

DOI: 10.11964/jfc.20230313950

基于全基因组重测序技术探究亚洲草鱼遗传多样性及其适应机制



谢玲莉^{1,2,3}, 沈玉帮^{1,2,3}, 桂朗^{1,2,3}, 徐晓雁^{1,2,3*}, 李家乐^{1,2,3*}

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, 水产养殖工程研究中心, 上海 201306; 3. 上海海洋大学, 水产科学国家级实验教学示范中心, 上海 201306

摘要:

【目的】探讨亚洲地区草鱼种群的遗传多样性和适应性进化, 以期全面理解草鱼的遗传资源和免疫适应的分子进化机制。

【方法】对来自中国4个省、直辖市的5个种群(沅江种群、安乡种群、万州种群、天津种群和肇庆种群)及与中国相邻的3个亚洲国家(尼泊尔、越南、印度)种群共计8个草鱼种群进行全基因组重测序。采用群体遗传学方法进行分析, 包括群体结构、主成分分析、系统进化树、遗传分化指数和选择清除分析。

【结果】实验共检测到6548523个SNP位点, 这些位点主要分布在基因间区, 并在细胞黏附分子、造血干细胞、DNA重组修复等信号通路中富集。万州种群的遗传多样性最高, 尼泊尔种群的遗传多样性最低。聚类分析发现, 越南与中国5个种群间表现出较高的遗传相似性。选择清除分析鉴定到的受选择基因主要富集在免疫反应调节、心肌细胞调节、钠离子通道等信号通路。结合SNP注释信息, 在印度和尼泊尔种群中发现受选择基因 *gimap8*, 该基因在T细胞和B细胞发育过程中发挥重要作用。

【结论】中国安乡、沅江、万州草鱼种群遗传多样性丰富, 尼泊尔和印度与其他种群草鱼在遗传分化上有明显差异, 初步筛选出了 *gimap8* 等潜在适应性基因。本研究为进一步探讨草鱼在不同生境下的分子进化机制奠定了基础。

关键词: 草鱼; 全基因组重测序; 遗传多样性; 选择清除

草鱼 (*Ctenopharyngodon idella*) 在过去十年成为全球产量最高的鱼类, 2020年占全球内陆水产养殖总产量的11.8%^[1]。迄今为止, 草鱼已被引入中亚、美洲、欧洲和阿拉伯半岛等地^[2], 由于地理隔离、人工驯养和环境差异等因素, 不同地理种群草鱼的遗传多样性已出现显著差异^[3]。汪焕等^[4]利用微卫星分子标记对来自安徽省的4个草鱼养殖种群进行遗传分析, 结果发现池州和无为种群的遗传分化最大, 而怀远和滁州种群的亲缘关系最近。瞿东东等^[5]利用线粒体 *Cyt b* 基因分析了长江上游5个草鱼种群的遗传多样性, 发现长江草鱼种群的遗传多

第一作者: 谢玲莉, 从事水产动物基因组学研究, E-mail: xlqff@163.com



通信作者: 徐晓雁, 从事水产动物种质资源与遗传育种研究, E-mail: xyxu@shou.edu.cn;

李家乐, 从事水产动物种质资源与遗传育种研究, E-mail: jlli2009@126.com



资助项目: 现代农业产业技术体系(CARS-45-03)

收稿日期: 2023-03-24

修回日期: 2023-07-16

文章编号:

1000-0615(2025)01-019103-15

中图分类号: S 917.4

文献标志码: A

作者声明本文无利益冲突

©《水产学报》编辑部(CC BY-NC-ND 4.0)
Copyright © Editorial Office of Journal of Fisheries of China (CC BY-NC-ND 4.0)



样性较低, 且历史上发生过瓶颈效应和种群扩张。随着第二代测序技术的发展, 现在能够最大程度地挖掘单核苷酸多态性 (single-nucleotide polymorphism, SNP) 背后的高多态性信息。Wang 等^[6] 利用第二代测序技术开发了雌性牙鲮 (*Paralichthys olivaceus*) SNP 标记, 并基于选择性扫描鉴定出 580 个具有假定选择特征的基因, 这些基因富集到免疫和繁殖相关的 39 个 KEGG 通路中。因此, 利用 SNP 分析不同种群的遗传多样性以及在全基因组层面上揭示不同地理种群的差异, 已成为一种重要的研究策略^[7]。

种群的遗传变异信息和基因型变化往往受到地理位置和环境压力的影响^[8], 环境压力会通过物种的适应性在基因组上留下痕迹, 即选择信号。利用遗传分化指数 (fixation index, F_{ST})、核苷酸多样性和跨群体复合似然比检验 (cross-population composite likelihood ratio test, XP-CLR)^[9] 挖掘基因组上的选择信号, 可以深入地理解不同种群的种质特性等问题^[10]。例如, Clucas 等^[11] 利用 F_{ST} 发现热休克蛋白等功能基因在冬季和春季产卵的大西洋鳕 (*Gadus morhua*) 种群之间产生了遗传分化。Clucas 等^[11] 和 Narum 等^[12] 在对具有不同生活史特征的鲑种群进行全基因组重测序研究中发现, 28 号染色体的变异与性早熟性状相关, 这发生在不同性成熟表型与早熟洄游的种群中。此外, 他们发现两个种群的基因组中存在广泛的选择性变异, 表明生活史变化与基因适应紧密关联。全基因组重测序 (whole-genome resequencing, WGR) 技术已被广泛应用于研究种群的遗传多样性和遗传结构^[13-14]。例如, Wang 等^[15] 利用 WGR 技术发现大黄鱼 (*Larimichthys crocea*) 2 个养殖种

群遗传多样性较低, 而野生种群遗传多样性较高。Sun 等^[16] 使用 WGR 技术对 5 个大口黑鲈 (*Micropterus salmoides*) 种群的遗传多样性和种群结构进行分析, 发现改良后的种群遗传结构变动较大, 且遗传多样性急剧下降。Du 等^[17] 利用 WGR 技术分析美国佛罗里达和中国北方大口黑鲈的遗传多样性, 发现佛罗里达大口黑鲈的遗传多样性高于北方大口黑鲈。上述研究都充分展示了 WGR 技术在遗传多样性和种群结构研究中的重要应用价值。

为了深入理解草鱼种群遗传结构和环境适应性的分子机制, 需要充分利用全基因组数据来挖掘草鱼的遗传资源和变异信息。本研究中从亚洲 8 个地点收集草鱼样本, 并进行了全基因组重测序, 旨在揭示不同地理种群的草鱼在基因组水平上的变异。此外, 还利用选择清除分析来鉴定草鱼的受选择基因, 以探索其潜在的分子适应机制。本研究可为进一步解析草鱼的遗传资源多样性、开发重要性状的分子标记以及探索草鱼在不同环境下的分子适应机制提供有价值的参考。

1 材料与方法

1.1 草鱼 DNA 提取

草鱼种群样本来自中国 4 个省、直辖市 (湖南、重庆、天津和广东) 的 5 个种群 (沅江种群、安乡种群、万州种群、天津种群和肇庆种群) 及与中国相邻 3 个亚洲国家 (尼泊尔、越南、印度) 的 3 个种群 (表 1)^[18]。采用海洋动物组织基因组 DNA 提取试剂盒 [DP324, 天根生化科技 (北京) 有限公司] 提取保存在 75% 乙醇中的

表 1 草鱼样品信息

Tab. 1 Sample information of *C. idella*

种群 group	纬度 latitude	经度 longitude	城市 city	区域 Asia region	所处流域/水系 river system
尼泊尔 NP	27.42	85.03	尼泊尔黑道达市	南亚	根得格河
印度 IN	26.76	83.37	印度戈勒克布尔市	南亚	恒河水系
越南 VN	21.12	105.98	越南北宁市	东南亚	湄公河流域
沅江 YJ	28.85	112.36	湖南省沅江市	东亚	长江流域
安乡 AX	29.40	112.13	湖南省常德市安乡县	东亚	长江流域
万州 WZ	30.83	108.45	重庆市万州区	东亚	长江流域
天津 TJ	39.38	117.05	天津市武清区	东亚	海河流域
肇庆 ZQ	23.08	112.53	广东省肇庆市	东亚	珠江流域

草鱼鳍条 DNA。采用 1.2% 琼脂糖凝胶电泳检测所提取 DNA 的质量, 利用 Nanodrap D2000 微量分光光度计测定 DNA 浓度及纯度。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(20171024), 实验过程严格遵守上海海洋大学伦理规范, 并按照上海海洋大学伦理委员会制定的规章制度执行。

