

DOI: 10.12264/JFSC2023-0046

海湾扇贝南部亚种遗传多样性和选择特征的基因组学解析

林璐¹, 王浩², 曾启繁^{1,2}, 包振民^{1,2}

1. 中国海洋大学三亚海洋研究院, 海南省热带水产种质重点实验室, 海南 三亚 572000;
2. 中国海洋大学海洋生命学院, 海洋生物遗传学与育种教育部重点实验室, 山东 青岛 266003

摘要: 为了探究海湾扇贝南部亚种(*Argopecten irradians concentricus*)的遗传多样性特征, 通过全基因组重测序分析, 系统解析了其群体遗传结构以及其与海湾扇贝北部亚种(*Argopecten irradians irradians*)间的遗传差异, 共筛选出 15059961 个高质量单核苷酸多态性(single nucleotide polymorphisms, SNPs)位点。基于全基因组 SNP 的群体祖先成分推断、主成分分析和系统发生重构皆表明, 南北亚种之间存在较为显著的遗传分化。其中, 海湾扇贝南部亚种的近交水平和连锁不平衡程度更高。通过选择性清除分析, 在海湾扇贝南部亚种中共鉴定出 349 个正选择基因, 其中 *GHSR*、*HSF1*、*HABP2* 和 *DnaJ* 等基因可能与其生长速度较快、耐热性较强以及免疫功能较强等环境适应性特征相关。该研究揭示了海湾扇贝亚种间遗传分化和选择特征, 为扇贝的基因组研究和分子育种提供了数据支持。

关键词: 海湾扇贝南部亚种; 单核苷酸多态性; 遗传分化; 选择性清除

中图分类号: S917

文献标志码: A

文章编号: 1005-8737-(2023)06-0669-08

海产品富含蛋白质和微量营养素, 是全世界食品消费的重要组成部分^[1]。全球水产养殖产量在水产品总产量中的占比不断增加, 从 2000 年的 25.7% 上升到 2018 年的 46.0%, 已逐步成为食品行业中增长最快的产业^[2]。扇贝属于软体动物门(Mollusca), 双壳纲(Bivalvia), 翼形亚纲(Pteriomorpha), 为滤食性贝类, 广泛分布于世界各海域。作为重要的水产养殖物种, 扇贝闭壳肌味道鲜美且富含类胡萝卜素、不饱和脂肪酸等营养成分, 被誉为海产“八珍”之一, 是全球消费量最大的双壳类动物之一^[3]。此外, 扇贝壳主要成分为碳酸钙, 可作为优质钙源用作饲料添加, 研磨成粉后具有良好的吸附性, 可用于污水处理等工业, 具有较高的经济价值^[4-5]。

得益于人工繁育和养殖技术的逐步完善, 我

国扇贝养殖产业近年来发展迅速。目前, 国内水产养殖产业中主要的扇贝物种包括海湾扇贝(*Argopecten irradians*)、虾夷扇贝(*Mizuhopecten yessoensis*)、栉孔扇贝(*Azumapecten farreri*)等^[4,6-7]。其中, 海湾扇贝北部亚种和南部亚种分别于 1982 年和 1991 年由美国大西洋沿岸引入我国^[8-9]。由于海湾扇贝具有生长快、养殖 1 年就可以达到商品规格, 并且耐温范围广等特点, 引进后迅速在黄渤海区形成了国际上新兴的海湾扇贝养殖产业, 引领了海洋水产养殖业发展的第三次“浪潮”^[10]。海湾扇贝南北亚种在贝壳颜色、形态等外观特征上存在明显差异, 北部亚种除左壳(常为上壳)颜色与南部亚种相同都呈多态性外, 其右壳(常为下壳)颜色也呈多态性, 可能为白、黄或橙色, 而南部亚种右壳由于缺乏色素而呈白色^[11]。南北亚种对温度的

收稿日期: 2023-03-09; 修订日期: 2023-05-06.

基金项目: 国家重点研发计划项目(2021YFD1200805); 国家自然科学基金项目(32002388).

