的基因以及 lncRNA 参与豹纹鳃棘鲈人工育苗和中间培育过程的生长差异分化过程。此外,本研究分析的核心基因和 lncRNA 在豹纹鳃棘鲈中的生

物学功能尚不明确,有待后续分子生物学实验进一步验证。

参 考 文 献

- AHONGO Y D, LE CAM A, MONTFORT J, *et al.* Gene expression profiling of trout muscle during flesh quality recovery following spawning. *BMC Genomics*, 2022, 23(1): 9
- BOUDREAU H E, BROUSTAS C G, GOKHALE P C, *et al.* Expression of BRCC3, a novel cell cycle regulated molecule, is associated with increased phospho-ERK and cell proliferation. *International Journal of Molecular Medicine*, 2007, 19: 29–39
- CHEN H F, TIAN K C, HUANG X X, *et al.* Construction of co-expression network of lncRNA and mRNA related to hair follicle development of Subo Merino sheep. *Scientia Agricultura Sinica*, 2019, 52(19): 3471–3484 [陈华峰, 田可川, 黄锡霞, 等. 苏博美利奴羊毛囊发育相关 lncRNA 与 mRNA 共表达网络的构建. *中国农业科学*, 2019, 52(19): 3471–3484]
- DEVON R S, SCHWAB C, TOPP J D, *et al.* Cross-species characterization of the ALS2 gene and analysis of its pattern of expression in development and adulthood. *Neurobiology of Disease*, 2005, 18(2): 243–257
- ENBERG K, JØRGENSEN C, DUNLOP E S, *et al.* Fishing-induced evolution of growth: Concepts, mechanisms and the empirical evidence. *Marine Ecology*, 2012, 33: 1–25
- FAROOQ M, SULOCHANA K N, PAN X F, *et al.* Histone deacetylase 3 (hdac3) is specifically required for liver development in zebrafish. *Developmental Biology*, 2008, 317(1): 336–353
- FLOREA L, SONG L, SALZBERG S L. Thousands of exon skipping events differentiate among splicing patterns in sixteen human tissues. *F1000 Research*, 2013, 2: 188
- FU S Y, WANG Y B, ZHENG F. Experimental study on artificial breeding of leopard striped perch in high pond. *Scientific Fish Farming*, 2009, 12: 26–27 [符书源, 王永波, 郑飞. 豹纹鳃棘鲈高位池人工育苗试验. *科学养鱼*, 2009, 12: 26–27]
- GUERRERO-COZAR I, JIMENEZ-FERNANDEZ E, BERBEL C, *et al.* Genetic estimates for growth and shape-related traits in the flatfish *Senegalese sole*. *Animals (Basel)*, 2021, 11(5): 1206
- HENSLEY M R, CUI Z B, CHUA R F S, *et al.* Evolutionary and developmental analysis reveals KANK genes were co-opted for vertebrate vascular development. *Scientific Reports*, 2016, 6: 2781
- HU Y R, CHEN K S, HUANG Y S, *et al.* Comparative transcriptome analysis of skin color-associated genes in leopard coral grouper (*Plectropomus leopardus*). *BMC Genomics*, 2023, 24: 5
- IMAI F, YOSHIKAWA A, FUJIMORI-TONOU N, *et al.* The ubiquitin proteasome system is required for cell proliferation of the lens epithelium and for differentiation of lens fiber cells in zebrafish. *Development*, 2010, 137: 3257–3268
- KANG Y J, YANG D C, KONG L, *et al.* CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Research*, 2017, 45(1): 12–16
- KIM D, LANGMEAD B, SALZBERG S L. HISAT: A fast spliced aligner with low memory requirements. *Nature Methods*, 2015, 12(4): 357–360
- LANGFELDER P, HORVATH S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*, 2008, 9(1): 559
- LI B J, JIANG D L, MENG Z N, *et al.* Genome-wide identification and differentially expression analysis of lncRNAs in tilapia. *BMC Genomics*, 2018, 19: 729
- LI C Y, JIN H M, ZHANG W, *et al.* Whole-transcriptome analysis reveals long noncoding rnas involved in female floral development of Hickory (*Carya cathayensis* Sarg.). *Frontiers in Genetics*, 2022, 13: 910488
- LOVE M I, HUBER W, ANDERS S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, 2014, 15(12): 550
- NIU J J, SUN M M, LI Z Y, *et al.* Whole transcriptome analysis provides new insight on immune response mechanism of golden pompano (*Trachinotus ovatus*) to *Amyloodinium ocellatum* infestation. *Aquaculture*, 2022, 560: 738396
- PENG S F, LI W J, CHEN Z J, *et al.* Preliminary study on artificial selection technique of fish fry in *Epinephelus lanceolatus*. *Ocean Fishery*, 2012(10): 84–85 [彭树锋, 李文娟, 陈中杰, 等. 鞍带石斑鱼鱼苗标粗技术初步研究. *海洋与渔业*, 2012(10): 84–85]
- PERTEA M, PERTEA G M, ANTONESCU C M, *et al.* StringTie enables improved reconstruction of a transcriptome from RNAseq reads. *Nature Biotechnology*, 2015, 33(3): 290–295
- SHANNON P, MARKIEL A, OZIER O, *et al.* Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, 2003, 13(11): 2498–2504
- SHI T P, ZHANG L. Application of whole transcriptomics in animal husbandry. *Hereditas (Beijing)*, 2019, 41(3): 193–205 [石田培, 张莉. 全转录组学在畜牧业中的应用. *遗传*, 2019, 41(3): 193–205]
- SONG F B, YU Z Z, DONG Z J, *et al.* Research progress of lncRNA in aquatic animals. *Journal of Fisheries of China*, 2020, 44(8): 1371–1383 [宋飞彪, 俞贞贞, 董在杰, 等. lncRNA 在水产动物中的研究进展. *水产学报*, 2020, 44(8): 1371–1383]
- SU J H, ZHANG Y P, LOU Z Y, *et al.* Progress on fish growth traits improvement and its regulation. *Sichuan Journal of Zoology*, 2012, 31(1): 165–169 [苏军虎, 张艳萍, 姜忠玉,