1.2 样品测序、序列比对

样品送至百迈克生物科技有限公司(北京)测序。通过超声波破碎技术将合格的 DNA 样品随机打断成长 300~500 bp 的片段。经过末端修复、加 ployA 尾、加测序接头、纯化、PCR 扩增等步骤, 完成文库制备。构建好的重测序文库利用 Illumina HiSeq 2500 平台进行测序。原始数据经过质控筛选[过滤掉接头序列、低质量序列(MQ \geq 6)以及短片段(长度<500 bp)], 得到质控后数据(clean reads)。使用 BWA 软件^[19]将 clean reads 比对到本实验室组装的草鱼基因组, 采用 Samtools^[20]统计比对率和覆盖度。

1.3 变异检测及结构注释

使用 GATK^[21]进行 SNP 和插入缺失(insertions and deletions, INDEL)变异检测, 过滤参数如下, SNP: QD < 2.0, MQ < 40.0, FS > 60.0, MQRankSum < -12.5, ReadPosRankSum < -8.0; InDel: QD < 2.0, FS > 200.0 II ReadPosRankSum < -20.0。使用 lumpy^[22]进行结构变异检测, 使用 CNVnator^[23]进行拷贝数变异检测。利用 ANNOVAR 软件^[24]和 snpEff 软件^[25]对 SNP、INDEL、SV 和 CNV 进行位置和功能注释。

1.4 群体进化分析

使用群体聚类分析和主成分分析来探究 8 个草鱼种群间的群体结构和亲缘关系。利用 iqtree 软件^[26]构建进化树, 使用 iTol^[27]作图。使用 Plink 软件^[28]完成主成分分析(principal component analysis, PCA), 利用 R 语言包 ggplot2 包制图。采用 VCFtools 软件^[29]计算各种群核苷酸多样性。

1.5 群体受选择区域鉴定

采用计算 F_{ST} ^[30]与 XP-CLR 法^[9], 通过 Linux 系统 awk 命令截取受选择区域, 用前面分析中产生的分组进行两两分组计算, 二者皆以阈值 1% 截取受选择区域。使用 VCFtools 软

件^[29]将基因组以 50 kb 的滑动步长进行 500 kb 不重叠的窗口分割, 进行 F_{ST} 计算; 使用 XP-CLR 软件^[9]以 20 kb 的滑动步长进行 100 kb 不重叠的窗口分割计算 xpclr 分值。

1.6 基因富集分析

通过 eggNOG-mapper 对草鱼基因进行功能注释^[31], 使用在线软件 AnnotationForge (<https://bioconductor.org/packages/release/bioc/html/AnnotationForge.html>) 构建草鱼 OrgDB 数据库, 使用 clusterProfiler^[32]软件包的 enrichGO 和 enricher 函数进行 GO 和 KEGG 富集分析。

1.7 序列分析

利用 MEGA11^[33]软件的 ClustalW 方法进行 8 个种群 *gimap8* 基因的序列比对, 利用 GeneDoc 软件^[34]进行可视化, 使用 NCBI 的 Batch CD-Search 工具预测基因的功能结构域^[35]。采用 AlphaFold2 软件进行蛋白质结构预测^[36-37], 使用 PyMol 软件^[38]进行可视化。

2 结果

2.1 全基因组重测序数据分析

8 个种群的重测序数据共获得 1 329 489 107 个 clean reads。Q30 平均值为 92.05%, 表明测序质量较高。基因组的平均覆盖率达到 90.72%, 平均测序深度为 55 \times (表 2)。上述结果表明, 测序质量可靠, 数据适用于后续分析。

2.2 变异检测及其结构注释

在 8 个种群中共鉴定出 6 548 523 个单核苷酸多态性(SNP)、2 267 193 个插入或缺失(INDEL)、330 397 个结构变异(SV)以及 34 800 个拷贝数变异(CNV)。SNP 变异在 24 对染色体上均匀分布(图 1)。在 CNV 变异中, 8 个种群的缺失(deletion)变异占比最高, 约为 85%。SV 变异的数量变化规律与 CNV 相似, 约占 78%(表 3)。

对 8 个种群 SNP 进行结构注释统计, 发现 SNP 主要分布在基因间区, 共检测到 3 591 582 个; 在外显子中的非同义替换达到 108 829 个; SNP 突变导致的终止子获得有 1 867 个, 而丢失终止子有 334 个; 在可变剪切位点 2 bp 以内的基因组区域内检测到 SNP 位点 1 011 个(图 2-a)。对于 INDEL 注释统计, 变异主要分布在基因间区, 占比为 51.2%(图 2-b)。

表 2 重测序数据质量比对

Tab. 2 Whole-genome resequencing of the *C. idella* populations

种群 groups	碱基/个 bases no.	GC/% GC percent	Q20/% Q20 percent	Q30/% Q30 percent	reads/条 reads no.	比对率/% mapping rate	测序深度/ \times depth
尼泊尔 NP	18 154 596 600	38.52	97.53	95.50	121 030 644	88.50	40.231 6
印度 IN	22 843 501 800	38.79	96.83	93.80	152 290 012	94.02	50.619 0
越南 VN	24 907 575 750	38.43	95.47	91.14	166 050 505	84.52	55.047 2
沅江 YJ	24 928 015 200	38.02	95.52	91.27	166 186 768	93.55	54.060 1
安乡 AX	30 056 091 900	38.22	92.79	86.65	200 373 946	90.13	66.117 8
万州 WZ	37 424 878 950	38.62	94.06	88.73	249 499 193	93.57	80.322 8
天津 TJ	21 151 797 150	38.61	97.38	94.88	141 011 981	93.03	46.743 0
肇庆 ZQ	19 956 908 700	38.57	97.17	94.45	133 046 058	88.43	44.164 3

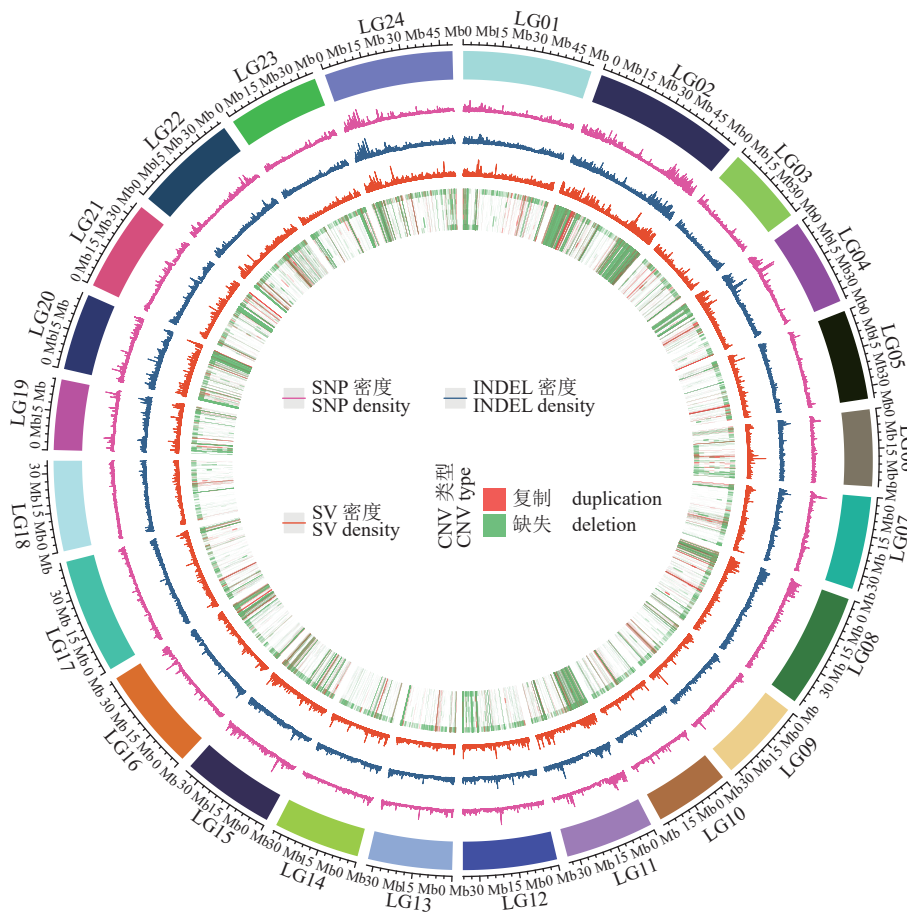


图 1 草鱼全基因组 SNP、INDEL、SV、CNV 分布图

从外到内各圈分别为草鱼染色体、SNP、INDEL、SV、CNV。

Fig. 1 SNP, INDEL, SV and CNV density of DNA level of the *C. idella* populations

The circles from outside to inside are *C. idella* chromosome, SNP, INDEL, SV, CNV variation.

