



亚洲草鱼群体的差异甲基化分析

魏 上^{1,2}, 谢玲莉^{1,3,4}, 朱 华⁵, 张清靖⁵, 沈玉帮^{1,3,4},
徐晓雁^{1,3,4*}, 李家乐^{1,3,4*}

(1. 上海海洋大学, 农业农村部淡水水产种质资源重点实验室, 上海 201306;

2. 上海海洋大学, 水产种质资源发掘与利用教育部重点实验室, 上海 201306;

3. 上海海洋大学, 水产养殖工程研究中心, 上海 201306;

4. 上海海洋大学, 水产科学国家级实验教学示范中心, 上海 201306;

5. 北京市农林科学院水产科学研究所, 渔业生物技术北京市重点实验室, 北京 100068)

摘要: 为探究水产养殖过程中 DNA 甲基化对亚洲草鱼群体人工驯化与环境选择适应的影响, 本实验对亚洲范围内 8 个地区的养殖和野生草鱼群体进行了全基因组甲基化测序, 测序序列经质控后比对到草鱼参考基因组上, 并进行差异甲基化分析和功能富集分析。结果显示, 甲基化测序共获得 308.15 Gbp 测序数据, 平均测序深度 31×, 平均比对率 62.97%。对 5 个养殖群体分别与野生背景群体间进行差异甲基化分析, 共发掘出 76 422 个差异甲基化位点、3 737 个差异甲基化区域、1 950 个差异甲基化基因。功能分析发现, 差异甲基化基因被注释到血管生成、神经嵴细胞迁移、T 细胞分化、颅骨系统发育、肌肉细胞发育等 GO 功能分类中。KEGG 代谢途径富集分析的结果显示, 差异甲基化基因主要参与细胞黏附连接、Notch 信号通路和 Wnt 信号通路等代谢途径。在这些功能注释的基因中, 有许多与神经、免疫、骨骼和肌肉等组织发育相关的重要基因, 如 *sema3d*、*sema5bb* 和 *nrg2a* 等。本研究进一步对表观遗传修饰在草鱼人工驯化和环境选择适应作用机制的探究奠定了基础, 为保存和科学利用草鱼种质资源提供了有价值的遗传基础数据。

关键词: 草鱼; 基因组; 甲基化测序; 差异甲基化

中图分类号: Q 347; S 965.112

文献标志码: A

世界上最早的水产养殖起源于我国, 并随着发展逐渐替代渔业捕捞成为水产品的主要供应来源^[1]。对水产养殖品种的驯化选育过程涉及在野外捕获成鱼个体, 并将其转移至与自然栖息地有很大不同的养殖环境中进行密集圈养^[2]。在这一过程中, 养殖鱼类不仅受到来自人类的生产性状选育方面的选择压力, 同时也经受着来自养殖地区气候环境等方面的选择压力, 使遗传多样性下

降和表型改变, 并留下多种可遗传的基因组特征^[3-4]。

以 DNA 甲基化为代表的表观遗传修饰是一类重要的变异来源^[5-6]。作为生物体内普遍发生的一种反应和修饰过程, 甲基化不改变 DNA 序列信息, 而是通过向 DNA 序列的胞嘧啶上添加甲基基团, 影响基因调控区与多种蛋白复合体的亲和力, 进而广泛参与维持基因组稳定与基因表达调控等方面^[7]。DNA 的甲基化水平与生物体内如基因表

收稿日期: 2022-02-25 修回日期: 2022-04-30

资助项目: 国家重点研发计划“蓝色粮仓科技创新”项目 (2020YFD0900103); 现代农业产业技术体系专项 (CARS-45-03); 上海市工程中心提升项目 (19DZ2284300)

第一作者: 魏上 (照片), 从事鱼类种质资源与种苗工程研究, E-mail: shangwei0126@163.com

通信作者: 徐晓雁, 从事鱼类种质资源与种苗工程研究, E-mail: xyxu@shou.edu.cn;

李家乐, 从事鱼类种质资源与种苗工程研究, E-mail: jlli2009@126.com



达调节与沉默、X 染色体失活、基因组印记和转座子活性抑制等多种遗传活动紧密相关^[8-9]。因此 DNA 甲基化同样可以促进产生恒定的表型改变并遗传给下一代^[10-13]。此外与 DNA 遗传变异相比, 甲基化容易受到如温度等环境因素的影响^[14-16]。特别是鱼类等低等脊椎动物的受精卵暴露在水中, 更容易受到外部环境的影响, 并且在这一时期普遍缺乏表观遗传重编程过程^[17-18]。因此, DNA 甲基化可能作为一类重要的变异来源, 在水产养殖鱼类对选择压力的适应过程中发挥了重要作用^[19-21]。目前已在其他多种养殖鱼类中对此进行了研究, 如已有的一项研究在罗非鱼野生亲本与养殖子代之间发掘出约 700 个差异甲基化位点, 参与肌肉生长、免疫、自噬、响应摄食、RNA 甲基化和组蛋白修饰等多种途径^[22]。一项针对经过 25 年选择性育种的舌齿鲈 (*Dicentrarchus labrax*) 的研究, 发掘出与解剖结构发育、胚胎发育调节、神经嵴细胞迁移等多个发育相关的差异甲基化基因^[23]。这些研究为水产养殖过程中鱼类驯化选择的表观遗传机制提供了新的见解。

草鱼 (*Ctenophyngodon idella*) 在中国的养殖已有 1 700 多年的历史, 自 20 世纪 80 年代以来已被直接或间接引种到美国、印度和墨西哥等 100 多个国家^[24-25]。作为一种广温性鱼类, 草鱼具有很强的地区适应性^[25-27]。此前一项基于重测序的在 9 个地区的 197 尾草鱼的研究, 发掘出与生物调节、细胞过程、发育过程、代谢过程、色素沉积和对

刺激的反应等多个功能相关的基因^[24]。该研究与之前的几项研究发现, 许多地区的养殖草鱼群体的遗传多样性降低, 但依然具有较强的环境适应性^[24, 28-29]。因此, 草鱼较强的环境适应性可能更多地来源于表观遗传层面。然而, 关于研究草鱼群体全基因组甲基化, 及其对水产养殖过程中人工驯化与环境选择适应有何影响, 迄今鲜见报道。

本研究利用全基因组 DNA 甲基化测序技术对亚洲范围内不同地区的 8 个养殖和野生草鱼群体进行差异甲基化分析, 旨在从表观遗传层面分析 DNA 甲基化对草鱼在水产养殖过程中人工驯化与环境选择适应的促进作用, 为其繁育、抗病以及生长性状方面的养殖实践与育种工作提供新的方向。

1 材料与amp;方法

1.1 实验材料

采集于亚洲 8 个地点的草鱼群体, 每个群体取 10 个个体, 共 80 尾草鱼样本进行全基因组 DNA 甲基化测序。8 个群体中, 5 个来自中国, 包括天津、万州、沅江、安兴和肇庆, 另外 3 个分别来自越南的 Bac Ninh、印度的 Gorakhpur 和尼泊尔的 Hetauda。其中, 中国万州、沅江和肇庆的群体为野生群体, 采样于当地野外河流。其余 5 个群体均为养殖群体, 采样自当地水产养殖场或原种场 (表 1)。

表 1 草鱼样本采样信息表

Tab. 1 Summary of sampling information of *C. idella*

群体 populations	采样经纬度 latitude, longitude	样本数量/尾 individual no.	群体类型 group types	所属水系 river system
尼泊尔 Nepal	27°25'12"N, 85°1'48"E	10	养殖	珠江水系
印度 India	26°45'36"N, 83°22'12"E	10	养殖	珠江水系
越南 Vietnam	21°7'12"N, 105°58'48"E	10	养殖	珠江水系
天津 Tianjin	39°22'48"N, 117°3'E	10	养殖	珠江水系
安兴 Anxing	29°24'N, 112°7'48"E	10	养殖	长江水系
万州 Wanzhou	30°49'48"N, 108°27'E	10	野生	长江水系
沅江 Yuanjiang	28°51'N, 112°21'36"E	10	野生	长江水系
肇庆 Zhaoqing	23°4'48"N, 112°31'48"E	10	野生	珠江水系