- 等. 鱼类生长性状改良及其调控研究进展. 四川动物, 2012, 31(1): 165–169]
- SUN L, LUO H T, BU D C, *et al.* Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic Acids Research*, 2013, 41(17): e166
- TOKUMASU Y, IIDA A, WANG Z, *et al.* ADAM12-deficient zebrafish exhibit retardation in body growth at the juvenile stage without developmental defects. *Development Growth and Differentiation*, 2016, 58: 409–421
- WANG G Y, YIN H Y, LI B Y, *et al.* Characterization and identification of long non-coding RNAs based on feature relationship. *Bioinformatics*, 2019, 35(17): 2949–2956
- WANG J L, YANG Q, HU Y R, *et al.* Identification of lncRNA-miRNA-mRNA network involved in sexual size dimorphism of Chinese Tongue Sole (*Cynoglossus semilaevis*). *Frontiers in Marine Science*, 2022, 9: 795525
- WANG T, JIANG Q H, PENG J J, *et al.* A review of the construction and analysis of gene co-expression network. *Intelligent Computer and Applications*, 2014, 4(6): 47–50 [汪涛, 蒋庆华, 彭佳杰, 等. 基因共表达网络的构建及分析方法研究综述. 智能计算机与应用, 2014, 4(6): 47–50]
- WANG T, WU X, SONG L L, *et al.* Identification of candidate growth-related SNPs and genes using GWAS and transcriptome analyses in leopard coral grouper (*Plectropomus leopardus*). *Aquaculture*, 2023, 574: 739677
- WU S J, HUANG J Q, LI Y J, *et al.* Integrated analysis of lncRNA and circRNA mediated ceRNA regulatory networks in skin reveals innate immunity differences between wild-type and yellow mutant rainbow trout (*Oncorhynchus mykiss*). *Frontiers in Immunology*, 2022, 13: 802731
- YANG M, GAO J, KE H J, *et al.* Transcriptome-based analysis of the response mechanism of leopard coral grouper liver at different flow velocities. *Fishes*, 2022, 7: 279
- YANG S H, SHRIVASTAV A, KOSINSKI C, *et al.* N-myristoyl-transferase 1 is essential in early mouse development. *Journal of Biological Chemistry*, 2005, 280(19): 18990–18995
- YANG X Y, LI T T, FANG Z Y, *et al.* ATP13A1 engages GET3 to facilitate substrate-specific translocation. *bioRxiv*, 2024, <https://doi.org/10.1101/2024.02.12.579870>
- YAO Y J, XIONG E H, QU X L, *et al.* WGCNA and transcriptome profiling reveal hub genes for key development stage seed size/oil content between wild and cultivated soybean. *BMC Genomics*, 2023, 24: 494
- YE W D, DUAN Y, ZHANG W T, *et al.* Comprehensive analysis of hub mRNA, lncRNA and miRNA, and associated ceRNA networks implicated in grass carp (*Ctenopharyngodon idella*) growth traits. *Genomics*, 2021, 113(6): 4004–4014
- YU G C, WANG L G, HAN Y Y, *et al.* clusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS*, 2012, 16(5): 284–287
- ZHANG X, CUI Y, DING X, *et al.* Analysis of mRNA-lncRNA and mRNA-lncRNA-pathway co-expression networks based on WGCNA in developing pediatric sepsis. *Bioengineered*, 2021, 12(1): 1457–1470
- ZHAO X F, LIANG L Q, LIEW H J, *et al.* Identification and analysis of long non-coding RNAs in *Leuciscus waleckii* adapted to highly alkaline conditions. *Frontiers in Physiology*, 2021, 12: 665268
- ZHOU Q, LU S, LIU Y, *et al.* Development of a 20 K SNP array for the leopard coral grouper, *Plectropomus leopardus*. *Aquaculture*, 2024, 578: 740079

