



基于简化基因组测序筛选白斑狗鱼耐热性状关联的 InDel 标记

海萨·艾也力汗¹, 张 钰¹, 杨博文¹, 咸玉兰¹, 高 攀¹, 沈玉帮^{2*}

(1. 新疆维吾尔自治区水产科学研究所, 新疆维吾尔自治区 乌鲁木齐 830000;

2. 上海海洋大学水产与生命学院, 上海 201306)

摘要: 为了能给白斑狗鱼耐高温性状的改良提供有效的分子标记, 本研究基于白斑狗鱼热敏感组 (109 尾) 和耐高温组 (103 尾) 进行简化基因组测序, 对 25 个染色体中的 InDel 标记, 以前两个 PCA 为协变量, 利用 MLM 模型 (Masked Language Mode) 与白斑狗鱼的耐热性状进行了关联分析。结果显示, 大部分 InDel 分布在内含子 (63.69%), 外显子分布的 InDel 位点较少 (1.30%)。通过 GWAS 分析, 发现 5 个位点与白斑狗鱼耐热性状显著关联, 分别位于 *sqfb* 基因、未知基因 LOC117593903 和 *CLSTN2* 基因内含子中。其中 9N del、4N del-1 和 4N del-2 三个 InDel 位点均位于 *CLSTN2* 基因第 3 内含子, 这 3 个位点在 212 尾个体中基因型分布高度连锁。本研究中发现的 5 个 InDel 突变可能会对白斑狗鱼的耐热性状产生显著的影响, 可作为白斑狗鱼耐热性状改良的候选分子标记。进一步在验证群体中利用 KASP 技术对部分位点进行了验证。发现 9N del 位点 DD 基因型个体在热敏感组中占优势, DI 基因型个体在耐高温组中占优势, 与简化基因组测序结果基本一致。位于 *CLSTN2* 基因第 2 内含子的 3 个 InDel 位点可能影响 *CLSTN2* 基因的转录, 该基因可能是白斑狗鱼耐高温性状相关的重要候选基因。研究结果为白斑狗鱼分子标记辅助育种提供了理论依据, 为白斑狗鱼耐热性状的改良提供了候选分子标记。

关键词: 白斑狗鱼; 耐热性状; InDel; 全基因组关联分析

中图分类号: Q 785; S 917

文献标志码: A

白斑狗鱼 (*Esox lucius*) 隶属狗鱼目 (Esociformes) 狗鱼科 (Esocidae) 狗鱼属 (*Esox*)。分布于北纬 45° 以北的亚洲、欧洲和北美洲的北极圈周边地区^[1], 中国境内仅分布于额尔齐斯河流域。该流域是我国唯一的白斑狗鱼种质资源宝库。白斑狗鱼生长速率快、肉质鲜美、营养价值高, 已成为新疆地区, 乃至河北等地主要养殖品种。由于额尔齐斯河位于白斑狗鱼自然分布区域的最南端, 因此全球气候变暖引起的极端天气对我国境内白斑狗鱼养殖和野生群体的影响较大。已有研究表明, 水温超过 18 °C 时白斑狗鱼受精卵孵

化率下降, 水温超过 27 °C 幼鱼的摄食和生长受到影响^[2]。因此白斑狗鱼热耐受性研究和耐高温性状的遗传改良已成为白斑狗鱼种质资源保护和产业化可持续发展亟待解决的问题之一。

基于高通量测序技术, 利用全基因组关联分析 (GWAS 分析) 方法筛选与目标性状关联 InDel 标记和候选基因已广泛应用于牛^[3]、山羊^[4]、鸡^[5]、羊^[6] 等动物的遗传改良研究中。在水产动物中, 伞利择等^[7] 基于卵形鲳鲀 (*Trachinotus ovatus*) 基因组重测序的 InDel 标记挖掘与耐低氧性状关联分析, 发现 3 个 InDel 位点接近显著阈值, 注释分

收稿日期: 2023-02-22 修回日期: 2023-04-19

资助项目: 国家自然科学基金 (32060826)

第一作者: 海萨·艾也力汗 (照片), 从事水产动物种质资源保护与选育研究, E-mail: hbahjan@126.com

通信作者: 沈玉帮, 从事鱼类种质资源与遗传育种研究, E-mail: ybshen@shou.edu.cn



析获得了9个候选基因。陈静^[8]基于全基因组重测序获得了鲂属 (*Megalobrama*) 鱼类高质量的 SNP、InDel 等变异位点信息, 并构建了鲂属鱼类群体的基因组变异数据库。该数据库在描述基因型-表型图谱以及亲鱼选择中发挥了重要的作用。黄智康等^[9]基于 InDel 标记建立了斑节对虾 (*Penaeus monodon*) 早期性别鉴定方法。竞争性等位基因特异性 PCR (Kompetitive allele specific PCR, KASP) 技术是英国政府化学家实验室 (Laboratory of the Government Chemist, LGC) 开发的检测 SNP 和 InDel 的第三代标记技术, 具有高通量、成本低、准确性高等特点^[10], 已成为利用 SNP 进行育种的首选标记^[11]。