1.2 文库构建与 DNA 甲基化测序

采样后, 剪取鱼鳍组织通过裂解法进行 DNA 提取, 通过超声波破碎法将 DNA 片段化。片段化的 DNA 经过纯化、末端修复、3'端加 A、加甲基化测序接头, 经重亚硫酸盐 (bisulfite) 处理并进行

PCR 扩增后, 完成 DNA 甲基化测序文库的构建, 于 Hiseq XTen 测序平台进行测序。

本研究获得了上海海洋大学实验动物管理和使用伦理委员会批准 (SHOU-DW-2022-029), 实验过程中操作人员严格遵守上海海洋大学实验动物相关伦理规范, 并按照上海海洋大学伦理委员

会制定的规章制度执行。

1.3 测序数据质控与序列比对

首先对下机数据切去接头序列, 并根据 barcode 分离样本得到不同样品的原始序列。通过 FastQC 软件^[30]对原始序列进行质量评估, 并通过 Trimomatic 软件^[31]在进行下游分析之前进行预处理, 以去除低质量的序列: ①切去位于序列头部的碱基比例异常的碱基; ②切去位于序列头尾的低质量碱基; ③以 4 个碱基长度对序列进行滑窗, 并切去平均测序质量低于 15 的后续碱基; ④去除经前几步过滤后长度低于 70 bp 的序列。最后再次通过 FastQC 软件对新序列进行质量评估, 以验证过滤效果。

通过 Bismark 软件^[32]中的 bismark_genome_preparation 工具, 对经过本实验室重新拼接的草鱼参考基因组^[33]构建基因组索引, 并通过 Bowtie2 软件^[34]将经过预处理的甲基化测序序列与基因组进行比对生成原始 bam 文件。然后通过 Bismark 中的 deduplicate_bismark 工具对甲基化测序 bam 文件进行了 PCR 去重复。最后, 通过 Bismark 中的 bismark_methylation_extractor 工具进行全基因组范围内的甲基化位点发掘。

1.4 差异甲基化分析

本研究将万州、沅江和肇庆 3 个野生群体合并为野生背景群体, 与印度、尼泊尔、天津、安兴与越南 5 个养殖群体通过使用 R 包 DSS^[35]进行差异甲基化分析。将平均甲基化率差异 ≥ 0.2 以及 $P\text{-value} < 0.05$ 的 CpG 位点标记为差异甲基化位点 (differentially methylated loci, DML)。把平均甲基化率差异 ≥ 0.2 、 $P\text{-value} < 0.05$ 并且 50 bp 内多于 3 个 CpG 位点的基因组区域标记为差异甲基化区域 (differentially methylated region, DMR)。并将

基因本体连同上游 1 kb 区域与差异甲基化区域两侧 1 kb 范围内相交的基因标记为差异甲基化基因 (differentially methylated gene, DMG)。对差异甲基化基因的功能注释和富集分析通过在线网站 KOBAS(<http://kobas.cbi.pku.edu.cn/kobas3>)、以斑马鱼 (*Danio rerio*) 为模式物种来完成。结果通过 Circos 软件^[36]和 R 语言软件包 ggplot2^[37]进行相应可视化。

2 结果

2.1 甲基化测序信息统计

本研究对 8 个草鱼群体共 80 个样本进行了全基因组甲基化测序, 共获得 308.15 Gb 测序数据。经过滤后, 平均每群体获得 128 396 030 条质控序列用于后续分析, 质控序列的平均测序深度为 $32.22 \times$, 测序质量值大于 Q30 的碱基比例平均为 90.89%, 序列 GC 含量平均为 18.72%。将质控后的序列比对到草鱼参考基因组上的比对率平均为 62.97% (表 2)。

2.2 全基因组甲基化水平分析

共计 8 个群体的 DNA 甲基化数据分析显示, 草鱼约 900 Mb 基因组基础上, 共发掘出潜在的甲基化位点 277 618 513 个, 其中平均每个群体的 CpG 类型位点约 26 010 547 个, 占总数的 8.42%; CHG 类型位点约 57 911 405 个, 占总数的 20.04%; CHH 类型位点约 193 696 561 个, 占总数的 71.54% (表 3)。发现 CpG 类型位点的甲基化程度最高, 平均为 79.57%; 其次是 CHG 类型位点, 甲基化程度平均为 0.69%; CHH 类型位点的甲基化程度最低, 平均为 0.68% (表 3)。在 8 个群体中, 尼泊尔群体的 3 种类型总体的甲基化程度最高, 为 7.57%; 安兴群体最低, 为 7.09%。所有草鱼个体的总体平均甲基化率约为 7.35% (表 3, 图 1)。

表 2 草鱼群体的全基因组甲基化测序信息统计表

Tab. 2 Summary of whole genome bisulfite sequencing of the *C. idella* populations

群体 populations	质控序列/条 clean reads	测序深度($\times$) depth ($\times$)	Q30比例/% Q30 percent age	GC含量/% GC content	比对率/% mapping rate
尼泊尔 Nepal	146 696 131	40.44	90.75	19.30	69.57
印度 India	123 079 682	33.11	90.91	18.50	67.85
越南 Vietnam	125 797 016	32.54	92.49	18.78	65.40
天津 Tianjin	144 343 439	34.39	90.42	18.61	60.28
安兴 Anxing	108 333 150	19.01	89.36	18.70	44.21
万州 Wanzhou	116 765 232	30.10	89.93	18.70	64.87
肇庆 Zhaoqing	132 233 009	34.09	90.96	18.66	65.23
沅江 Yuanjiang	129 920 581	34.07	92.28	18.54	66.36
平均 average	128 396 030	32.22	90.89	18.72	62.97

表 3 草鱼群体的甲基化相关信息统计表

Tab. 3 Overview of identified DNA methylation of the *C. idella* populations

群体 populations	CpG			CHG			CHH			总甲基 化率/% total methylation rate
	位点数/个 loci no.	占比/% percentage	甲基化率/% methylation rate	位点数/个 loci no.	占比/% percentage	甲基化率/% methylation rate	位点数/个 loci	占比/% percentage	甲基化率/% methylation rate	
尼泊尔 Nepal	26 286 027	8.64	80.57	58 412 108	20.36	0.67	195 071 487	71.00	0.66	7.57
印度 India	26 098 266	8.25	79.84	58 086 808	19.91	0.59	194 149 297	71.83	0.60	7.14
越南 Vietnam	26 267 205	8.58	78.61	58 376 666	20.12	0.71	195 081 390	71.30	0.70	7.39
天津 Tianjin	26 281 381	8.42	79.23	58 420 727	19.97	0.67	195 197 442	71.61	0.66	7.27
安兴 Anxing	25 452 413	8.19	79.03	56 983 858	20.00	0.69	191 131 229	71.80	0.66	7.09
万州 Wanzhou	25 116 787	8.46	79.84	56 130,558	20.00	0.69	188 387 352	71.53	0.68	7.38
肇庆 Zhaoqing	26 290 093	8.41	80.24	58 437 003	19.95	0.85	195 248 595	71.63	0.86	7.53
沅江 Yuanjiang	26 292 207	8.40	79.16	58 443 510	20.00	0.66	195 305 692	71.60	0.66	7.26
平均 average	26 010 547	8.42	79.57	57 911 405	20.04	0.69	193 696 561	71.54	0.68	7.35