(编辑 冯小花)

Weighted Gene Co-Expression Network of Growth Differentiation-Related lncRNAs and mRNAs in *Plectropomus leopardus*

GAO Jin^{1,2}, LIU Jinye^{1,3}, CHEN Fuxiao^{1,2,3}, WANG Yongbo^{1,2}, FU Shuyuan^{1,2,3①}

(1. Hainan Academy of Ocean and Fisheries Sciences, Haikou 571126, China;

2. Hainan Tropical Ocean University, Yazhou Bay Innovation Institute, Sanya 572025, China;

3. Hainan Provincial Engineering Research Center for Tropical Sea-Farming, Haikou 571126, China)

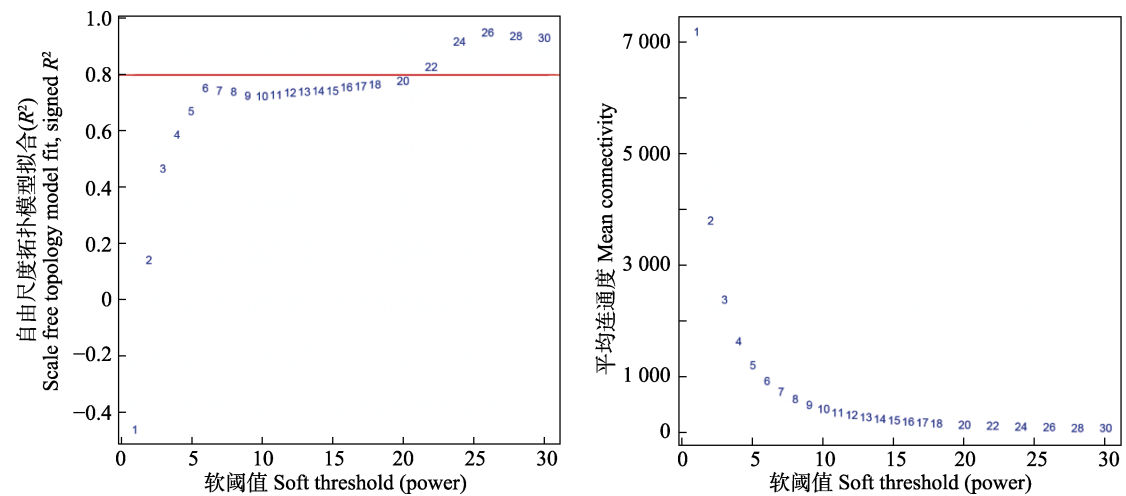
Abstract The leopard coral grouper (*Plectropomus leopardus*) exhibits captivating chromatic patterns and boasts substantial nutritional content. This species garners significant consumer preference owing to its aesthetic appeal and nutritional richness, thereby conferring considerable economic and aquacultural significance. The seedling and intermediate breeding of the *P. leopardus*, much like many other groupers,