本研究为了筛选白斑狗鱼耐热性状关联的分子标记和候选基因, 通过热胁迫获得2个极端群体样品, 结合简化基因组测序和全基因组关联分析筛选出与目标性状关联的 InDel 位点, 利用 KASP 技术对显著关联位点进行验证, 挖掘与白斑狗鱼耐热性状显著关联的 InDel 标记和候选基因, 为白斑狗鱼耐热性状的遗传改良和分子机制研究提供理论依据。

1 材料与amp;方法

1.1 实验鱼

简化基因组测序材料: 实验鱼亲鱼采自额尔齐斯河北屯段、布尔津县段以及乌伦古湖。2021年4月12日至4月14日通过人工采集和授精获得受精卵。受精卵孵化、苗种培育在同一循环系统的14个矩形槽(长×宽×高: 400 cm×60 cm×80 cm)中进行培育至幼鱼, 随机挑选450尾(平均全长90.45 mm, 平均体重5.85 g)作为下一步热胁迫材料。标记验证材料: 2022年4月通过人工繁育获得受精卵, 与上一步相同的方法培育至幼鱼期, 随机挑选400尾(平均全长83.31 mm, 平均体重3.96 g/尾)作为下一步的热胁迫材料。本研究在实验过程中, 严格按照实验动物管理条例和使用伦理规范执行。

1.2 热胁迫实验

苗种培育至幼鱼期(2月龄)后随机挑选450尾在一个长×宽×高为400 cm×60 cm×80 cm的矩形槽中暂养7 d后开展热胁迫实验。热胁迫实验期间利用3个1 000 W的智能PID变频加热棒控温, 实验全过程持续充气, 提供足够的活体饵料, 每

天换1/3相同温度的水。升温策略: 暂养期间水温(26.0±0.5) °C, 在30.0 °C之前以1 °C/12 h的速率升温, 在30.0 °C之后以1 °C/24 h的速率升温至实验结束。每2小时测量一次水温和溶解氧。我们对实验鱼进行了24 h连续监测, 及时捞出死鱼, 记录体长、全长、体重和累计存活时间, 收集肌肉组织, 浸泡于无水乙醇, 保存于-20 °C。最早死亡的110尾和最后死亡的110尾实验鱼分别被认为是热敏感组和耐高温组, 进行简化基因组测序。KASP验证材料的热胁迫实验同上, 实验鱼在暂养期死亡13尾, 对387尾实验鱼开展热胁迫实验。最先死亡的24尾视为热敏感组, 最后死亡的24尾视为耐高温组, 采集鳍条组织, 用酚-氯仿法抽提基因组DNA, -20 °C保存备用。

1.3 简化基因组测序

热胁迫实验筛选出的热敏感组和耐高温组肌肉组织送上海欧易生物医学科技有限公司, 通过随机测序式基因型检测 (genotyping-by-sequencing, GBS) 技术进行简化基因组测序。采用限制性内切酶 PstI-HF/MspI 对 DNA 进行酶切, 酶切后的片段两端用 T4 连接酶加接头和 barcode; 使用改进的磁珠回收系统, 通过调整磁珠溶液与连接产物的体积比来回收 300~700 bp 的片段; 对回收片段使用高保真酶进行 PCR 扩增; 使用 Qubit 测定 PCR 产物浓度, 浓度需大于 5 ng/μL; 将混好的文库上机 (Illumina Nova, PE150) 测序。

1.4 数据分析

使用软件 Stacks 过滤原始测序条目, 以去除接头序列和低质量碱基。去除接头序列, 去除 N(非 AGCT) 碱基大于或等于 5 的读数, 去除低质量读数 (质量值 $Q \leq 20$)。利用 BWA 软件将 clean reads 比对到白斑狗鱼基因组 Eluc_v4 版本 (PRJNA221548) 上, 根据比对结果统计样品的测序深度和参考基因组覆盖度。使用 GATK4 软件的 Haplotypecaller 模块进行 InDel 检测。为了降低 InDel 检测的错误率, 选用 $QD \geq 2.0$ 的标准进行过滤, 只保留同时满足该条件的突变位点。其中 QD 是突变质量值 (quality) 除以覆盖深度 (depth) 得到的比值, 实际上就是单位深度的突变质量值, 大部分假阳性突变的 QD 值都小于 2。使用 vcftools 进一步过滤: 保留 reads 支持深度不小于 4 的位点, 删除最小等位基因频率 (MAF) 小于 0.01 的位点, 保留个体检出率大于 80% 的位点。

使用 SnpEff 软件对得到的 InDel 进行注释, 以确定 InDel 在基因元件的位置、对氨基酸的变化影响等。过滤后的 InDel 数据利用 TASSEL5.2.80 软件进行亲缘关系分析, 利用 R 语言可视化。使用 GAPIT3 程序包进行关联分析。关联分析时白斑狗鱼耐高温性状 (UTT) 记录为二元表型数据, 分别为 0=耐高温组和 1=热敏感组。前 2 个 PCA 为协变量, 以 MLM 模型 (Masked Language Model) 进行 GWAS 分析, 以假阳性检出率 (false discovery rate, FDR) 校正的 $P < 0.05$ 为阈值, 筛选显著的 SNP 位点。IGV 软件可视化分析显著位点的具体位置, 统计 InDel 位点在每个个体中的分布情况 (表 1)。