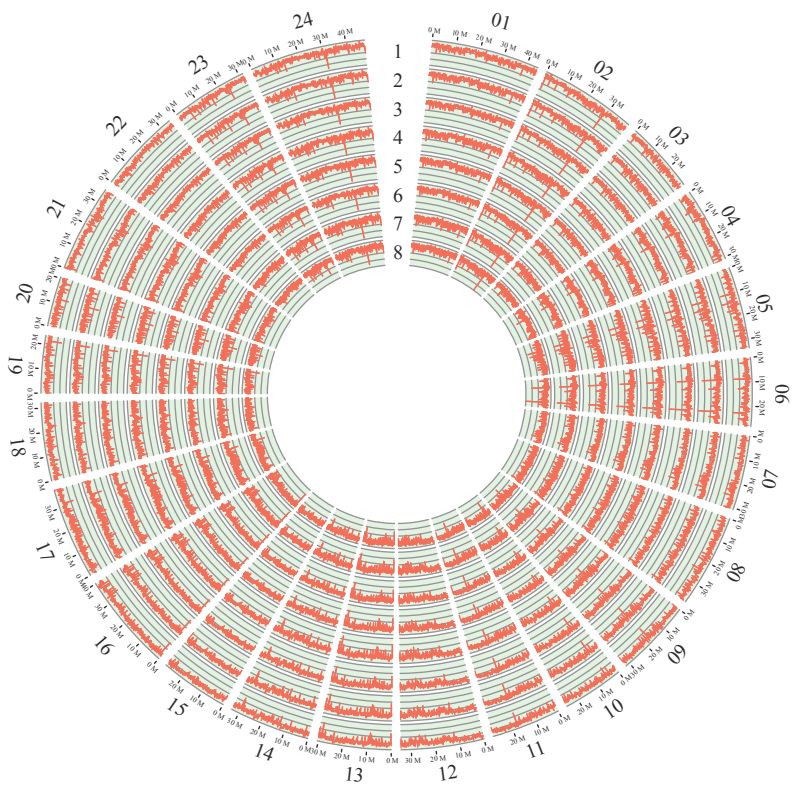


图 1 草鱼群体的全基因组甲基化水平分布图

1. 印度, 2. 尼泊尔, 3. 越南, 4. 肇庆, 5. 天津, 6. 安兴, 7. 沅江, 8. 万州; 最外圈的 01~24 是草鱼染色体编号; 下同。

Fig. 1 Genomic landscape of DNA methylation level of the *C. idella* populations

1. India, 2. Nepal, 3. Vietnam, 4. Zhaoqing, 5. Tianjin, 6. Anxing, 7. Yuanjiang, 8. Wanzhou; the outermost numbers 01-24 indicate the chromosome number; the same below.

2.3 群体间差异甲基化分析

在 CpG 类型位点进行的群体间差异甲基化分析中, 将万州、沅江和肇庆这 3 个野生群体合并为野生背景群体组合, 其余 5 个养殖群体均分别与此野生背景群体组合进行了比较, 来分析水产养殖条件与自然生长条件对草鱼表观遗传学所产生的不同影响。分析中, 共检测到 76 422 个 DML, 其中超过五分之二 (41.56 %) 是在印度群体中发现的, 所占比例最高; 而在越南群体中发现的 DML 数量最少。根据这些 DML, 进一步发掘出了 3 737 个 DMR, 总长度为 0.85 Mb, 约占草鱼

基因组的 0.09 %。同样, 接近一半数量 (46.53 %) 的 DMR 存在于印度群体中, 所占比例最高 (表 4, 图 2)。位于 DMR 两侧各 1 kb 范围内共发掘出 1 950 个 DMG (表 4)。

2.4 DMG 的功能注释和富集分析

本研究对 DMG 进行了 GO 功能分类和 KEGG 通路注释和富集分析, 共有 73 个 GO 功能分类和 8 个 KEGG 通路有显著的富集 (表 5)。其中, 印度群体的 DMG 富集到的 GO 功能分类数量, 占有这些分类数量的一半, 为 53.42 %。而在其他群体中, 富集到的 GO 功能分类数量都明显较少,

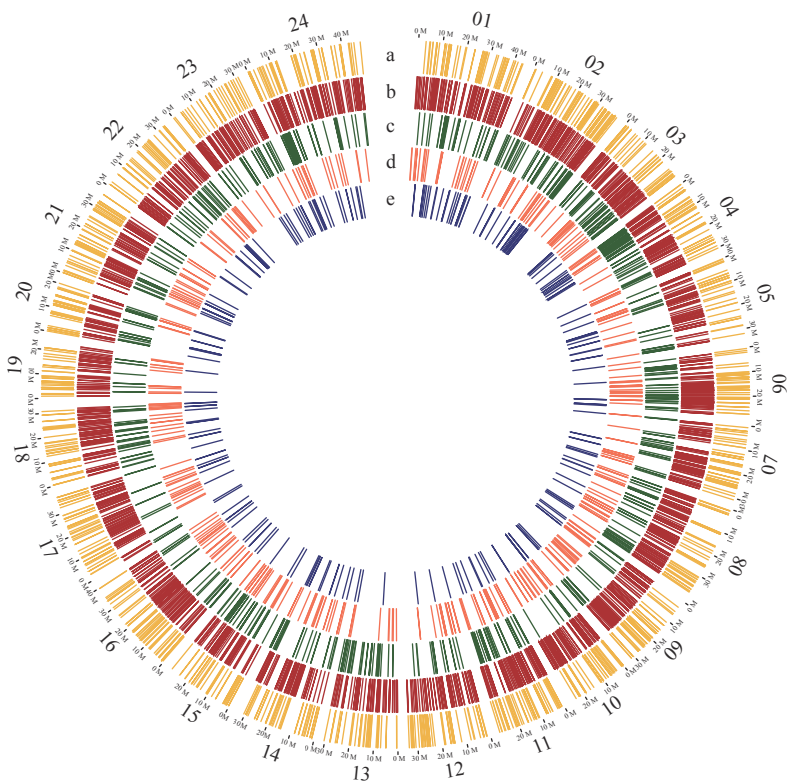


图 2 草鱼群体的差异甲基化区域分布图

a. 安兴 vs. 野生, b. 印度 vs. 野生, c. 尼泊尔 vs. 野生, d. 天津 vs. 野生, e. 越南 vs. 野生。

Fig. 2 Circos of DMR distribution of the *C. idella* populations

a. Anxing vs. wild, b. India vs. wild, c. Nepal vs. wild, d. Tianjin vs. wild, e. Vietnam vs. wild.

表 4 草鱼群体的差异甲基化相关信息统计表

Tab. 4 Summary of differential methylation information of the *C. idella* populations

尾

差异甲基化信息 differentially methylated information	安兴 vs. 野生 Anxing vs. wild	印度 vs. 野生 India vs. wild	尼泊尔 vs. 野生 Nepal vs. wild	天津 vs. 野生 Tianjin vs. wild	越南 vs. 野生 Vietnam vs. wild	总计 total
DML	17 738	31 763	14 496	7 556	4 869	76 422
DMR	862	1 739	554	371	211	3 737
DMG	442	944	281	184	99	1 950

表 5 草鱼群体的差异甲基化基因 GO 功能分类与 KEGG 富集信息统计表

Tab. 5 Summary of enriched GO terms and KEGG pathways of DMGs of the *C. idella* populations

	项目 items	安兴 vs. 野生 Anxing vs. wild	印度 vs. 野生 India vs. wild	尼泊尔 vs. 野生 Nepal vs. wild	天津 vs. 野生 Tianjin vs. wild	越南 vs. 野生 Vietnam vs. wild	总计 total
GO功能分类 GO term	数量/尾 no.	28	39	0	0	6	73
	占比/% percent	38.36	53.42	0.00	0.00	8.22	100.00
KEGG通路 KEGG pathway	数量/尾 no.	0	8	0	0	0	8
	占比/% percent	0.00	100.00	0.00	0.00	0.00	100.00