2.3 遗传突变的 KEGG 富集分析

对涉及 4 种变异的非同义突变基因进行 KEGG 富集分析后发现, SNP 变异在细胞黏附分子、造血干细胞、DNA 重组修复等信号通路

中富集 (图 3-a); INDEL 变异在细胞黏附分子、造血干细胞、B 细胞受体信号通路中富集 (图 3-b); SV 变异在细胞黏附分子、造血干细胞、B 细胞受体等信号通路中富集 (图 3-c); CNV 变异在细胞黏附分子、纤毛信号与相关蛋白、造

表 3 序列变异信息

Tab. 3 The information of SV and CNV

种群 group	SV数量 no. of SV			CNV数量 no. of CNV	
	缺失 deletion	复制 duplication	颠换 inversion	缺失 deletion	复制 duplication
尼泊尔 NP	28 821	4 695	410	2 539	552
印度 IN	32 412	5 385	453	2 401	634
越南 VN	33 967	5 152	444	3 355	878
沅江 YJ	33 723	5 331	445	4 070	1 180
安乡 AX	42 453	6 984	677	4 419	1 009
万州 WZ	47 793	7 982	692	5 099	1 462
天津 TJ	30 718	4 784	385	2 869	811
肇庆 ZQ	31 277	4 989	425	2 614	908

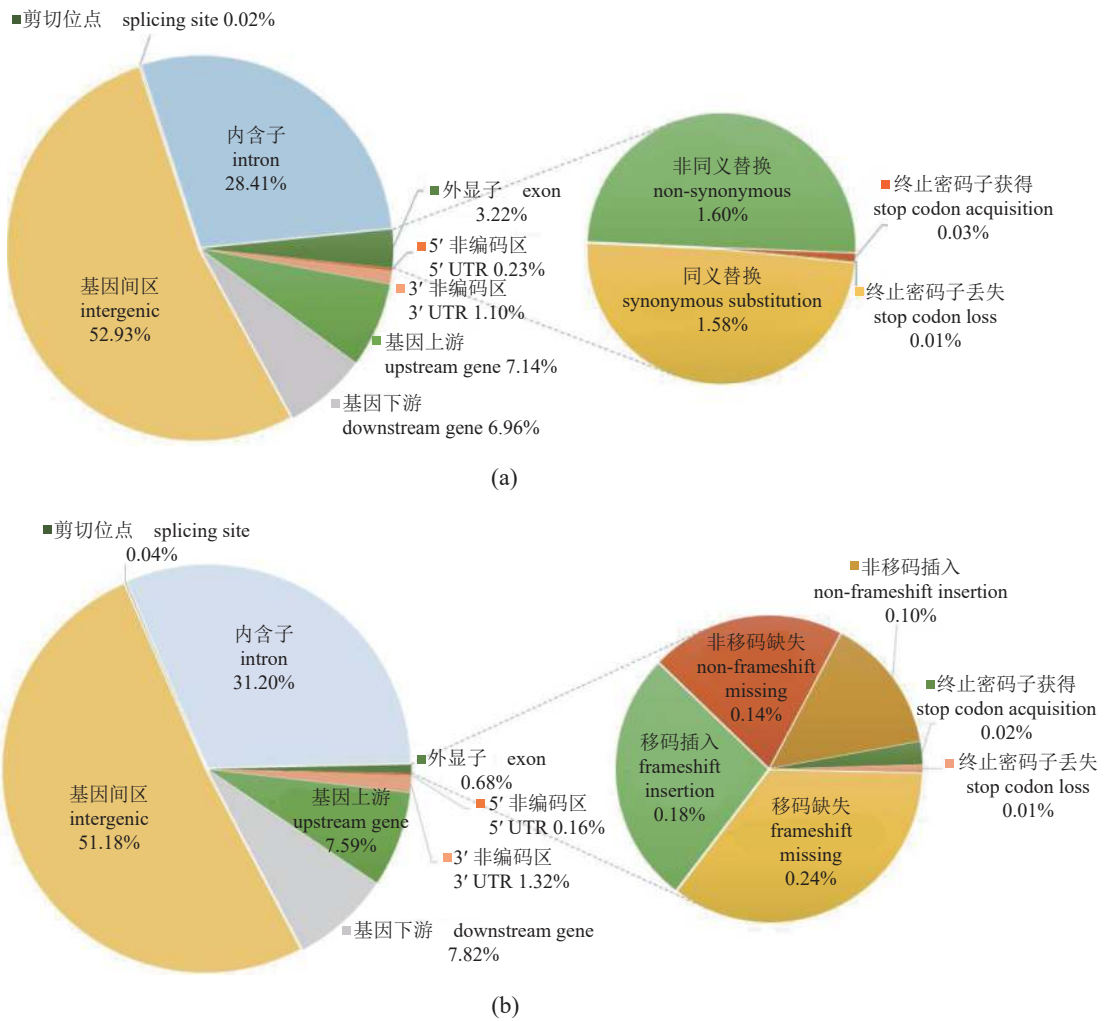


图 2 SNP、INDEL 注释结果

(a) SNP 注释结果, (b) INDEL 注释结果。

Fig. 2 The annotations of SNPs and INDELs

(a) the annotations of SNPs, (b) the annotations of INDELs.

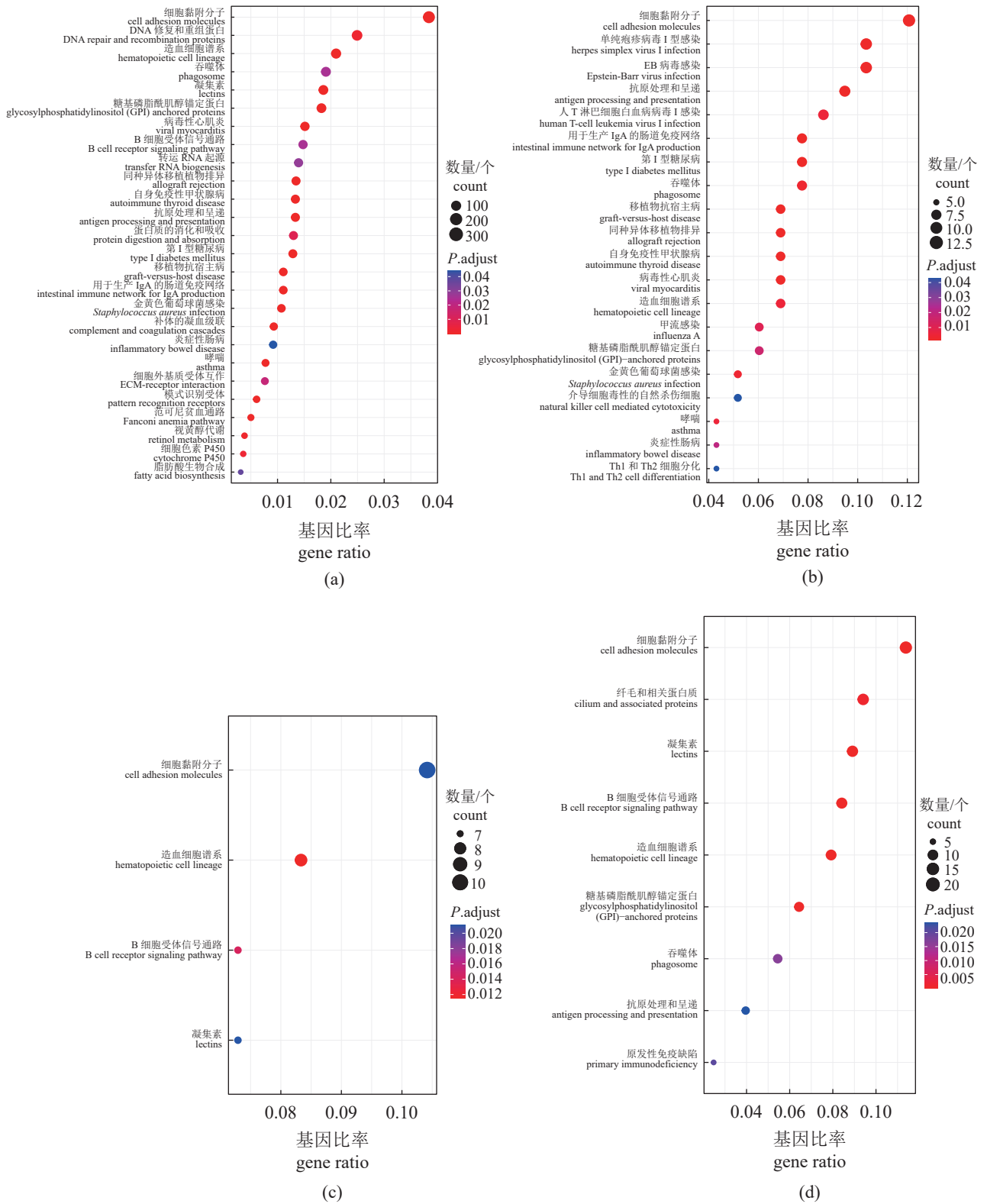


图3 四种变异区域基因的 KEGG 富集气泡图

(a) SNP 变异区域, (b) NDEL 变异区域, (c) V 变异区域, (d) CNV 变异区域。

Fig. 3 Dot plot of enriched KEGG pathway of genes in four variable regions

(a) SNP, (b) INDEL, (c) SV, (d) CNV.