1.5 KASP 验证候选 InDel 位点

与白斑狗鱼耐高温性状显著关联的 InDel 位点前后 100 bp 的序列, 利用 Primer premier 5.0 软件设计引物 (表 2)。每个位点设计两条正向引物和一条反向引物。两条正向引物 3' 为 InDel 位点, 5' 段分别加 FAM (5'-GAAGGTGACCAAGTTCATGCT-3') 和 HEX (5'-GAAGGTCGGAGTCAACGGATT-3') 荧光序列标签。引物由生工生物工程 (上海) 股份有限公司合成。PCR 反应体系按 LGC 公司 KASP TF V4.0 2×Master Mix 96 Std Rox 试剂盒说明配置, 反应程序包括 94 °C 预变性 15 min; 94 °C 变性 20 s, 63~57 °C 复性/延伸 60 s (-0.6/循环), 循环 10 次; 57 °C 复性/延伸 60 s, 循环

25 次; 37 °C 荧光数据读取 60 s。根据分型结果再次运行一下程序: 57 °C 复性/延伸 60 s, 循环 1~6 次; 37 °C 荧光数据读取 60 s。反应使用 CFX96 Real-Time System 完成。

2 结果

2.1 测序结果分析

热敏感组中第 1 尾实验鱼在 33 °C 存活 8 h 0 min 后死亡, 最后 1 尾实验鱼在 34 °C 存活 15 h 40 min 后死亡, 时间跨度为 31 h 40 min。耐高温组中的第 1 尾实验鱼 35 °C 存活 6 h 15 min 后死亡, 最后 1 尾实验鱼在 35 °C 存活 13 h 56 min 后死亡, 时间跨度尾 6 h 15 min。

本研究利用 Illumina 平台获得白斑狗鱼热敏感组 109 尾和耐高温组 103 尾个体的简化基因组数据。全部样本比对率范围为 91.32%~99.79%, 平均测序深度为 7.98×, 覆盖度范围为 10.52%~9.9%。经过突变位点检测并过滤, 共获得 12 821 个 InDel 位点, 其中插入 3 683 个, 缺失为 9 138 个。通过对 InDel 位点在基因组上的位置分析, 发现大部分 InDel 分布在内含子, 占有 InDel 位点的 63.69%, 基因间 InDel 位点占 9.80%, 外显子只有 1.30% (图 1)。获得的 InDel 平均密度为 36.67 InDel/Mb, NC_047579 号染色体密度最大 55.41 InDel/Mb, NC_047593 号染色体密度最小 22.59 InDel/Mb (图 2)。

表 1 MLM 模型 GWAS 分析与耐热性状显著关联的 InDel 位点

Tab. 1 InDel sites significantly associated with heat tolerance traits in GWAS analysis of the MLM models

插入缺失 InDel	染色体 chromosome	物理位置 location	InDel 序列 InDel sequence	基因 genes	P 值 P-values	MAF 值 MAF	FDR 值 FDR
2N del	NC_047592.1	22 216 786-22 216 787	CA	LOC117593903	3.47E-10	0.351 4	4.44E-06
9N del	NC_047571.1	27 885 959-27 885 967	ATGTTACCG	Clstn2	1.68E-08	0.434 0	5.62E-05
4N del-1	NC_047571.1	27 886 007-27 886 010	TAAG	Clstn2	1.68E-08	0.434 0	5.62E-05
4N del-2	NC_047571.1	27 886 256-27 886 259	TATC	Clstn2	1.76E-08	0.462 3	5.62E-05
3N ins	NC_047576.1	10 432 752	ACT	Safb	1.48E-05	0.441 0	0.037 9

注: FDR:FDR_Adjusted_P-values。
Note: FDR=FDR_Adjusted_P-values.

表 2 9N del 位点 KASP 引物序列

Tab. 2 KASP primer sequence of 9N del site

引物名称 primers name	荧光染料 fluorescent dyes	引物序列 primer sequences	位点 allele
CL IN2 F11	FAM	CCTCAGTGAGAGCACGGTAACA	In
CL IN2 F12	HEX	AGCCCCTCAGTGAGAGCATG	Del
CL IN2 R12		TGTATGATGCTGCAGGAACCTAAAC	

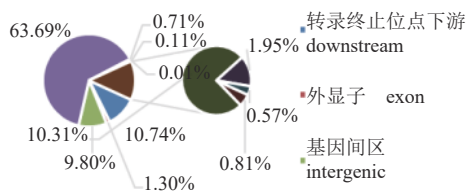


图 1 InDel 在基因组上的分布

Fig. 1 The distribution of InDel across the genome

2.2 全基因组关联分析

亲缘关系分析结果显示, 大多数个体的亲缘关系介于 0.1~0.4, 群体中亲缘关系值较均匀 (图 3-a), 表明群体结构对 GWAS 分析结果的影响小。QQ 图检验本研究分析的变异位点的有效性, 结果显示, 在横坐标呈现的均匀分布近似于基因组上的随机漂变, 当 P -value 小于 e^{-3} 时, 真实的 GWAS 结果与均匀分布出现分离 (图 3-b), 说明遗传变异与耐热性状相关性, 同时表明元分析的