① Corresponding author: FU Shuyuan, Email: fulank23@sohu.com

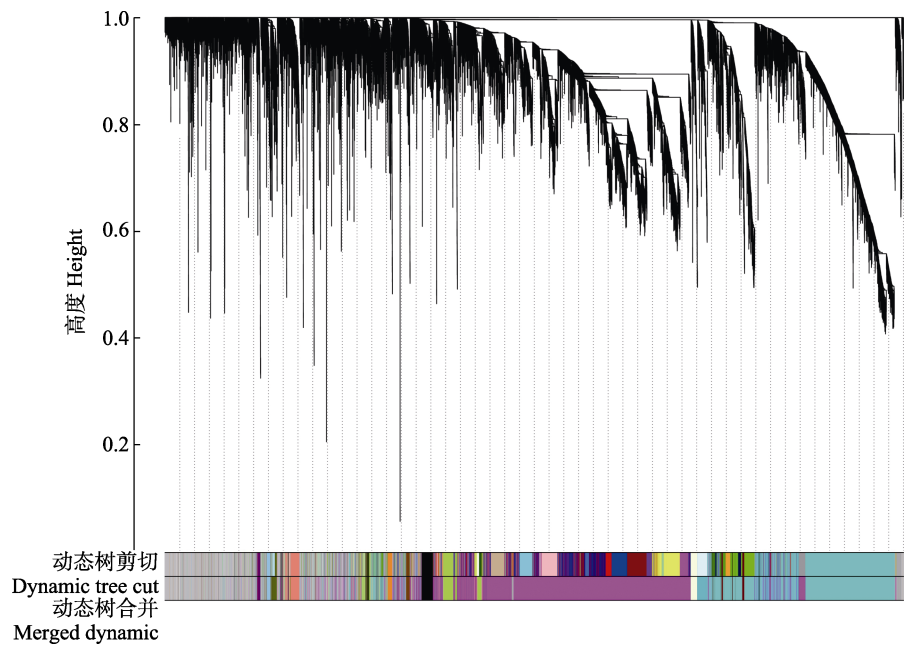
undergoes a process involving the screening of fries and acclimatization to artificial feed. During this process, repeated occurrences of growth differentiation among the same batch of fries, juveniles, and fingerlings are observed, which constitutes one of the principal factors contributing to prolonged cultivation cycles and impacting the economic efficiency of aquaculture. To address this issue, a majority of grouper species undergo a “size grading” process during seedling cultivation, wherein fries are segregated according to different size categories post-screening, thereby enhancing fry survival rates and shortening the overall cultivation cycle. However, “size grading” effectiveness is not only significantly influenced by human factors but also by inherent variations among fry themselves. Despite the considerable influence of environmental factors on this phenomenon of growth differentiation, undefined potential genetic factors remain. Yet, there is a noticeable absence of comprehensive research addressing this pressing issue. However, the quality of growth traits serves as a critical indicator for evaluating the economic value of fish aquaculture. Influenced by both genetic predispositions and environmental factors, these traits warrant investigation into the genetic mechanisms and regulatory strategies governing growth-related characteristics in farmed fish. Such research offers insights into variety improvement and breeding practices within the field. Here, during the artificial breeding and intermediate rearing process of *P. leopardus*, early juveniles exhibited relatively minor susceptibility to non-genetic effects such as water temperature and quality changes, along with significantly differentiated growth and shorter intervals of divergence. Consequently, we selected muscle tissues of individuals exhibiting differential divergence within three consecutive time points of size screening for early juveniles (42, 70 and 91 days post-hatching) for whole transcriptome sequencing, aiming to elucidate genetic mechanisms. The expression profiles of lncRNAs and mRNAs were acquired, and a Weighted Gene Co-expression Network Analysis (WGCNA) method was used to construct a co-expression network of growth differentiation related lncRNAs and mRNAs. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment analyses of differentially expressed genes within specific gene modules were conducted to analyze the biological functions of the modules. Additionally, Cytoscape software was used to construct a gene interaction networks to identify key candidate genes associated with growth differential differentiation in *P. leopardus*. The results showed that a total of 6,252 differentially expressed mRNAs and 367 differentially expressed lncRNAs were identified through gene differential expression analysis. Co-expression network analysis revealed three growth differentiation specific gene modules (ME_{magenta}, ME_{greenyellow}, and ME_{black}). GO and KEGG analyses indicated that differentially expressed genes in these specific modules were significantly enriched in biological processes related to muscle protein formation, regulation of skeletal muscle tissue development, and axial tissue formation, participating in metabolic pathways associated with growth differential differentiation such as proteasome, cytoskeletal regulation by actin proteins, and inositol phosphate metabolism. High-weight and high-connectivity lncRNAs and mRNAs were screened from the three specific modules to construct a gene interaction network, revealing key genes associated with growth differentiation including *psmd6*, *nmt1*, *als2*, *brcc3*, *kank1*, *ada12*, *at131*, *hdac3*, as well as 20 lncRNA transcription factors such as *MSTRG.15660*, *MSTRG.8694*, and *MSTRG.1896*. Wherein, *Psmd6* in zebrafish is encoded by the Volvox mutant, which regulates cellular function by degrading polyubiquitinated proteins, impacting the proliferation of zebrafish lens epithelial cells and the differentiation of lens fiber cells. *Mmt1* is an essential gene for mammalian development, serving as a major N-myristoyltransferase during early embryogenesis. *Als2* plays a significant role in normal muscle development, and mutations in this gene can lead to autosomal recessive amyotrophic lateral sclerosis and related disorders. The *Brcc3*, as a cell cycle regulatory gene, participates in the generation of phosphorylated cyclin-dependent kinase activator and biological processes such as cell proliferation. *Kank1* regulates actin polymerization, actin stress fiber formation, and cell migration through RhoA signaling. *Ada12* (*Adam12*) plays a crucial role in murine