血干细胞、血凝素、B 细胞受体等信号通路中富集(图 3-d)。

2.4 群体遗传结构分析

通过主成分分析研究 8 个样本间的遗传背景相似性和聚类关系, 发现 8 个草鱼种群被分为 3 个谱系。越南与中国的 5 个种群聚集在一起, 而印度和尼泊尔种群各自分散(图 4-a)。基于检测到的 SNP 信息对 8 个种群构建进化树, 结果与主成分分析一致, 同样聚类成 3 个分支(图 4-b)。计算草鱼 8 个种群的核苷酸多样性

(π) 和 F_{ST} 值, 探讨其遗传多样性及差异。结果显示, WZ 种群的 π 值最高, 为 4.89×10^{-3} , 而 NP 种群的 π 值最低, 为 3.64×10^{-3} (表 4)。此外, 长江流域种群相对于第三亚群(珠江亚群, ZJ)及南亚流域种群相对于 ZJ 的遗传分化指数 (F_{ST}) 较高, 分别达 0.033 0 和 0.026 6, 而第三亚群(ZJ)相对于长江流域(DY) F_{ST} 仅 0.000 6。

2.5 受选择基因挖掘

适应过程中的选择信号 为了探索 8 个草鱼种群在适应不同环境时的受选择信号, 实

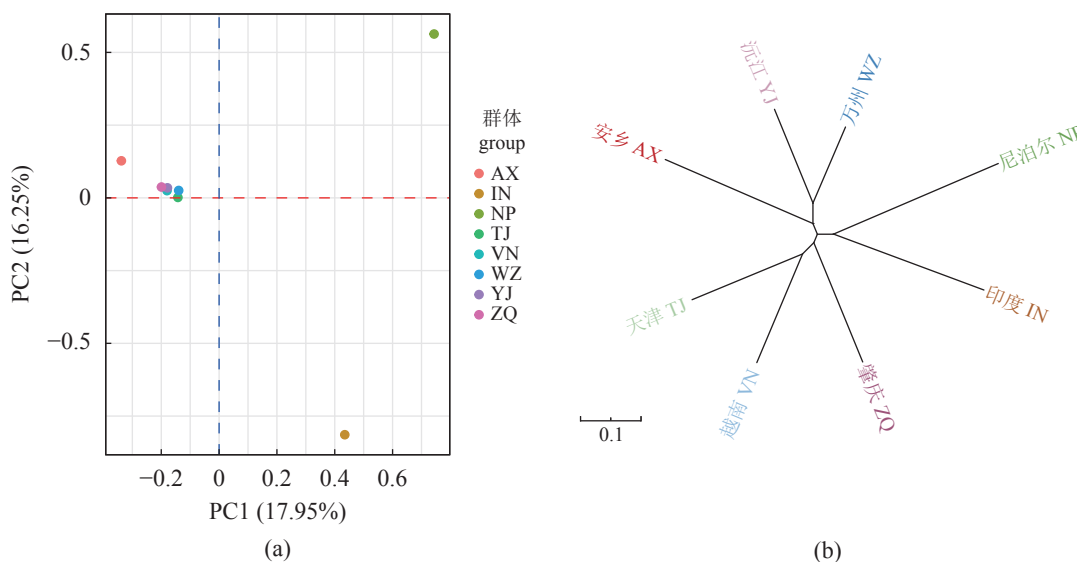


图 4 8 个草鱼种群遗传结构分析图

(a) 种群 PCA 分析图, (b) 种群进化树分析图。

Fig. 4 Population genetic structure of *C. idella*

(a) principal component analysis of eight groups, (b) phylogenetic tree of eight groups.

表 4 草鱼 8 个种群的核苷酸多样性

Tab. 4 Nucleotide diversity of eight populations of *C. idella*

种群 group	核苷酸多样性 π
安乡 AX	4.48E-03
印度 IN	3.76E-03
尼泊尔 NP	3.64E-03
天津 TJ	4.58E-03
越南 VN	4.54E-03
万州 WZ	4.89E-03
沅江 YJ	4.82E-03
肇庆 ZQ	4.49E-03

验根据种群聚类结果将上述种群划分为 3 个亚群。采用 F_{ST} 值和 XP-CLR 方法, 均以前 1% 为阈值, 对受选择区域进行筛选。结果显示, 相对于第三亚群, 长江流域和南亚流域种群分别筛选出 177 个区间, 其中分别包含 1 633 个和 686 个基因。使用 XP-CLR 法筛选受选择区域时, 长江流域和南亚流域相对于第三亚群各筛选出 471 个和 874 个区间, 分别包含 1 664 个和 2 909 个基因(图 5)。

受选择基因 GO 富集分析 通过对 2 种方法截取到的基因取交集, 发现长江流域和南亚流域种群相对于第三亚群分别存在 123 个和 90 个受选择基因。GO 富集分析结果显示, 长江流域种群的受选择基因主要富集在免疫反应

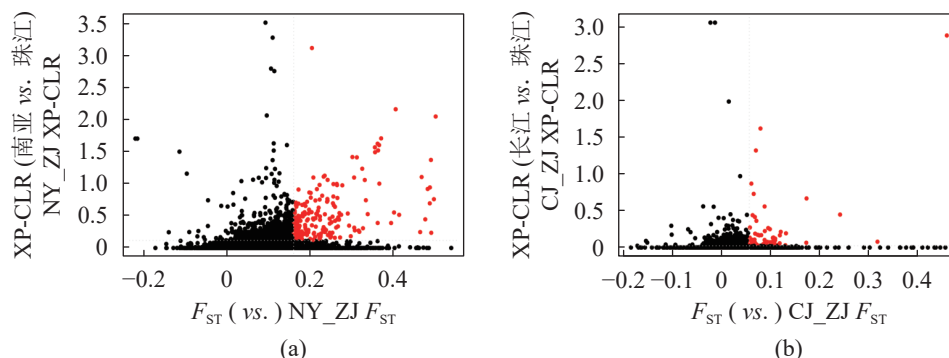


图5 受选择区域分析图

(a) 南亚流域种群相对于珠江流域种群选择性区域图, (b) 长江流域种群相对于珠江流域种群选择性区域图。NY. 南亚, ZJ. 珠江, CJ. 长江。

Fig. 5 Manhattan plot of selective analysis

(a) the selective region of the South Asian basin relative to the ZJ group, (b) the selective region of the Yangtze River basin relative to the ZJ group. NY. South Asia, ZJ. Zhujiang River, CJ. the Yangtze River.