GWAS 数据的可靠性和代表性。

曼哈顿图呈现了 GWAS 分析结果, 即 12 821 个 InDel 位点所在的染色体位置和 FDR 值大小, 虚线分隔为 FDR 校正后的阈值 $FDR < 0.05$, 实线分割线为 FDR 校正后的阈值 $FDR < 0.01$, 5 个 InDel 位点通过阈值检验, 表现出与耐热性状的显著相关性 (图 4)。

实验对 25 个染色体中的 InDel 标记, 以前两个 PCA 为协变量, 利用 MLM 模型与白斑狗鱼的耐热性状进行了关联分析。结果发现, 4 个位点与目标性状极显著关联 ($FDR P$ -value <0.01), 1 个位点为显著关联 ($FDR P$ -value <0.05)。这些位点位于未知基因 LOC117593903、*CLSTN2*、*safb* 基因的内含子中。其中, 9N del、4N del-1 和 4N del-2 三个位点位于 NC_047571.1 号染色体的 *CLSTN2* 基因第二个内含子中, 9N del 和 4N del-1 间有 40 个碱基, 4N del-1 和 4N del-2 间有 246 个碱基。

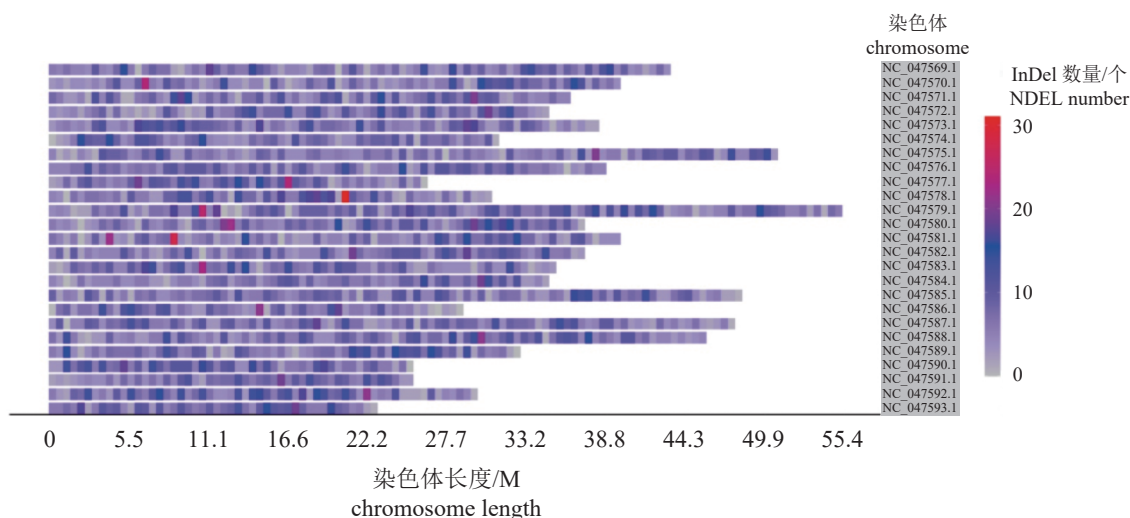


图 2 InDel 标记在 25 条染色体上的分布密度图

Fig. 2 The distribution density diagram of InDel markers on 25 chromosomes

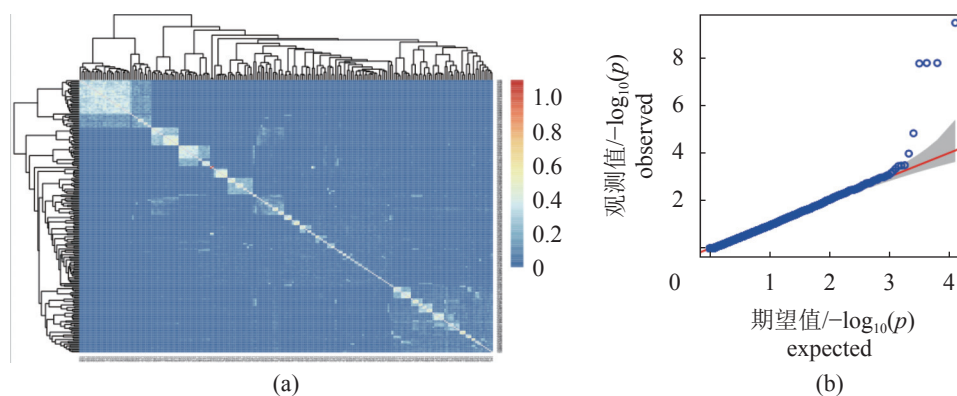


图 3 InDel 标记亲缘关系分析图 (a) 和 QQ 图 (b)