在尼泊尔和天津群体没有显著富集的 GO 功能分类。而在 KEGG 通路分析中, 只有印度群体的差异甲基化基因富集到 8 个 KEGG 通路(表 5)。

在安兴、印度和越南群体中显著富集的生物途径类 GO 功能分类中, 均主要与系统发育相关。安兴群体富集到的 14 个生物途径类 GO 功能分类中有 9 个与系统发育相关(图 3-a), 主要包括血管生成 (GO:0001525)、肝脏发育 (GO:0001889)、多细胞组织发育 (GO:0007275)、前后体轴模式规范 (GO:0009952)、细胞分化 (GO:0030154)、胚胎骨骼系统形态发生 (GO:0048704)、T 细胞分化 (GO:0030217)、内分泌腺发育 (GO:0031018)、肌肉细胞发育 (GO:0055001) 等。印度群体富集到的 22 个生物途径类 GO 功能分类中, 有 15 个与系统发育相关(图 3-b), 主要包括神经嵴细胞迁移 (GO:0001755)、多细胞组织发育 (GO:0007275)、神经嵴细胞分化 (GO:0014033)、神经嵴细胞命运规范 (GO:0014036)、脊索动物胚胎发育 (GO:0043009)、解剖结构发育 (GO:0048856)、胚胎面部颅骨的形态发生 (GO:0048703)、颅骨系统发育 (GO:1904888)、神经系统发育 (GO:0007399)、轴突导向 (GO:0007411)、树突发育 (GO:0016358)、突触的组织组装或分解 (GO:0050808)、信号素-丛素信号通路 (GO:0071526)、细胞黏附 (GO:0007155)、鳍再生 (GO:0031101) 等。而越南群体富集到的 5 个生物途径类 GO 功能分类中有 3 个与系统发育相关(图 3-c), 主要包括多细胞组织发育 (GO:0007275)、前后体轴模式规范 (GO:0009952) 和胚胎骨骼系统形态发生 (GO:0048704) 等。

在印度群体中显著富集的 8 个 KEGG 通路主要包括 Notch 信号通路 (dre04330)、黏着连接 (dre04520)、泛素介导的蛋白质水解 (dre04120)、肌动蛋白细胞骨架的调节 (dre04810)、Wnt 信号通路 (dre04310)、黏着斑 (dre04510)、紧密连接 (dre04530)、胞吞作用 (dre04144) 等(图 4)。

此外, 在印度群体富集到的 15 个与系统发

育相关的 GO 功能分类中, 有 3 个直接或间接与神经嵴细胞正常发育有关, 包括神经嵴细胞迁移 (GO:0001755)、神经嵴细胞命运规范 (GO:0014036) 和神经嵴细胞分化 (GO:0014033)。其中神经嵴细胞迁移 GO 功能分类包含 9 个 DMG: *atp6v0a1a*、*chd7*、*col18a1a*、*masp1*、*nrg2a*、*pard3ab*、*plod3*、*sema3d* 和 *sema5bb* (图 5-a)。神经嵴细胞命运规范 GO 功能分类包含 3 个 DMG: *tfap2a*、*tfap2c* 和 *kctd15b* (图 5-b)。而神经嵴细胞分化 GO 功能分类包含 3 个 DMG: *pard3ab*、*sema3d* 和 *alcama* (图 5-c)。

3 讨论

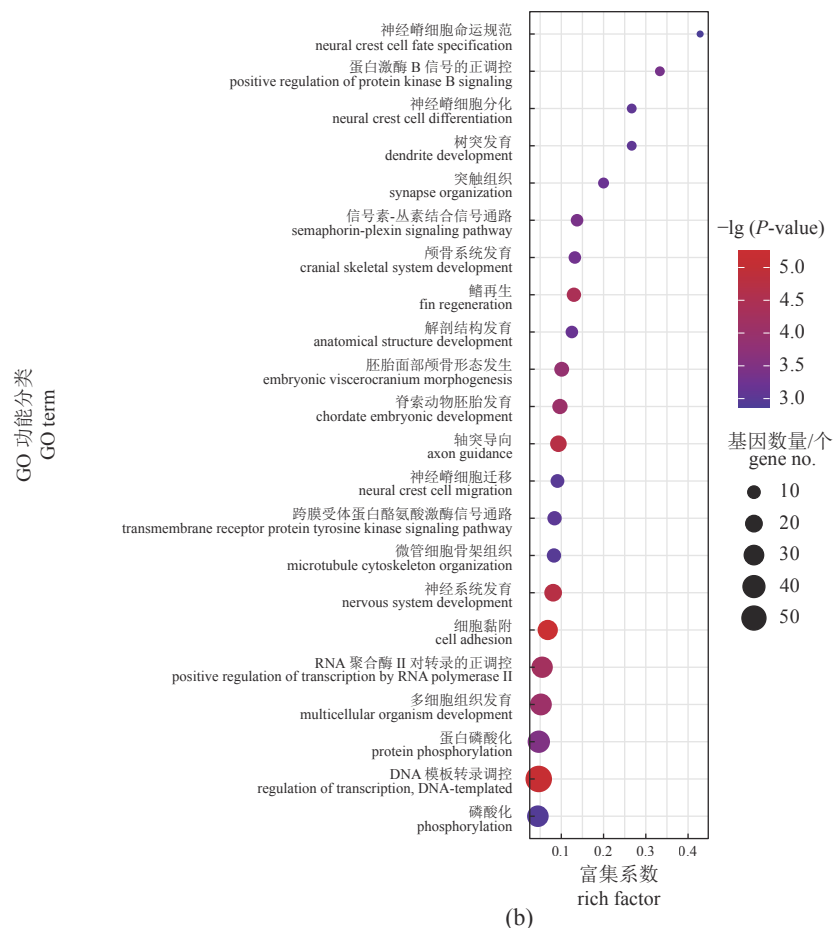
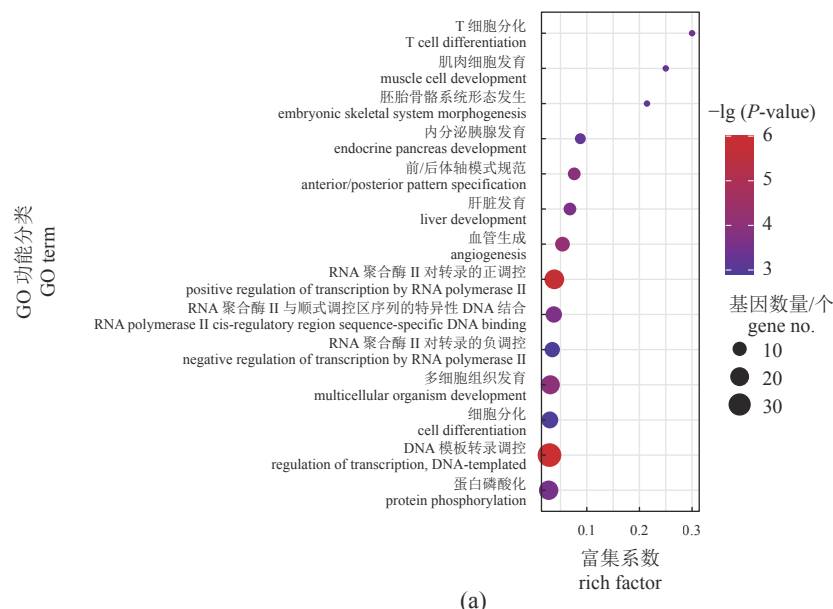
3.1 草鱼群体的 DNA 甲基化水平和差异甲基化模式