调节、细胞电位调节、心肌细胞调节等生物过程; 细胞组分主要富集在钠离子通道复合物、PcG 蛋白复合体等; 分子功能主要富集在控制钠离子通道活性、成纤维细胞生长因子结合等(图 6-a); 南亚流域种群的受选择基因主要富集在免疫反应调节、代谢等生物过程; 细胞组分主要富集在线粒体外膜、脂滴、细胞器外膜等; 分子功能主要富集在核糖核苷酸结合等(图 6-b)。

***gimap8* 基因分析** 通过受选择分析, 实验共鉴定出 158 个基因, 通过 snpEFF 注释结果, 发现 13 个基因的结构受到了重要影响。实验观察到, *gimap8* 基因在不同流域结构差异显著(表 5), LG2 的 50 137 315 bp 碱基突变导致基因发生可变剪切; 而在 LG2 的 50 137 464 bp 突变可导致提前终止(图 7)。

在越南和印度种群中, P-loop NTPase 超家族结构域和 GTP/Mg²⁺连接位点出现了丢失, 而长江流域种群仅丢失了 GTP/Mg²⁺连接位点(图 8)。对上述序列进行蛋白三维结构预测, 发现南亚以外的种群均出现了不同程度的 α 螺旋和 β 折叠缺失(图 9), 越南与沅江种群均缺少 P-loop_NTPase super family 结构域, 同时沅江群体缺少 2 个 β 折叠, 这些折叠构成一环状结构(图 8)。

3 讨论

草鱼以其广泛的生态适应性而闻名, 超过一半的引入国家已形成了各自独特的种群^[39]。种群分化通常是由地理隔离、人工选择和遗传漂变等因素引起的^[40-41], 已有研究证实不同纬

度的草鱼种群存在遗传差异^[42]。为验证草鱼各种群是否因地理分布而分化, 并探究其分子机制, 本研究对中国 4 个省、直辖市(湖南、重庆、天津和广东)的 5 个种群、越南、印度和尼泊尔 3 个种群, 共计 8 个种群进行全基因组重测序分析。系统发育树分析结果显示, 印度和尼泊尔种群聚为一支, 这可能与它们的地理位置相近有关。主成分分析(PCA)结果显示, 2 种群间距离较远, 暗示它们已经发生了遗传分化。推测是由于它们位于不同河流中, 种群间的连通性较差导致。有研究证实, 水系的差异会影响产卵生境, 从而限制基因交流^[43], 8 个种群被分为 3 个亚群也能体现这一点。李文俊等^[44]证明了珠江和长江水系间复杂的地形地貌形成了明显的地理隔离, 导致 2 个光倒刺鲃(*Spinibarbus hollandi*)种群间发生明显分化。本研究结果揭示了印度和尼泊尔种群因地理隔离而与其他种群产生独立的遗传分化。

基因变异与环境压力的显著关联性已被广泛研究^[45-46], 例如, 针对大西洋鳕种群个体多样性的研究发现, 染色体重排区域反映了环境选择压力^[47]。本研究发现的基因主要富集在细胞黏附分子信号通路。细胞黏附分子(CAM)被认为在免疫反应中起着重要的作用, 主要参与炎症、伤口修复、止血和免疫反应等多种重要生理功能^[48]。在免疫反应的过程中, CAM 促进免疫细胞的迁移^[49], 鳃弧菌(*Vibrio anguillarum*)感染红鲷(*Red snapper*)的转录组分析中, 发现头肾和脾脏的差异表达基因均显著富集于细胞黏附分子通路^[50]。同样, 有研究发现嗜水气单胞菌

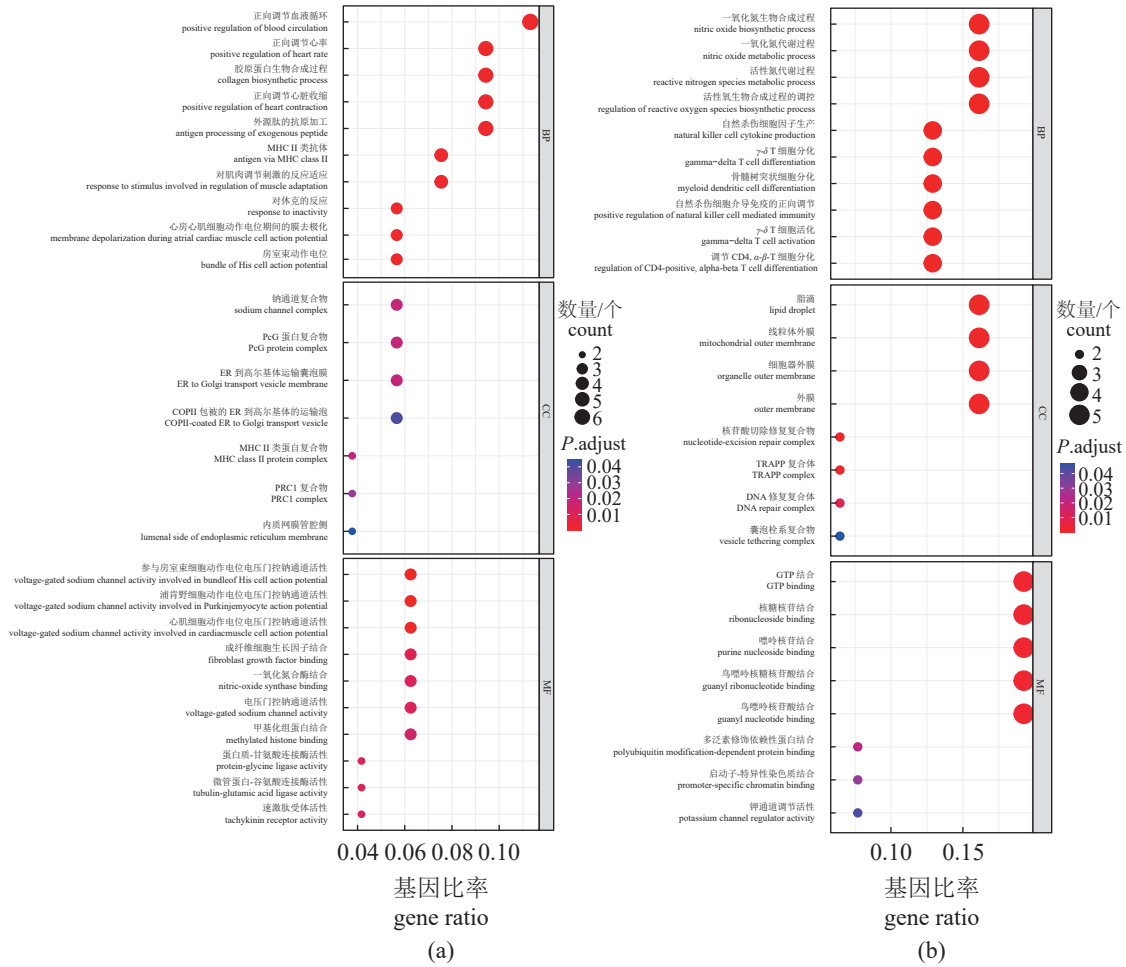


图 6 受选择基因 GO 富集图

(a) 长江流域相对于第三亚群受选择基因, (b) 南亚流域相对于第三亚群受选择基因。

Fig. 6 Dot plot of enriched GO pathway of selective genes

(a) the Yangtze River basin relative to the third group, (b) the South Asian basin relative to the third group.

表 5 筛选基因列表

Tab. 5 Filter gene list

基因ID gene ID	描述 description	基因缩写 abbreviation
Cide001884RA	low affinity immunoglobulin gamma Fc region receptor	<i>fcrl5</i>
Cide002127RA	GTPase IMAP family member 8-like	<i>gimap8</i>
Cide002130RA	GTPase IMAP family member 7-like	<i>gimap7</i>
Cide002149RA	GTPase IMAP family member 8-like	<i>gimap8</i>
Cide002150RA	GTPase IMAP family member 7-like	<i>gimap7</i>
Cide002151RA	GTPase IMAP family member 8-like	<i>gimap8</i>
Cide002154RA	GTPase IMAP family member 8-like	<i>gimap8</i>
Cide003438RA	cAMP responsive element binding protein 3-like 1	<i>creb3l1</i>
Cide007973RA	upstream binding transcription factor	<i>ubtfl</i>
Cide020016RA	class II histocompatibility antigen	H2-Aa
Cide020445RA	dynactin 2 (p50)	<i>dctn2</i>
Cide030887RA	protogenin homolog b (<i>Gallus gallus</i>)	<i>prtg</i>
Cide035525RA	DNA/RNA non-specific endonuclease	