Fig. 3 Phylogenetic analysis diagram (a) and QQ diagram (b) of InDel markers

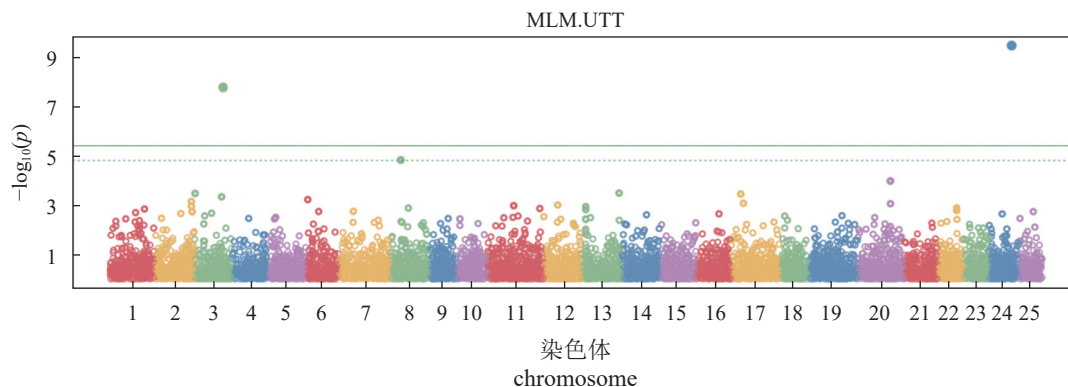


图4 InDel 标记的曼哈顿图

数字 1~25 依次代表 NC_047569.1 至 NC_047593.1 的染色体。

Fig. 4 Manhattan diagram of InDel markers

The numbers 1-25 represent chromosomes from NC_047569.1 to NC_047593-1.

9N del 和 4N del-1 位点在热敏感组中的 DI 基因型个体为 37 尾、DD 为 72 尾, 耐高温组中 DI 基因型个体为 83 尾、DD 为 20 尾。4N del-2 位点 DI 基因型个体为 34 尾、DD 为 75 尾, 耐高温组中 DI 基因型个体为 80 尾、DD 为 23 尾。2N del 位点在热敏感组中的 DD 基因型个体为 21 尾、DI 为 35 尾、II 为 53 尾, 耐高温组中 DD 基因型个体为 20 尾、DI 为 87 尾、II 为 1 尾。3N ins 位点在热敏感组中 DD 基因型个体为 89 尾、DI 为 20 尾, 耐高温组中 DD 基因型个体为 94 尾、DI 为 7 尾、II 为 2 尾。

2.3 KASP 验证

IGV 软件可视化分型为 20 尾, 4N del-2 位点 DI 基因型为 80 尾、DD 基因型为 23 尾。9N del 和 4N del-1 位点在 212 尾个体析显示, 热敏感组中 9N del 和 4N del-1 位点 DI 基因型个体为 37 尾、DD 基因型为 72 尾, 4N del-2 位点 DI 基因型为 34 尾、DD 基因型为 75 尾。耐高温组中 9N del 和 4N del-1 位点 DI 基因型个体为 83 尾、DD 基中的基因型分布无差异, 4N del-2 位点与前两者的差异仅为 2.83%, 212 尾个体中 3 个位点基因型分布高度连锁。因此我们针对 9N del、2N del 和 3N ins 位点设计引物, 以验证群体 DNA 为模板, 利用 KASP 技术进行了验证, 分型结果见图 5。结果显示, 2N del 和 3N ins 位点未通过验证, 9N del 位点在热敏感组 DD 基因型占 37.5%(9 尾), DI 基因型为 25%(6 尾), 未知占 37.5%(9 尾), 在耐高温组中 DD 基因型占 8.3%(2 尾), DI 基因型为 58.3%(14 尾), 未知占 33.3%(8 尾)。9N del 位点 DD 基因型个体在热敏感组中占优势, DI 基因

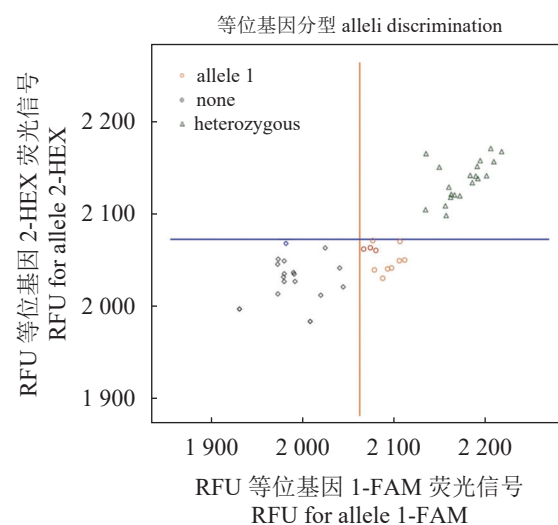


图5 KASP 基因分型图

橘色圆圈为缺失纯合 FAM 基因型 (DD), 绿色三角为杂合基因型 (DI), 黑色棱形为对照或未知基因型。

Fig. 5 Genotyping graph based on KASP

The orange circle represents the missing homozygous FAM genotype (DD), the green triangle represents the heterozygous genotype (DI), and the black prisms represent control or unknown genotypes.