对 8 个草鱼群体进行的 DNA 甲基化测序和分析结果显示, 尼泊尔群体的 CpG、CHG 和 CHH 等 3 种类型甲基化水平最高, 安兴群体最低, 但总体上相差不大, 表明草鱼群体间的总体甲基化水平差异较小。而在上述 3 种类型的甲基化之间则有明显差异, CHG 和 CHH 类型的甲基化位点在草鱼群体全基因组范围内分布更加广泛, 但总体甲基化水平均较低, 表明其可能侧重于维持基因组的稳定^[38-39]。CpG 类型的甲基化位点尽管数量较少, 但是具有更高的甲基化水平, 表明其可能主要在基因表达的表现调控方面发挥重要作用^[40-41]。

有研究通过分析草鱼群体结构和基因流动事件, 发现印度、尼泊尔和中国等地的草鱼养殖群体起源于中国长江和珠江的多个水系^[24]。因此本研究将来自长江水系的万州、沅江群体以及来自珠江水系的肇庆这 3 个野生群体合并, 作为野生背景群体与其余 5 个养殖群体进行后续差异甲基化分析。5 个草鱼养殖群体与野生背景群体的差异甲基化模式, 反映了前者相较于自然栖息地的

野生群体而言, 对各地不同养殖环境和驯化选择适应的差异^[42]。印度养殖群体与 3 个野生群体之间具有最多的差异甲基化区域, 并同时富集到最多的 GO 功能分类, 表明草鱼从中国引种至印度

并在当地养殖的过程中经历了较强的养殖环境和驯化选择, 在表观遗传层面发生的许多甲基化变异可能促进了对这一过程的适应。



(图 3 Fig. 3)

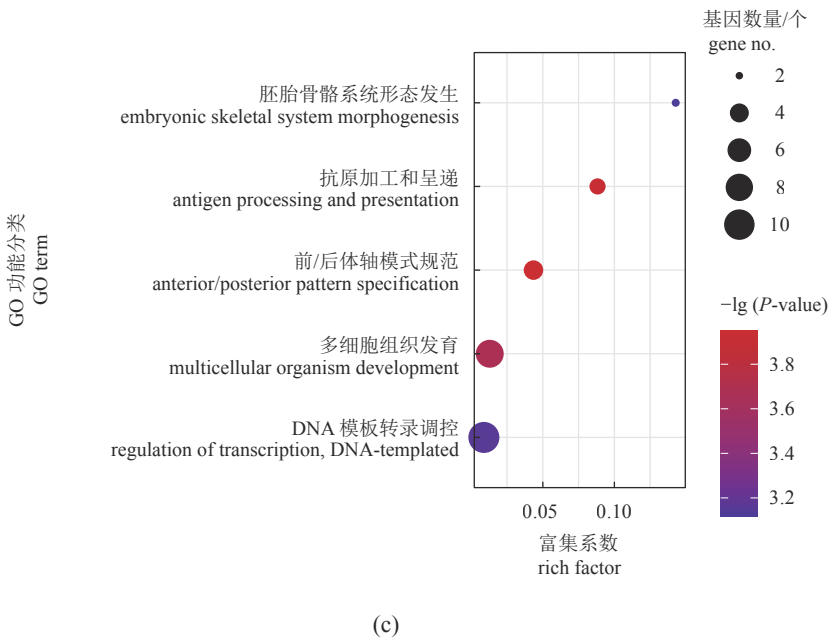


图 3 差异甲基化基因的生物途径类 GO 功能富集气泡图

(a) 安兴 vs. 野生; (b) 印度 vs. 野生; (c) 越南 vs. 野生。

Fig. 3 Dot plot of enriched biology process GO terms of DMGs

(a) Anxing vs. wild; (b) India vs. wild; (c) Vietnam vs. wild.