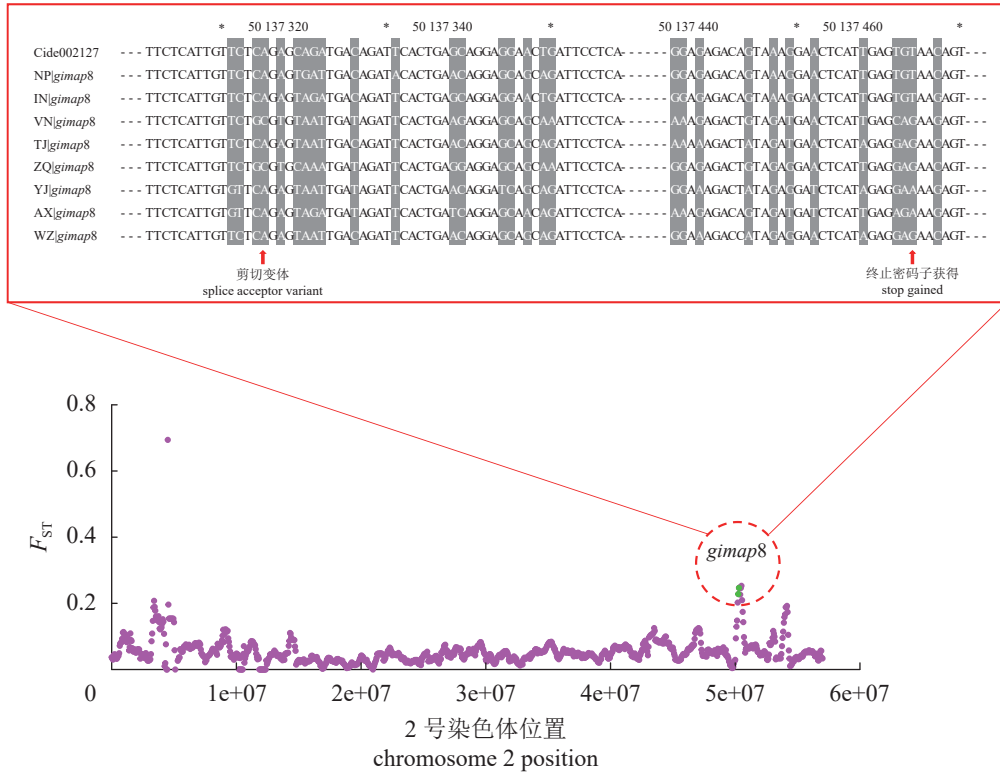


图 7 *gimap8* 基因关键变异点图

Fig. 7 Key variation point of *gimap8* gene



图 8 3 个代表种群 GIMAP8 蛋白序列比对图

图中波浪线为 GTP/Mg²⁺ 连接位点; 实线框为 P-loop_NTPase super family 结构域; 虚线框为 DUF5401 super family 结构域; 灰色阴影为氨基酸相似区域; 黑色阴影为氨基酸一致序列。

Fig. 8 Alignment of *gimap8* protein sequences of three representative populations

The wavy line in the figure is the GTP/Mg²⁺ binding site; the solid frame is the P-loop_NTPase super family domain; the dotted box is the DUF5401 super family domain; shaded gray is an area of amino acid-likeness; black shade is an amino acid consistent sequence.

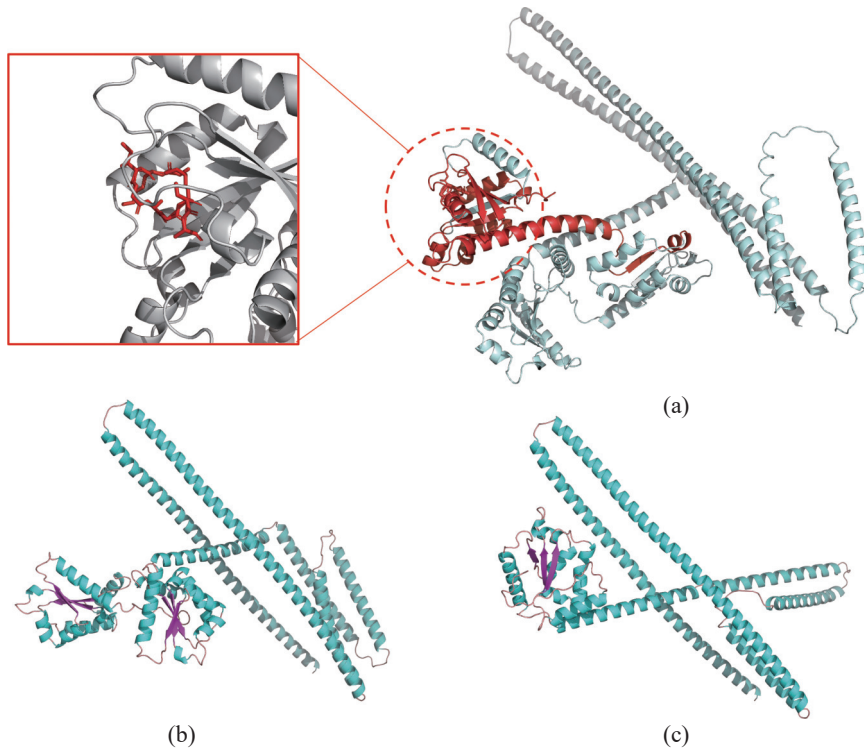


图9 蛋白三维结构对比图

(a) 印度 (IN), (b) 越南 (VN), (c) 沅江 (YJ); 红色为 VN 缺失部分, 红色支架图像为 GTP/Mg²⁺ 连接位点, 紫色为 α 螺旋, 蓝色为 β 折叠。

Fig. 9 Comparison of three-dimensional protein structures

(a) the group of IN, (b) the group of VN, (c) the group of YJ; red is the missing part of VN, the red scaffold image is the GTP/Mg²⁺ junction site, the purple is the α -helix, and the blue is the β -sheet.

(*Aeromonas hydrophila*) 感染青鱼 (*Mylopharyngodon piceus*) 后, 脾脏差异表达基因也富集在细胞黏附分子和病原体识别等通路^[51]。因此, 实验推测草鱼在适应各个环境过程中产生了免疫相关基因的变异, 这类变异提升草鱼适应不同环境的能力。

本研究截取阈值 1% 的受选择区域, 通过 GO 富集发现, 受选择基因显著富集在免疫反应调节、细胞电位调节、心肌细胞调节、代谢等生物过程之中。与四种变异的非同义突变涉及基因的 KEGG 富集结果相比, 二者均体现了环境选择压力的适应。同时, 实验利用 F_{ST} 和 XP-CLR 共同定位受选择区域, 结合 SNP 注释结果和群体结构信息发现了 *gimap* (GTPase IMAP family) 家族基因, 其中有 4 个被注释为 *gimap8* 基因, 这些基因在 T 细胞发育的早期和晚期、B 细胞发育的晚期以及外周 T 细胞和 B 细胞中大量表达^[52], 它们的缺失将导致成熟 T 细胞在体外的凋亡延迟。研究发现, 野生大西洋鲑 (*Salmo salar*) 和养殖群体之间 SV 等位基因

频率的变化与驯化过程中基因受选择相关, 这些 SV 影响了多种功能的基因, 如精子游动力基因 *catsperb*、免疫 B 细胞存活和信号传导基因 *gimap8*、昼夜节律控制代谢基因 *nr1d2* 等^[53]。Yáñez 等^[54] 通过全基因组关联研究分析, 在鲑鱼抗立克次体综合征 (salmon rickettsial syndrome, SBS) 的群体中鉴定获得 *gimap* 基因。上述研究均表明 *gimap8* 对抗性性状具有重要作用。实验中发现 *gimap8* 具有 2 个变异位点, 使长江流域、珠江流域、越南群体缺少部分 P 环 NTP 酶结构域 (P-loop NTPase)。该结构域的酶水解 β - γ 磷酸键, NTP 水解的自由能通常用于诱导其他分子的构象变化, 进而在调节信号转导、细胞分裂、DNA 复制、RNA 转录和翻译等关键生物过程中起作用^[55]。同时, P-loop NTPase 还涉及一种柔性结构, 该结构可以影响底物结合, 或者与另一种结构域产生相互作用而发生构象变化。反之, 构象变化可以调控进入活性位点的底物, 进而可以在不同程度上促进激酶活性位点和硫化酶位点的催化作用。缺少 P-

loop 结构域可能会导致许多生物过程的异常或中断^[56]。例如, ATP 合成酶缺失 P-loop 结构域可能无法催化 ATP 合成, 影响能量代谢^[57]; 小 G 蛋白缺失 P-loop 结构域可能无法正确结合 GTP, 影响信号转导^[58]。因此 *gimap8* 基因 P-loop 结构域的变动对于下游信号的传递产生影响。

4 结论

本研究通过对天津、肇庆、沅江、安乡、万州、尼泊尔、越南和印度 8 个草鱼种群进行全基因组重测序分析, 揭示了这些种群的遗传结构和适应性变异。实验发现, 4 种类型的基因变异主要富集在细胞黏附分子通路。种群结构分析进一步揭示了该 8 个草鱼种群可分为 3 个亚群, 这种分化可能是受地理环境不同所影响。通过结合选择清除分析和 SNP 注释结果, 实验确定了与免疫相关的候选基因 *gimap8*。其变异反映了草鱼种群在适应不同环境压力时的分子机制。研究结果揭示了不同草鱼种群在适应性免疫方面的分子机制差异, 这些发现不仅有助于深入理解草鱼面对不同环境的适应性进化方式, 而且为草鱼育种工作提供了重要的理论依据和新的思路。

参考文献 (References):

[1] FAO. The state of world fisheries and aquaculture 2022. Towards blue transformation[M]. Rome: FAO, 2022.