型个体在耐高温组中占优势, 未发现插入纯合 (II), 验证结果与简化基因组测序结果基本一致。

3 讨论

本研究对白斑狗鱼 109 尾热敏感组和 103 尾耐高温组个体进行简化基因组测序, 经过滤后获得 12 821 个 InDel 位点。基因组上的位置分析发现, 大部分 InDel 分布在内含子 (63.69%), 外显子分布的 InDel 位点较少 (1.30%)。这一结果符合非编码区的 InDel 变异多于编码区的生物学特性^[12]。

GWAS 分析, 未发现与白斑狗鱼的耐高温性状显著关联的编码区 InDel 位点, 与耐热性状显著关联的 5 个 InDel 位点均分布于内含子。由于内含子不是编码区序列, 一度认为基因组中的“垃圾序列”。但近几年研究证实, 内含子与基因表达、细胞骨架构建和动态变化密切相关^[13-14]。在饥饿条件下酵母中大多数内含子可促进细胞的生存^[15], 内含子保留的 mRNA 不仅在细胞的正常发育中调控基因表达, 还在胁迫的应答中起着重要的作用^[16-17]。随着内含子生物学功能的不断证实, 内含子区域的变异也被重视。虽然鱼类 InDel 标记研究较少, 但在牛、绵羊、山羊、鸡等经济动物研究中发现了与经济性状关联的基因内含子区 InDel 位点, 并影响了所在基因的表达。Shi 等^[3]在发现中国黄牛品种的 *SMAD3* 基因内含子中检测到 1 个 17 bp 的插入片段与 *SMAD3* 转录水平显著相关; 大连雪龙牛 *SREBP1* 基因第 5 内含子 InDel 变异与背膘厚具有显著相关^[18]; Li 等^[6]在绵羊 *PRNP* 基因中发现了 4 个内含子多态性, 通过关联分析结果显示, 4 个内含子 InDel 均与绵羊 13 个不同生长性状显著相关; 陕西白绒山羊 *KDM6A* 基因内含子的一个 16 bp 插入/缺失显著影响 *KDM6A* 基因的表达, 并与陕西白绒山羊生长相关性状之间的关联^[4]; Wei 等^[19]在鸡 *LDB2* 基因第 2 内含子区域中发现了一个 31 bp InDel 突变位点, 通过关联分析发现该 InDel 与多个生长性状和胴体性状显著相关, 并影响 *LDB2* 基因在肌肉组织中的表达。由此推断, 本研究中发现的 5 个 InDel 突变可能会对白斑狗鱼的耐热性状产生显著的影响, 可作为白斑狗鱼耐热性状改良的候选分子标记。

对所发现的 5 个与白斑狗鱼耐热性状显著关联 InDel 位点注释和 IGV 可视化结果显示, 2N del 位点位于未知基因 LOC117593903 内含子中, 3N ins 位于 *sa7b* 基因内含子中。与目标性状极显著关联 (FDR < 0.01) 的 9N del、4N del-1 和 4N del-2 三个 InDel 位点均位于 *CLSTN2* 基因第 3 内含子, 并 9N del 和 4N del-1 间仅有 40 个碱基, 4N del-1 和 4N del-2 间有 246 个碱基。上述 3 个位点在 212 尾个体中的基因型分布高度连锁。在验证群体中 KASP 基因分型结果显示, 9N del 位点 DD 基因型个体在热敏感组中占优势, DI 基因型个体在耐高温组中占优势。有意思的是我们^[20]之前发现的白斑狗鱼耐热性状显著关联的 SNP 位点 HT1

位于 4N del-1 和 4N del-2 间。钙离子作为细胞内重要的第二信使, 参与包括盐胁迫、冷和热刺激等非生物胁迫应答, 并胁迫信号能改变细胞内钙离子水平^[21]。过高的体温具有直接的细胞毒性, 影响细胞膜的稳定性和跨膜转运蛋白的功能。因此, 跨膜离子转运被中断, 导致胞内钙离子的浓度增加。蛋白质和 DNA 的合成也在该途径中不同阶段被中断^[22]。*CLSTN2* (Calsyntenin-2, *CLSTN2*) 是钙结合蛋白家族 (Calsyntenins, *CLSTNs*) 成员之一, 是 I 型跨膜蛋白, 分布于脑、心脏、骨骼肌、肝脏、胰腺、胎盘和肺组织中^[23]。*CLSTNs* 对 Ca^{2+} 有高度亲和性的蛋白质, 其能够通过结合参与 Ca^{2+} 的运输、吸收与分泌^[24]。此外, 白斑狗鱼 *CLSTN2* 基因第 2 内含子为超长内含子 (34 365 bp)。王玲平^[25]认为, 超长的内含子通常采用外显子限定的剪接模式, 因此更容易发生可变剪接; 此外, 由于内含子越长, 伴随的基因转录时间也就越长, 转录时受到干扰而终止的可能性也就越高, 因此越长的内含子, 其处理的不可预测性就越大。由此推断, *CLSTN2* 可能是白斑狗鱼耐高温性状相关的重要候选基因。与白斑狗鱼耐热性状显著关联的 3 个 InDel 和 1 个 SNP 位点均位于 *CLSTN2* 第二内含子, 而 *CLSTN2* 第二内含子为超长内含子, 容易发生可变剪切和转录终止从而影响 *CLSTN2* 基因的转录。