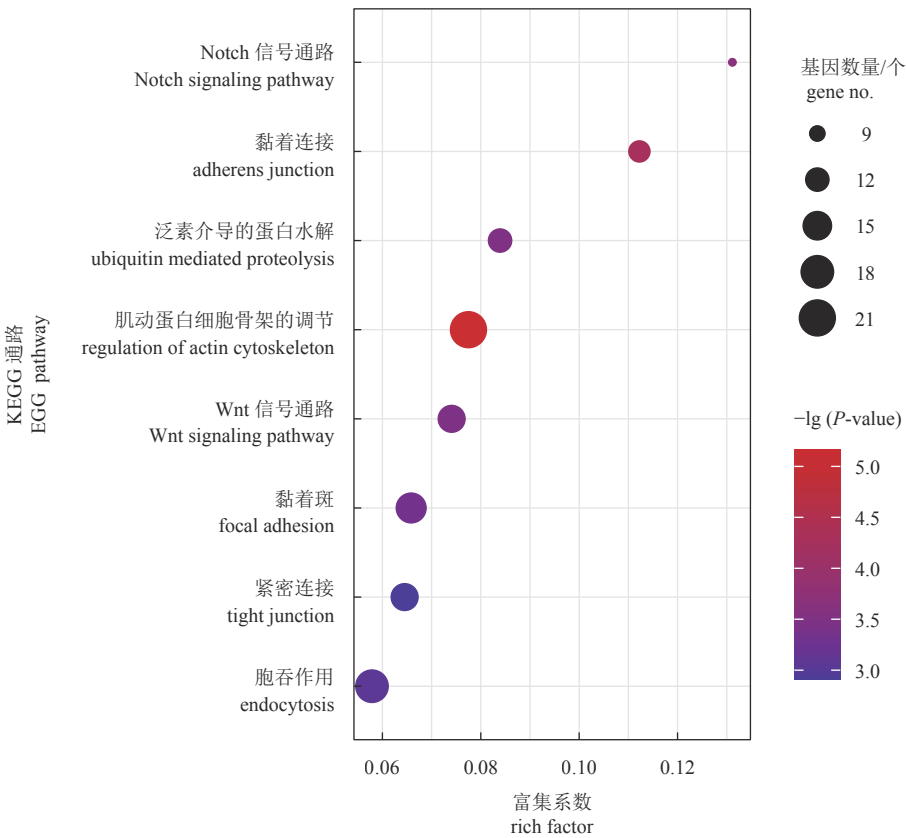


图 4 印度 vs. 野生群体的差异甲基化基因的 KEGG 通路富集气泡图

Fig. 4 Dot plot of enriched KEGG pathways of DMGs in India vs. wild

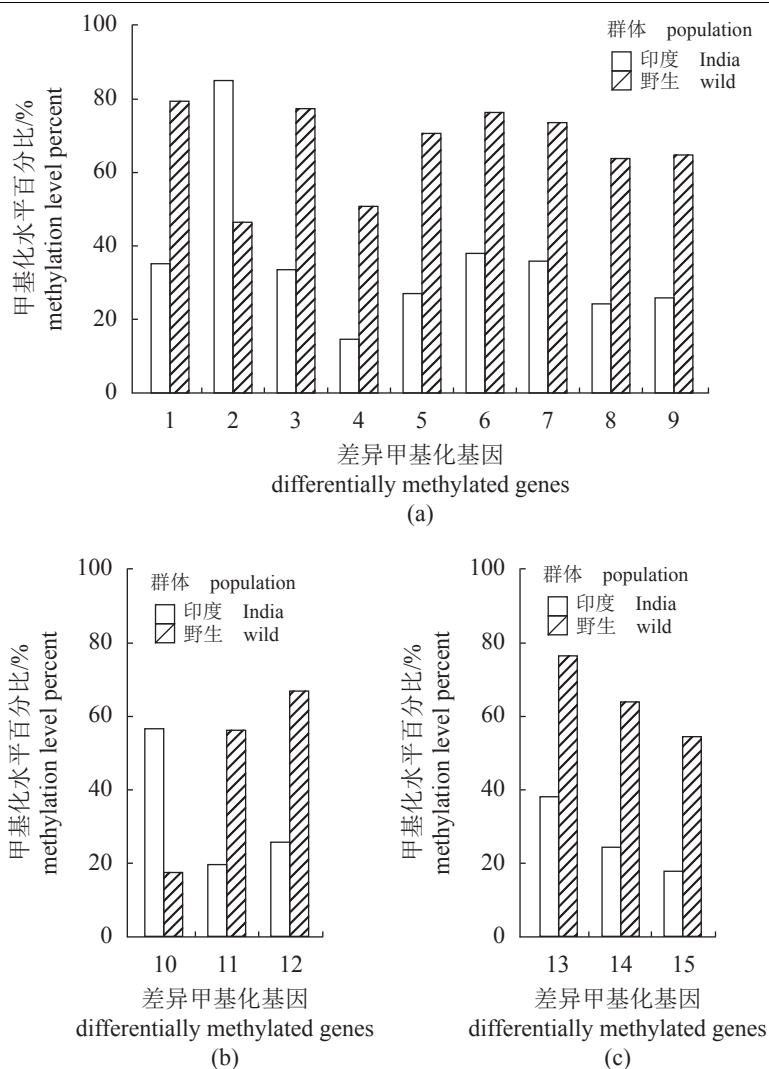


图 5 印度 vs. 野生群体中神经嵴细胞相关富集基因的甲基化水平柱状图

(a) 神经嵴细胞迁移相关基因 (GO:0001755); (b) 神经嵴细胞命运规范相关基因 (GO:0014036); (c) 神经嵴细胞分化相关基因 (GO:0014033); 1. *atp6v0a1a*, 2. *chd7*, 3. *col18a1a*, 4. *masp1*, 5. *nrg2a*, 6. *pard3ab*, 7. *plod3*, 8. *sema3d*, 9. *sema5bb*, 10. *tfap2a*, 11. *tfap2c*, 12. *kctd15b*, 13. *pard3ab*, 14. *sema3d*, 15. *alcama*.

Fig. 5 Bar plot of methylation level of enriched neural crest cell migration genes in India vs. wild

(a) the genes enriched with neural crest cell migration (GO:0001755); (b) the genes enriched with neural crest cell fate specification (GO:0014036); (c) the genes enriched with neural crest cell differentiation (GO:0014033); 1. *atp6v0a1a*, 2. *chd7*, 3. *col18a1a*, 4. *masp1*, 5. *nrg2a*, 6. *pard3ab*, 7. *plod3*, 8. *sema3d*, 9. *sema5bb*, 10. *tfap2a*, 11. *tfap2c*, 12. *kctd15b*, 13. *pard3ab*, 14. *sema3d*, 15. *alcama*.

3.2 草鱼养殖中发育相关基因甲基化的影响

富集分析的结果表明, 印度、越南和中国安兴这 3 个养殖群体与野生参考群体之间的 DMG 主要与系统发育相关, 涉及神经、免疫、骨骼和肌肉等多种组织的发育过程。这一结果与多项针对其他养殖鱼类的研究结果类似^[19, 23, 42-43], 表明不同鱼类在相似的养殖过程中对选择适应可能具有相似的表观遗传基础; 同时表明发育相关基因可能通过参与表型改变, 进而广泛参与养殖鱼类对水产养殖过程中人工驯化与环境选择的适应过程。

神经嵴细胞 (neural crest cell) 是胚胎发育阶段暂时存在的一类细胞群体。在胚胎发育过程中, 位于神经管背壁的神神经嵴细胞被诱导开始迁移, 沿着胚层间缝隙以细胞流的形式迁移到全身各处, 并进一步分化成如软骨、骨骼、肌肉、色素、结缔组织和神经等多种类型的细胞^[44]。在这一过程中神经嵴细胞的正常迁移尤为重要, 其不正常迁移会导致发育缺陷, 进而引起表型变化^[45]。维持神经嵴细胞的正常迁移涉及一系列信号通路以及细胞间黏附的动态调节^[44, 46]。已有多项研究证明, 神经嵴细胞在鱼类对环境适应过程中具有重要作

用^[23, 47-48]。值得注意的是, 本研究同样在印度养殖群体与野生参考群体之间的 DMG 中显著富集到 3 个与神经嵴细胞相关的 GO 功能分类。其中 *atp6v0a1a*、*chd7*、*col18a1a*、*masp1*、*nrg2a*、*pard3ab*、*plod3*、*sema3d* 和 *sema5bb* 等 9 个基因参与神经嵴细胞的迁移过程: *atp6v0a1a*、*col18a1a* 和 *plod3* 基因参与神经嵴细胞与周围细胞间黏附的调节^[49-51]; *masp1* 基因引导神经嵴细胞的迁移方向^[52]; *pard3ab*、*sema3d* 和 *sema5bb* 基因抑制神经嵴细胞的不当迁移^[44, 53-54]; *chd7* 与 *nrg2a* 基因促进神经嵴细胞迁移与分化过程^[55-57]。而 *tfap2a*、*tfap2c* 和 *kctd15b* 等 3 个基因参与神经嵴细胞的命运规范过程: *tfap2a* 和 *tfap2c* 基因作为转录因子 AP-2 家族成员, 诱导神经嵴细胞的命运规范^[58]; *kctd15b* 基因通过抑制 AP-2 转录因子活性, 参与对神经嵴细胞命运规范的调节过程^[59-60]。此外还有 *pard3ab*、*sema3d* 和 *alcama* 等 3 个基因参与神经嵴细胞的分化过程: *pard3ab* 和 *sema3d* 基因除了参与神经嵴细胞迁移过程外, 也分别促进神经和鳍组织的正常分化^[61-63]; *alcama* 基因通过参与 *edn1* 信号通路, 促进神经嵴细胞向咽部软骨组织的正常分化^[64]。因此, 这些基因的差异甲基化模式表明, 表观遗传修饰可能影响了这些基因的后续表达, 进而参与并促进了草鱼在印度当地水产养殖过程中对人工驯化与环境选择的适应。

4 结论

通过对亚洲范围内 8 个地区的养殖和野生草鱼群体进行了全基因组 DNA 甲基化测序, 共获得 308.15 Gb 测序数据, 发现 277 618 513 个甲基化位点。在养殖和野生群体之间共检测出 76 422 个 DML、总长度为 0.846 Mb 的 3 737 个 DMR, 以及 1 950 个 DMG。通过富集分析, 发现 DMG 主要与发育相关, 其中包括了一类与表型改变相关的重要发育基因。本研究初步探索了表观遗传修饰在鱼类对水产养殖过程中人工驯化与环境选择适应的影响, 为养殖鱼类进一步改良育种工作提供了新的思路, 同时也为保存和科学利用草鱼种质资源提供了有价值的遗传数据。

(作者声明本文无实际或潜在的利益冲突)

参考文献 (References):

- [1] Fabrice T. Fish domestication: an overview[M]//Teletchea F. Animal domestication. Intech Open, 2018.