[2] Wilson R P. Handbook of nutrient requirements of finfish (1991)[M]. Boca Raton: CRC Press, 1991.

[3] 丁雪梅, 颜岳辉, 李强, 等. 南盘江破碎生境中子陵吻虾虎鱼的遗传多样性[J]. 水产科学, 2020, 39(6): 852-862.

Ding X M, Yan Y H, Li Q, *et al.* Genetic diversity of goby *Rhinogobius giurinus* in the fragmented habitat in nanpan river[J]. Fisheries Science, 2020, 39(6): 852-862 (in Chinese).

[4] 汪焕, 江河, 段国庆, 等. 安徽省草鱼养殖群体遗传多样性及遗传结构分析[J]. 安徽农业大学学报, 2020, 47(1): 25-29.

Wang H, Jiang H, Duan G Q, *et al.* The genetic diversity and population structure of cultured *Ctenopharyngodon idella* in Anhui Province[J]. Journal of Anhui Agricultural University, 2020, 47(1): 25-29 (in Chinese).

[5] 翟东东, 蔡金, 喻记新, 等. 长江上游 5 个草鱼群体的遗传多样性[J]. 淡水渔业, 2020, 50(5): 81-87.

Zhai D D, Cai J, Yu J X, *et al.* Genetic diversity of five *Ctenopharyngodon idella* populations in the upper reaches of the

Yangtze River[J]. Freshwater Fisheries, 2020, 50(5): 81-87 (in Chinese).

[6] Wang L J, Wu Z H, Zou C C, *et al.* Genetic diversity and signatures of selection in the mito-gynogenetic olive flounder *Paralichthys olivaceus* revealed by genome-wide SNP markers[J]. Aquaculture, 2022, 553: 738062.

[7] 罗慧, 方弟安, 何苗, 等. 基于 SNP 标记的青海湖裸鲤遗传多样性及种群结构研究[J]. 南方水产科学, 2023, 19(1): 86-96.

Luo H, Fang D A, He M, *et al.* Genetic diversity and population structure of *Gymnocypris przewalskii* based on SNP markers[J]. South China Fisheries Science, 2023, 19(1): 86-96.

[8] 薛艳会, 刘尊雷, 李圣法, 等. 南黄海和东海中南部小黄鱼种群形态分化[J]. 中国水产科学, 2021, 28(9): 1162-1174.

Xue Y H, Liu Z L, Li S F, *et al.* Morphological differentiation of small yellow croaker populations in the south Yellow Sea and the south-central East China Sea[J]. Journal of Fishery Sciences of China, 2021, 28(9): 1162-1174 (in Chinese).

[9] Chen H, Patterson N, Reich D. Population differentiation as a test for selective sweeps[J]. Genome Research, 2010, 20(3): 393-402.

[10] 何金明, 孙智禹, 刘玉玲, 等. 新疆黑蜂基因组选择信号分析[J]. 中国畜牧兽医, 2022, 49(5): 1786-1795.

He J M, Sun Z Y, Liu Y L, *et al.* Genome-wide selection signal analysis of Xinjiang black bee[J]. China Animal Husbandry & Veterinary Medicine, 2022, 49(5): 1786-1795 (in Chinese).

[11] Clucas G V, Lou R N, Therkildsen N O, *et al.* Novel signals of adaptive genetic variation in northwestern Atlantic cod revealed by whole-genome sequencing[J]. Evolutionary Applications, 2019, 12(10): 1971-1987.

[12] Narum S R, Di Genova A, Micheletti S J, *et al.* Genomic variation underlying complex life-history traits revealed by genome sequencing in Chinook salmon[J]. Proceedings of the Royal Society B: Biological Sciences, 2018, 285(1883): 20180935.

[13] Iannucci A, Benazzo A, Natali C, *et al.* Population structure, genomic diversity and demographic history of Komodo dragons inferred from whole-genome sequencing[J]. Molecular Ecology, 2021, 30(23): 6309-6324.

[14] Fujino K, Shirasawa K. Fine-scale genetic structure of the rice landrace population in Japan[J]. Molecular Genetics and Genomics, 2022, 297(3): 711-718.

[15] Wang G, Yuan Y, Li X, *et al.* Genome-wide association and genomic selection studies on feed efficiency with fishmeal-free feed in large yellow croaker (*Larimichthys crocea*)[J/OL].

- SSRN, 2024, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4947329.
- [16] Sun C F, Zhang X H, Dong J J, *et al.* Whole-genome resequencing reveals recent signatures of selection in five populations of largemouth bass (*Micropterus salmoides*)[J]. *Zoological Research*, 2023, 44(1): 78-89,
- [17] Du J X, Li S J, Shao J Q, *et al.* Genetic diversity analysis and development of molecular markers for the identification of largemouth bass (*Micropterus salmoides* L.) based on whole-genome re-sequencing[J]. *Frontiers in Genetics*, 2022, 13: 936610, doi: 10.3389/fgene.2022.936610.
- [18] 魏上, 谢玲莉, 朱华, 等. 亚洲草鱼群体的差异甲基化分析 [J]. *水产学报*, 2023, 47(3): 99-111.
Wei S, Xie L L, Zhu H, *et al.* Differential methylation analysis of *Ctenopharyngodon idella* populations in Asia[J]. *Journal of Fisheries of China*, 2023, 47(3): 99-111 (in Chinese).
- [19] Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform[J]. *Bioinformatics*, 2009, 25(14): 1754-1760,
- [20] Li H, Handsaker B, Wysoker A, *et al.* The sequence alignment/map format and SAMtools[J]. *Bioinformatics*, 2009, 25(16): 2078-2079,
- [21] McKenna A, Hanna M, Banks E, *et al.* The genome analysis toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data[J]. *Genome Research*, 2010, 20(9): 1297-1303,
- [22] Layer R M, Chiang C, Quinlan A R, *et al.* LUMPY: a probabilistic framework for structural variant discovery[J]. *Genome Biology*, 2014, 15(6): R84,
- [23] Abyzov A, Urban A E, Snyder M, *et al.* CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing[J]. *Genome Research*, 2011, 21(6): 974-984,
- [24] Wang K, Li M Y, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data[J]. *Nucleic Acids Research*, 2010, 38(16): e164,
- [25] Cingolani P, Platts A, Wang L L, *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff[J]. *Fly*, 2012, 6(2): 80-92,
- [26] Nguyen L T, Schmidt H A, von Haeseler A, *et al.* IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies[J]. *Molecular Biology and Evolution*, 2015, 32(1): 268-274,
- [27] Letunic I, Bork P. Interactive tree of life (iTOL) v5: an online tool for phylogenetic tree display and annotation[J]. *Nucleic Acids Research*, 2021, 49(W1): W293-W296,
- [28] Purcell S, Neale B, Todd-Brown K, *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses[J]. *The American Journal of Human Genetics*, 2007, 81(3): 559-575,
- [29] Danecek P, Auton A, Abecasis G, *et al.* The variant call format and VCFtools[J]. *Bioinformatics*, 2011, 27(15): 2156-2158.
- [30] Weir B S, Cockerham C C. Estimating *F*-statistics for the analysis of population structure[J]. *Evolution*, 1984, 38(6): 1358-1370,
- [31] Cantalapiedra C P, Hernández-Plaza A, Letunic I, *et al.* eggNOG-mapper v2: functional annotation, orthology assignments, and domain prediction at the metagenomic scale[J]. *Molecular Biology and Evolution*, 2021, 38(12): 5825-5829,
- [32] Wu T Z, Hu E Q, Xu S B, *et al.* clusterProfiler 4.0: a universal enrichment tool for interpreting omics data[J]. *The Innovation*, 2021, 2(3): 100141,
- [33] Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11[J]. *Molecular Biology and Evolution*, 2021, 38(7): 3022-3027,
- [34] Nicholas K B, Nicholas Jr H B, Deerfield II D W. GeneDoc: analysis and visualization of genetic variation[N]. *Embnew News*, 1997-04-14.
- [35] Lu S N, Wang J Y, Chitsaz F, *et al.* CDD/SPARCLE: the conserved domain database in 2020[J]. *Nucleic Acids Research*, 2020, 48(D1): D265-D268,
- [36] Mirdita M, Schütze K, Moriwaki Y, *et al.* ColabFold: making protein folding accessible to all[J]. *Nature Methods*, 2022, 19(6): 679-682,
- [37] Cramer P. AlphaFold2 and the future of structural biology[J]. *Nature Structural & Molecular Biology*, 2021, 28(9): 704-705,
- [38] DeLano W L. PyMOL: an open-source molecular graphics tool[N]. *CCP4 News Protein Crystallogr*, 2002.
- [39] Xia J H, Liu F, Zhu Z Y, *et al.* A consensus linkage map of the grass carp (*Ctenopharyngodon idella*) based on microsatellites and SNPs[J]. *BMC Genomics*, 2010, 11(1): 135,
- [40] Van Oosterhout C, Joyce D A, Cummings S M, *et al.* Balancing selection, random genetic drift, and genetic variation at the major histocompatibility complex in two wild populations of guppies (*Poecilia reticulata*)[J]. *Evolution*, 2006, 60(12): 2562-2574.