(作者声明本文无实际或潜在的利益冲突)

参考文献 (References):

- [1] 海萨·艾也力汗, 贺疆滔, 张钰, 等. 白斑狗鱼早期发育阶段的耐热性及耐热性状关联 SNP 筛选 [J]. 中国水产科学, 2023, 30(6):677-684.
Haysa A, He J T, Zhang Y, et al. Heat resistance and screening of heat resistance-associated SNPs in the early development stage of *Esox lucius*[J]. Journal of Fishery Sciences of China, 2023, 30(6):677-684.
- [2] 齐遵利, 张秀文, 韩叙. 温度对白斑狗鱼早期发育和生长的影响[J]. 水产科学, 2010, 29(12): 708-710.
Qi Z L, Zhang X W, Han X. The effect of temperature on early development and growth in northern pike *Esox lucius*[J]. Fisheries Sciences, 2010, 29(12): 708-710 (in Chinese).
- [3] Shi T, Peng W W, Yan J Y, et al. A novel 17 bp indel in the *SMAD3* gene alters transcription level, contributing

- to phenotypic traits in Chinese cattle[J]. *Archives Animal Breeding*, 2016, 59(1): 151-157.
- [4] Wang K, Cui Y, Wang Z, *et al.* One 16 bp insertion/deletion (indel) within the *KDM6A* gene revealing strong associations with growth traits in goat[J]. *Gene*, 2019, 686: 16-20.
- [5] Wang R, Wang T, Lu W, *et al.* Three indel variants in chicken *LPIN1* exon 6/flanking region are associated with performance and carcass traits[J]. *British Poultry Science*, 2015, 56(6): 621-630.
- [6] Li J, Erdenee S, Zhang S L, *et al.* Genetic effects of *PRNP* gene insertion/deletion (indel) on phenotypic traits in sheep[J]. *Prion*, 2018, 12(1): 42-53.
- [7] 伞利择, 刘宝锁, 张楠, 等. 基于卵形鲳鲹基因组重测序的InDel标记挖掘与耐低氧性状关联分析[J]. *南方水产科学*, 2022, 18(5): 100-109.
- San L Z, Liu B S, Zhang N, *et al.* Mining of InDel marker and association analysis of hypoxia tolerance traits in *Trachinotus ovatus* based on resequencing[J]. *South China Fisheries Science*, 2022, 18(5): 100-109 (in Chinese).
- [8] 陈静. 基于全基因组重测序的鲟属鱼类群体遗传学研究 [D]. 武汉: 华中农业大学, 2021.
- Chen J. Population genetics of *Megalobrama* species based on the whole genome re-sequencing[D]. Wuhan: Huangzhong Agricultural University, 2021 (in Chinese).
- [9] 黄智康, 江世贵, 周发林, 等. 基于InDel标记的斑节对虾早期性别鉴定方法的建立[J]. *南方水产科学*, 2020, 16(3): 113-118.
- Huang Z K, Jiang S G, Zhou F L, *et al.* Establishment of a sex identification method for black tiger shrimp (*Penaeus monodon*) in embryonic and early postembryonic development based on an InDel molecular marker[J]. *South China Fisheries Science*, 2020, 16(3): 113-118 (in Chinese).
- [10] Semagn K, Babu R, Hearne S, *et al.* Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): Overview of the technology and its application in crop improvement[J]. *Molecular Breeding*, 2014, 33(1): 1-14.
- [11] Kaur B, Mavi G S, Gill M S, *et al.* Utilization of KASP technology for wheat improvement[J]. *Cereal Research Communications*, 2020, 48(4): 409-421.
- [12] 王慧芳, 周光现, 孙永峰, 等. 基于全基因组重测序技术的狮头鹅Indel标记分析[J]. *畜牧兽医学报*, 2021, 52(3): 662-675.
- Wang H F, Zhou G X, Sun Y F, *et al.* Analysis of Indel markers of Shitou goose based on whole genome resequencing technology[J]. *Acta Veterinaria et Zootechnica Sinica*, 2021, 52(3): 662-675 (in Chinese).
- [13] Jeong Y M, Mun J H, Lee I, *et al.* Distinct roles of the first introns on the expression of arabidopsis profilin gene family members[J]. *Plant Physiology*, 2006, 140(1): 196-209.
- [14] 唐馨, 苟萍. 内含子的功能[J]. *生命的化学*, 2019, 39(4): 772-777.
- Tang X, Gou P. The introns function[J]. *Chemistry of Life*, 2019, 39(4): 772-777 (in Chinese).
- [15] Parenteau J, Maignon L, Berthoumieux M, *et al.* Introns are mediators of cell response to starvation[J]. *Nature*, 2019, 565(7741): 612-617.
- [16] Jacob A G, Smith C W J. Intron retention as a component of regulated gene expression programs[J]. *Human Genetics*, 2017, 136(9): 1043-1057.
- [17] Naro C, Sette C. Timely-regulated intron retention as device to fine-tune protein expression[J]. *Cell Cycle*, 2017, 16(14): 1321-1322.
- [18] 徐丽. 牛 *SREBP1* 基因与牛肉肌内脂肪沉积的研究 [D]. 北京: 中国农业科学院, 2013.
- Xu L. Study on beef intramuscular fat deposition of *SREBP1* gene in cattle[D]. Beijing: Chinese Academy of Agricultural Sciences, 2013 (in Chinese).
- [19] Wei C J, Hou D, Feng Y, *et al.* Molecular characterization and a duplicated 31-bp indel within the *LDB2* gene and its associations with production performance in chickens[J]. *Gene*. 2020, 761: 145046.
- [20] Jiang Y C, Guo J M, Haisa A, *et al.* Genome-wide association analysis of heat tolerance in the northern pike (*Esox lucius*)[J]. *Aquaculture*, 2022, 559: 738459.
- [21] Ranty B, Aldon D, Galaud J P. Plant calmodulins and calmodulin-related proteins[J]. *Plant Signaling & Behavior*, 2006, 1(3): 96-104.
- [22] Walter E J, Hanna-Jumma S, Carraretto M, *et al.* The pathophysiological basis and consequences of fever[J]. *Critical Care*, 2016, 20(1): 200.
- [23] 任艳红, 赵福利, 李婧, 等. 铅对仔鼠学习记忆及海马组织中钙结合蛋白-2mRNA表达的影响[J]. *中国兽医学报*, 2013, 33(1): 98-101.