<https://www.china-fishery.cn>

- [2] Milla S, Pasquet A, El Mohajer L, et al. How domestication alters fish phenotypes[J]. *Reviews in Aquaculture*, 2021, 13(1): 388-405.
- [3] Houston R D, Bean T P, Macqueen D J, et al. Harnessing genomics to fast-track genetic improvement in aquaculture[J]. *Nature Reviews Genetics*, 2020, 21(7): 389-409.
- [4] López M E, Neira R, Yáñez J M. Applications in the search for genomic selection signatures in fish[J]. *Frontiers in Genetics*, 2015, 5: 458.
- [5] Biber C, Kawam B, Chapelle V, et al. The role of stochasticity in the origin of epigenetic variation in animal populations[J]. *Integrative and Comparative Biology*, 2020, 60(6): 1544-1557.
- [6] Sarkies P. Molecular mechanisms of epigenetic inheritance: possible evolutionary implications[J]. *Seminars in Cell & Developmental Biology*, 2020, 97: 106-115.
- [7] Moore L D, Le T, Fan G P. DNA methylation and its basic function[J]. *Neuropsychopharmacology*, 2013, 38(1): 23-38.
- [8] Meng H, Cao Y, Qin J Z, et al. DNA methylation, its mediators and genome integrity[J]. *International Journal of Biological Sciences*, 2015, 11(5): 604-617.
- [9] Illingworth R S, Bird A P. CpG Islands-'a rough guide'[J]. *FEBS Letters*, 2009, 583(11): 1713-1720.
- [10] Chamorro-Garcia R, Diaz-Castillo C, Shoucri B M, et al. Ancestral perinatal obesogen exposure results in a transgenerational thrifty phenotype in mice[J]. *Nature Communications*, 2017, 8(1): 2012.
- [11] Knecht A L, Truong L, Marvel S W, et al. Transgenerational inheritance of neurobehavioral and physiological deficits from developmental exposure to benzo[a]pyrene in zebrafish[J]. *Toxicology and Applied Pharmacology*, 2017, 329: 148-157.
- [12] Leroux S, Gourichon D, Leterrier C, et al. Embryonic environment and transgenerational effects in quail[J]. *Genetics Selection Evolution*, 2017, 49(1): 14.
- [13] Skjærven K H, Jakt L M, Fernandes J M O, et al. Parental micronutrient deficiency distorts liver DNA methylation and expression of lipid genes associated with a fatty-liver-like phenotype in offspring[J]. *Scientific Reports*, 2018, 8(1): 3055.
- [14] Norouzitallab P, Baruah K, Vanrompay D, et al. Can epigenetics translate environmental cues into phenotypes?[J]. *Science of the Total Environment*, 2019, 647: 1281-1293.
- [15] Skinner M K, Guerrero-Bosagna C, Haque M M. Environmental factors and epigenetics in the development of the reproductive system[J]. *Environmental Health Perspectives*, 2012, 120(12): 1603-1610.

中国水产学会主办 sponsored by China Society of Fisheries

- onmentally induced epigenetic transgenerational inheritance of sperm epimutations promote genetic mutations[J]. *Epigenetics*, 2015, 10(8): 762-771.
- [16] Kamstra J H, Hurem S, Martin L M, *et al.* Ionizing radiation induces transgenerational effects of DNA methylation in zebrafish[J]. *Scientific Reports*, 2018, 8(1): 15373.
- [17] Jiang L, Zhang J, Wang J J, *et al.* Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos[J]. *Cell*, 2013, 153(4): 773-784.
- [18] Potok M E, Nix D A, Parnell T J, *et al.* Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern[J]. *Cell*, 2013, 153(4): 759-772.
- [19] Wellband K, Roth D, Linnansaari T, *et al.* Environment-driven reprogramming of gamete DNA methylation occurs during maturation and is transmitted intergenerationally in Atlantic salmon[J]. *G3 Genes| Genomes| Genetics*, 2021, 11(12): jkab353.
- [20] Vogt G. Facilitation of environmental adaptation and evolution by epigenetic phenotype variation: insights from clonal, invasive, polyploid, and domesticated animals[J]. *Environmental Epigenetics*, 2017, 3(1): dvx002.
- [21] Herráez M P, Lombó M, González-Rojo S. The role of epigenetics in fish biology and reproduction: an insight into the methods applied to aquaculture[M]//Monzón I F, Fernandes J M O. Cellular and molecular approaches in fish biology. Academic Press, 2022: 69-104.
- [22] Podgorniak T, Dhanasiri A, Chen X Q, *et al.* Early fish domestication affects methylation of key genes involved in the rapid onset of the farmed phenotype[J]. *Epigenetics*, 2022, 17(10): 1281-1298.
- [23] Anastasiadi D, Piferrer F. Epimutations in developmental genes underlie the onset of domestication in farmed European sea bass[J]. *Molecular Biology and Evolution*, 2019, 36(10): 2252-2264.
- [24] Shen Y B, Wang L, Fu J J, *et al.* Population structure, demographic history and local adaptation of the grass carp[J]. *BMC Genomics*, 2019, 20(1): 467.
- [25] Cui Y T, Liu B, Xie J, *et al.* Effect of heat stress and recovery on viability, oxidative damage, and heat shock protein expression in hepatic cells of grass carp (*Ctenopharyngodon idellus*)[J]. *Fish Physiology and Biochemistry*, 2014, 40(3): 721-729.
- [26] Cudmore B, Mandrak N E. Biological synopsis of grass carp (*Ctenopharyngodon idella*)[R]. Burlington: Fisheries and Oceans Canada, Great Lakes Laboratory for Fisheries and Aquatic Sciences, 2004.
- [27] Shi M J, Zhang Q X, Li Y M, *et al.* Global gene expression profile under low-temperature conditions in the brain of the grass carp (*Ctenopharyngodon idellus*)[J]. *PLoS One*, 2020, 15(9): e0239730.
- [28] Chen Q, Wang C H, Lu G Q, *et al.* Microsatellite genetic diversity and differentiation of native and introduced grass carp populations in three continents[J]. *Genetica*, 2012, 140(4): 115-123.
- [29] 宋晓, 李思发, 王成辉, 等. 草鱼中国土著群体与欧美日移居群体遗传差异的线粒体序列分析[J]. *水生生物学报*, 2009, 33(4): 709-716.
- Song X, Li S F, Wang C H, *et al.* Grass carp (*Ctenopharyngodon idellus*) genetic structure analysis among native populations in China and introduced populations in USA, Europe and Japan based on mitochondrial sequence[J]. *Acta Hydrobiologica Sinica*, 2009, 33(4): 709-716 (in Chinese).
- [30] Andrews S. FastQC A quality control tool for high throughput sequence data[Z]. 2014.
- [31] Bolger A M, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data[J]. *Bioinformatics*, 2014, 30(15): 2114-2120.
- [32] Krueger F, Andrews S R. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications[J]. *Bioinformatics*, 2011, 27(11): 1571-1572.
- [33] Wang Y P, Lu Y, Zhang Y, *et al.* The draft genome of the grass carp (*Ctenopharyngodon idellus*) provides insights into its evolution and vegetarian adaptation[J]. *Nature Genetics*, 2015, 47(6): 625-631.
- [34] Langmead B, Salzberg S L. Fast gapped-read alignment with bowtie 2[J]. *Nature Methods*, 2012, 9(4): 357-359.
- [35] Park Y, Wu H. Differential methylation analysis for BS-seq data under general experimental design[J]. *Bioinformatics*, 2016, 32(10): 1446-1453.
- [36] Krzywinski M, Schein J, Birol I, *et al.* Circos: an information aesthetic for comparative genomics[J]. *Genome Research*, 2009, 19(9): 1639-1645.
- [37] Wickham H. ggplot2: elegant graphics for data analysis[M]. New York: Springer-Verlag, 2016.
- [38] 罗荣松. 奶绵羊与蒙古羊全基因组选择信号和 DNA 甲基化差异研究 [D]. 呼和浩特: 内蒙古大学, 2020.
- Luo R S. Differences in genomic selection signals and DNA methylation between dairy sheep and Mongolian sheep[D]. Hohhot: Inner Mongolia University, 2020 (in Chinese).
- [39] Nardo L, Lamperti M, Salerno D, *et al.* Effects of non-