- [41] Wang X D, Tong L T, Deng J J, *et al.* Insights into historical drainage evolution based on the phylogeography of *Schizopygopsis malacanthus* Herzenstein (Cypriniformes, Cyprinidae) across the upper and middle Yalong River drainage in the Hengduan Mountains region, southwest China[J]. *Global Ecology and Conservation*, 2022, 35: e02084.
- [42] Wang L, Shen Y B, Fu J J, *et al.* Genomic divergence, introduction history and latitudinal adaptation of grass carp[J/OL]. bioRxiv, 2016, <https://www.biorxiv.org/content/10.1101/066092>.
- [43] Nordahl O, Koch-Schmidt P, Sunde J, *et al.* Genetic differentiation between and within ecotypes of pike (*Esox lucius*) in the Baltic Sea[J]. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 2019, 29(11): 1923-1935.
- [44] 李文俊, 李强, 钟良明, 等. 基于线粒体 DNA 控制区序列的珠江和长江水系光倒刺鲃群体遗传变异分析 [J]. *南方农业学报*, 2021, 52(11): 3121-3129.
Li W J, Li Q, Zhong L M, *et al.* Analysis of genetic variation among *Spinibarbus hollandi* in the Pearl River and the Yangtze River based on mitochondrial DNA control region sequences[J]. *Journal of Southern Agriculture*, 2021, 52(11): 3121-3129 (in Chinese).
- [45] Chen L, Hu Y F, He J, *et al.* Responses of the proteome and metabolome in livers of zebrafish exposed chronically to environmentally relevant concentrations of microcystin-LR[J]. *Environmental Science & Technology*, 2017, 51(1): 596-607.
- [46] Pereiro P, Rey-Campos M, Figueras A, *et al.* An environmentally relevant concentration of antibiotics impairs the immune system of zebrafish (*Danio rerio*) and increases susceptibility to virus infection[J]. *Frontiers in Immunology*, 2023, 13: 1100092.
- [47] Kess T, Bentzen P, Lehnert S J, *et al.* Modular chromosome rearrangements reveal parallel and nonparallel adaptation in a marine fish[J]. *Ecology and Evolution*, 2020, 10(2): 638-653.
- [48] Ley K, Laudanna C, Cybulsky M I, *et al.* Getting to the site of inflammation: the leukocyte adhesion cascade updated[J]. *Nature Reviews Immunology*, 2007, 7(9): 678-689.
- [49] Meager A. Cytokine regulation of cellular adhesion molecule expression in inflammation[J]. *Cytokine & Growth Factor Reviews*, 1999, 10(1): 27-39.
- [50] Song L F, Dong X Z, Hu G B. Transcriptome analysis of red sea bream (*Pagrus major*) head kidney and spleen infected by *Vibrio anguillarum*[J]. *Aquaculture Reports*, 2021, 21: 100789.
- [51] Zhang X S, Shen Y B, Xu X Y, *et al.* Transcriptome analysis and histopathology of black carp (*Mylopharyngodon piceus*) spleen infected by *Aeromonas hydrophila*[J]. *Fish & Shellfish Immunology*, 2018, 83: 330-340.
- [52] Webb L M C, Pascall J C, Hepburn L, *et al.* Generation and characterisation of mice deficient in the multi-GTPase domain containing protein, GIMAP8[J]. *PLoS One*, 2014, 9(10): e110294.
- [53] Bertolotti A C, Layer R M, Gundappa M K, *et al.* The structural variation landscape in 492 Atlantic salmon genomes[J]. *Nature Communications*, 2020, 11(1): 5176.
- [54] Yáñez J M, Yoshida G M, Parra Á, *et al.* Comparative genomic analysis of three salmonid species identifies functional candidate genes involved in resistance to the intracellular bacterium *Piscirickettsia salmonis*[J]. *Frontiers in Genetics*, 2019, 10: 665.
- [55] Saraste M, Sibbald P R, Wittinghofer A. The P-loop-a common motif in ATP- and GTP-binding proteins[J]. *Trends in Biochemical Sciences*, 1990, 15(11): 430-434.
- [56] Deyrup A T, Krishnan S, Cockburn B N, *et al.* Deletion and site-directed mutagenesis of the ATP-binding motif (P-loop) in the bifunctional murine ATP-sulfurylase/adenosine 5'-phosphosulfate kinase enzyme[J]. *Journal of Biological Chemistry*, 1998, 273(16): 9450-9456.
- [57] Koonin E V. A superfamily of ATPases with diverse functions containing either classical or deviant ATP-binding motif[J]. *Journal of Molecular Biology*, 1993, 229(4): 1165-1174.
- [58] Bourne H R, Sanders D A, McCormick F. The GTPase superfamily: conserved structure and molecular mechanism[J]. *Nature*, 1991, 349(6305): 117-127.

Genetic diversity and adaptation mechanism of *Ctenopharyngodon idella* in Asian by whole-genome resequencing

XIE Lingli^{1,2,3}, SHEN Yubang^{1,2,3}, GUI Lang^{1,2,3}, XU Xiaoyan^{1,2,3*}, LI Jiale^{1,2,3*}

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

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

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

Abstract: This study examines the genetic diversity and adaptive evolution of *Ctenopharyngodon idella* populations across Asia, aiming to enhance our comprehension of the molecular evolution mechanisms and immune adaptations in *C. idella*. We conducted whole-genome resequencing of *C. idella* populations from four Chinese provinces (Hunan, Chongqing, Tianjin, Guangdong) and three Asian countries (Nepal, Vietnam, India). Employing population genetics methodologies, we analysed population structure, performed principal component analysis, constructed phylogenetic trees, assessed genetic differentiation, and conducted selective sweep analyses. The Wanzhou population displayed the highest genetic diversity, whereas the Nepal population exhibited the lowest. Cluster analysis revealed significant genetic similarities among the five populations in Vietnam and China. A total of 6 548 523 SNP sites were identified, predominantly in intergenic regions and enriched in signaling pathways such as cell adhesion molecules, hematopoietic stem cells, and DNA recombination repair. Selective sweep analysis highlighted genes enriched in pathways related to immune response regulation, cardiomyocyte regulation, and sodium ion channel activity. Utilizing SNP annotation data, we identified *gimap8* in Indian and Nepal populations, crucial for T cells and B cells development. *C. idella* populations in Anxiang, Yuanjiang, and Wanzhou in China, are genetically diverse, while those in Nepal and India show significant differentiation from other populations. We preliminarily identified the potential adaptive genes *gimap8*. Concurrently, mutations in *gimap8*'s base affected the gene's structure in three populations. This study lays a foundation for further exploration of the molecular evolution mechanism of *C. idella* in various habitats, reflecting the molecular mechanisms by which *C. idella* populations adapt to different environmental stresses. The results reveal the differences in the molecular mechanisms of adaptive immunity in different *C. idella* populations. These findings not only deepen our understanding of *C. idella*'s adaptive evolution in diverse environments but also provide a significant theoretical basis for *C. idella* breeding and offer new insights.

Key words: *Ctenopharyngodon idella*; whole-genome resequencing; genetic diversity; selective sweep

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

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

Funding projects: China's Agricultural Research System (CARS-45-03)