- Ren Y H, Zhao F L, Li J, *et al.* Effect of lead (Pb²⁺) on ability of learning and memory and expression of calsyn-tenin-2 in hippocampus of mice[J]. Chinese Journal of Veterinary Science, 2013, 33(1): 98-101 (in Chinese).
- [24] 马翔. 二花脸猪产仔数变异的生理特征挖掘和关键基因鉴别 [D]. 南京: 南京农业大学, 2019.
- Ma X. Detection of physiological features and identification of key candidate gene affecting variation of litter size of Erhualian pigs[D]. Nanjing: Nanjing Agricultural University, 2019 (in Chinese).
- [25] 王玲平. 不同生物中内含子大小和丰度的系统分析 [D]. 北京: 中国科学院北京基因组研究所, 2014.
- Wang L P. Systematic analysis of intron size and abundance parameters in diverse lineages[D]. Beijing: Beijing Institute of Genomics, Chinese Academy of Sciences, 2014 (in Chinese).

Screening of InDel markers associated with heat tolerance traits in *Esox lucius* based on simplified genome sequencing

AYELHAN Haisa¹, ZHANG Yu¹, YANG Bowen¹, XIAN Yulan¹, GAO Pan¹, SHEN Yubang^{2*}

(1. Xinjiang Fishery Research Institute, Urumqi 830000;

2. College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306)

Abstract: To provide effective molecular markers for the improvement of high temperature tolerance of *Esox lucius*, simplified genome sequencing was conducted based on the heat sensitive group (109) and high temperature tolerant group (103) of *E. lucius*. For InDel markers in 25 chromosomes, the first two PCA were covariables. The Masked Language Mode (MLM) was used to analyze the correlation between the heat tolerance characteristics of *E. Lucius*. The results showed that most InDel was distributed in introns (63.69%), while exons had few InDel sites (1.30%). Five loci were found to be significantly associated with heat tolerance traits (FDR *P*-value<0.05), which were located in the intron of *saftb* gene, unknown gene LOC117593903 and *CLSTN2* gene, respectively. The three InDel sites of 9N del, 4N del-1 and 4N del-2 were all located in the 3rd intron of *Clstn2* gene, and the genotypes of the three sites were highly linked in 212 individuals. The five InDel mutations found in this study may have a significant effect on the heat tolerance traits of *E. lucius*, and can be used as candidate molecular markers for the improvement of heat tolerance traits of *E. lucius*. Further, some loci were verified by KASP technique in validation population. It was found that the 9N del site DD genotype was dominant in the heat sensitive group, and the DI genotype was dominant in the high temperature tolerance group, which was basically consistent with the simplified genome sequencing results. Three InDel sites located in the second intron of *CLSTN2* gene may affect the transcription of *CLSTN2* gene. This gene may be an important candidate gene related to high temperature tolerance in *E. lucius*. The results provide theoretical basis for marker assisted breeding and candidate molecular markers for the improvement of *E. lucius* heat tolerance traits.

Key words: *Esox lucius*; heat resistance; InDel; Genome-wide association analysis

Corresponding author: SHEN Yubang. E-mail: ybshen@shou.edu.cn

Funding projects: National Nature Science Foundation of China (32060826)