myogenesis and adipogenesis. *Adam12* synthesis sequence and amino acid sequence in zebrafish are consistent with mammals and is strongly expressed in the cardiovascular system, erythroid progenitor cells, brain, and jaw cartilage. Zebrafish with knocked out *Ada12* genes exhibited reduced body size in infancy without significant morphological defects. *At131* (*Atp13a1*) plays a crucial role in stabilizing the MAVS virus protein in mice. *At131* knockout in mice resulted in issues such as growth retardation and embryonic lethality. *Hdac3* is essential for liver formation in zebrafish, primarily by inhibiting the growth differentiation factor 11 (*Gdf11*), which is a negative regulator of cell proliferation and a specific transcription target of *Hdac3*. *Ada12* and *At131* knockout leads to defects in growth in zebrafish and mice, directly affecting body size and growth rate. Therefore, *Ada12* and *At131* are considered as important candidate genes for further research on the molecular regulatory mechanisms of growth differentiation in *P. leopardus*. This study provides insights into the molecular mechanisms underlying growth differentiation in *P. leopardus*.

Key words *Plectropomus leopardus*; Growth differentiation; WGCNA; Co-expression network



附图1 软阈值的筛选
Fig.S1 Screening of soft threshold (power)



附图2 基因聚类树及共表达模块构建
Fig.S2 Clustering dendrograms of genes and co-expression module division