- CpG site methylation on DNA thermal stability: a fluorescence study[J]. *Nucleic Acids Research*, 2015, 43(22): 10722-10733.
- [40] Suzuki M M, Bird A. DNA methylation landscapes: provocative insights from epigenomics[J]. *Nature Reviews Genetics*, 2008, 9(6): 465-476.
- [41] Jones P A. Functions of DNA methylation: islands, start sites, gene bodies and beyond[J]. *Nature Reviews Genetics*, 2012, 13(7): 484-492.
- [42] Venney C J, Sutherland B J G, Beacham T D, *et al.* Population differences in Chinook salmon (*Oncorhynchus tshawytscha*) DNA methylation: genetic drift and environmental factors[J]. *Ecology and Evolution*, 2021, 11(11): 6846-6861.
- [43] Nilsson E, Sadler-Riggelman I, Beck D, *et al.* Differential DNA methylation in somatic and sperm cells of hatchery vs. wild (natural-origin) steelhead trout populations[J]. *Environmental Epigenetics*, 2021, 7(1): dvab002.
- [44] Szabó A, Mayor R. Mechanisms of neural crest migration[J]. *Annual Review of Genetics*, 2018, 52: 43-63.
- [45] Wilkins A S, Wrangham R W, Fitch W T. The “domestication syndrome” in mammals: a unified explanation based on neural crest cell behavior and genetics[J]. *Genetics*, 2014, 197(3): 795-808.
- [46] Scarpa E, Szabó A, Bibonne A, *et al.* Cadherin switch during EMT in neural crest cells leads to contact inhibition of locomotion via repolarization of forces[J]. *Developmental Cell*, 2015, 34(4): 421-434.
- [47] Yoshizawa M, Hixon E, Jeffery W R. Neural crest transplantation reveals key roles in the evolution of cavefish development[J]. *Integrative and Comparative Biology*, 2018, 58(3): 411-420.
- [48] Luo M K, Lu G Q, Yin H R, *et al.* Fish pigmentation and coloration: molecular mechanisms and aquaculture perspectives[J]. *Reviews in Aquaculture*, 2021, 13(4): 2395-2412.
- [49] Tuttle A M, Hoffman T L, Schilling T F. Rabconnectin-3a regulates vesicle endocytosis and canonical Wnt signaling in zebrafish neural crest migration[J]. *PLoS Biology*, 2014, 12(5): e1001852.
- [50] Banerjee S, Isaacman-Beck J, Schneider V A, *et al.* A novel role for Lh3 dependent ECM modifications during neural crest cell migration in zebrafish[J]. *PLoS One*, 2013, 8(1): e54609.
- [51] Schneider V A, Granato M. The myotomal *diwanka* (*lh3*) glycosyltransferase and type X VIII collagen are critical for motor growth cone migration[J]. *Neuron*, 2006, 50(5): 683-695.
- [52] Leslie J D, Mayor R. Complement in animal development: unexpected roles of a highly conserved pathway[J]. *Seminars in Immunology*, 2013, 25(1): 39-46.
- [53] Moore R, Theveneau E, Pozzi S, *et al.* Par3 controls neural crest migration by promoting microtubule catastrophe during contact inhibition of locomotion[J]. *Development*, 2013, 140(23): 4763-4775.
- [54] Yu H H, Moens C B. Semaphorin signaling guides cranial neural crest cell migration in zebrafish[J]. *Developmental Biology*, 2005, 280(2): 373-385.
- [55] Patten S A, Jacobs-McDaniels N L, Zaouter C, *et al.* Role of Chd7 in zebrafish: a model for CHARGE syndrome[J]. *PLoS One*, 2012, 7(2): e31650.
- [56] Liu Z Z, Wang Z L, Choi T I, *et al.* Chd7 is critical for early T-cell development and thymus organogenesis in zebrafish[J]. *The American Journal of Pathology*, 2018, 188(4): 1043-1058.
- [57] Zhu S Q, Liu W H, Ding H F, *et al.* BMP4 and Neuregulin regulate the direction of mouse neural crest cell differentiation[J]. *Experimental and Therapeutic Medicine*, 2019, 17(5): 3883-3890.
- [58] Li W, Cornell R A. Redundant activities of Tfap2a and Tfap2c are required for neural crest induction and development of other non-neural ectoderm derivatives in zebrafish embryos[J]. *Developmental Biology*, 2007, 304(1): 338-354.
- [59] Chambers B E, Clark E G, Gatz A E, *et al.* Kctd15 regulates nephron segment development by repressing Tfap2a activity[J]. *Development*, 2020, 147(23): dev191973.
- [60] Heffer A, Marquart G D, Aquilina-Beck A, *et al.* Generation and characterization of Kctd15 mutations in zebrafish[J]. *PLoS One*, 2017, 12(12): e0189162.
- [61] Govindan J, Tun K M, Iovine M K. Cx43-dependent skeletal phenotypes are mediated by interactions between the hapln1a-ECM and Sema3d during fin regeneration[J]. *PLoS One*, 2016, 11(2): e0148202.
- [62] Hong E, Jayachandran P, Brewster R. The polarity protein Pard3 is required for centrosome positioning during neurulation[J]. *Developmental Biology*, 2010, 341(2): 335-345.
- [63] Tep C, Kim M L, Opincariu L I, *et al.* Brain-derived neurotrophic factor (BDNF) induces polarized signaling of small GTPase (Rac1) protein at the onset of Schwann

cell myelination through partitioning-defective 3 (Par3) protein[J]. *Journal of Biological Chemistry*, 2012, 287(2): 1600-1608.

[64] Choudhry P, Joshi D, Funke B, *et al.* Alcama mediates Edn1 signaling during zebrafish cartilage morphogenesis[J]. *Developmental Biology*, 2011, 349(2): 483-493.

Differential methylation analysis of *Ctenophryngodon idella* populations in Asia

WEI Shang^{1,2}, XIE Lingli^{1,3,4}, ZHU Hua⁵, ZHANG Qingjing⁵, SHEN Yubang^{1,3,4},
XU Xiaoyan^{1,3,4*}, LI Jiale^{1,3,4*}

(1. Key Laboratory of Freshwater Aquatic Genetic Resources, Ministry of Agriculture and Rural Affairs, Shanghai Ocean University, Shanghai 201306, China;

2. Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources, Ministry of Education, Shanghai Ocean University, Shanghai 201306, China;

3. Shanghai Engineering Research Center of Aquaculture, Shanghai Ocean University, Shanghai 201306, China;

4. National Demonstration Center for Experimental Fisheries Science Education, Shanghai Ocean University, Shanghai 201306, China;

5. Beijing Key Laboratory of Fishery Biotechnology, Fisheries Research Institute, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100068, China)

Abstract: *Ctenophryngodon idella* is one of the most important aquaculture species in the world, and suffered from local adaptation during the process of introduction and aquaculture. However, most studies of *C. idella* local adaptation were based on genetic variation, and the effect of epimutation was still widely unknown. In order to explore how DNA methylation affects the domestication and environmental adaptation of the *C. idella* populations in Asia. A whole-genome bisulfite sequencing was performed between 5 farmed populations and the wild background population, followed by further differential methylation analysis and enrichment analysis. A total of 308.15 Gb sequencing data were generated with a sequencing depth of 31× and a mapping rate of 62.97 % in average, where 76 422 differentially methylated loci (DML), 3 737 differentially methylated regions (DMR), and 1 950 differentially methylated genes (DMG) were identified. Gene Ontology enrichment analysis showed that the DMGs were significantly enriched in development process, such as angiogenesis, neural crest cell migration, and cranial skeletal system development. Annotation with the KEGG pathway database exhibited these genes were mainly involved in adhesion junction, Notch signaling pathway, and Wnt signaling pathway. Among the DMG, some might be functional in the developmental process of neural, immunity, bone, and muscle tissues, like *sema3d*, *sema5bb* and *nrg2a*. This study enhances the understanding of the epigenetic mechanism of domestication and environmental adaption during the process of aquaculture and provides valuable data for the conservation and utilization of the *C. idella* germplasm resources.

Key words: *Ctenophryngodon idella*; genome; whole-genome bisulfite sequencing; differential methylation

Corresponding authors: XU Xiaoyan. E-mail: xyxu@shou.edu.cn;

LI Jiale. E-mail: jlli2009@126.com

Funding projects: National Key R & D Program of China (2020YFD0900103); China Agriculture Research System (CARS-45-03); Shanghai Engineering and Technology Center for Promoting Ability (19DZ2284300)