作者简介: 林璐(1999-), 女, 硕士研究生, 主要从事贝类遗传育种. E-mail: linlu19991214@163.com

通信作者: 王浩, 博士研究生, 研究方向为海洋生物基因组学, E-mail: wanghao5533@126.com; 曾启繁, 副教授, 研究方向为海洋生物基因组学与遗传育种, E-mail: zengqifan@ouc.edu.cn

适应性也不同,北部亚种扇贝对寒冷的北方水域更耐受,但不能在温暖的南方水域生存;南部亚种扇贝更适应温暖的南方水域,但在 12℃或更低的温度下会停止生长^[12]。因此,对南北亚种生长和适应性性状遗传差异的解析,对海湾扇贝种质资源的遗传改良具有潜在的重要意义。

本研究对海湾扇贝南北亚种进行了全基因组重测序分析,利用全基因组范围的单核苷酸多态性(single nucleotide polymorphism, SNP)数据揭示其群体结构和选择特征。研究结果扩充了扇贝基因组研究的数据资源,为扇贝种质资源调查和分子育种工作提供了重要参考。

1 材料与方法

1.1 样品采集及全基因组测序

本研究共采集 30 只海湾扇贝用于全基因组测序和遗传分析,包括 6 只海湾扇贝南部亚种(*Argopecten irradians concentricus*)(取自广东湛江)和 24 只海湾扇贝北部亚种(*Argopecten irradians irradians*)(取自山东青岛胶南和山东烟台莱州)。通过活体解剖收集其横纹肌组织,并用常规苯酚-氯仿萃取法提取了全基因组 DNA^[13]。使用 DNA 文库试剂盒(诺唯赞, ND617-02),利用酶切打断、磁珠分选的方法构建了插入片段大小为 250~350 bp 的 DNA 双端测序文库,并于 Illumina NovaSeq 平台完成 DNA 测序。之后利用 Trimmomatic 软件对原始下机数据进行质量控制,过滤低质量的 reads,参数为 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36^[14]。

1.2 SNP 标记开发和功能注释

本研究使用 Bowtie2 (版本 2.3.4.1)默认参数将高质量 reads 比对到海湾扇贝的参考基因组序列^[12,15]。利用 Samtools (版本 1.6)软件计算每个样本的基因组覆盖深度,并构建参考索引、格式转换和按基因组坐标排序得到储存测序数据比对的 BAM 文件^[16]。使用 Picard (<http://broadinstitute.github.io/picard/version1.92>)软件去除 BAM 文件中 PCR 重复数据。利用 Genome Analysis Toolkit (GATK)中的 HaplotypeCaller 模块获取每个样本的全基因组 SNP 信息^[17]。利用 GATK 中的 GenotypeGVCFs

模块对每个样本的 gVCF 文件进行联合基因分型。随后使用 GATK 的 VariantFiltration 和 SelectVariants 模块来获取高质量的双等位基因 SNPs。为了确保变异的位点和个体均一性,所有的 SNP 都经 Vcftools (版本 0.1.16)筛选,筛选标准为:次要等位基因频率(MAF)<0.05,位点的个体缺失比例(max-missing)>0.8,最小测序深度(minDP)<4^[18]。利用 SnpEff 软件注释了每个 SNP 所在的基因组元件,确定了发生突变的类型和突变的影响^[19]。

1.3 遗传多态性和连锁规律解析

本研究基于全基因组 SNPs,利用 CMplot 绘制 SNP 分布密度图;采用 ADMIXTURE 软件进行群体结构分析^[20];借助 IQ-TREE 软件构建了最大似然法(maximum likelihood, ML)系统发生树^[21];采用 PLINK 软件对所有测序样本进行主成分分析(principal component analysis, PCA)^[22];采用 Vcftools 计算两个亚种的核苷酸多态性($\theta\pi$)、群体遗传分化指数(F_{ST})^[18];采用 Stacks 对两个亚种的杂合度进行统计分析^[23];利用 PopLDdecay 对两个亚种进行了连锁不平衡(linkage disequilibrium, LD)衰减分析^[24];借助 PLINK 软件鉴定了每个亚种的连续性纯合片段(Runs of homozygosity, ROHs)的总数(ROHs 最小长度设定为 100 kb)^[22]。

1.4 选择性清除分析

基因组中的正选择区域具有较高遗传分化和低核苷酸多态性的特征,本研究采用全基因组中前 5%极值的 $\theta\pi$ 比值和前 1%极值的 F_{ST} 作为显著受选择区域^[25-27]。将正选择区域内的基因定义为受选择的候选基因。利用 TBtools 进行 Gene Ontology (GO)富集分析^[28]。利用费舍尔精确检验确定候选基因所富集的 GO 条目,显著性水平 $\alpha=0.05$ 。

2 结果与分析

2.1 海湾扇贝 SNP 标记开发和功能注释

将本研究来自 2 个亚种的 30 只海湾扇贝的全基因组重测序文库比对到参考基因组,每个文库基因组的平均覆盖深度为 21×,其中大约 90%读长为唯一比对。经过 SNP 质量控制后,在两个亚种中共鉴定出 15059961 个高质量 SNP,其中

1616072 个 SNP 为南部亚种特有, 3319537 个 SNP 为北部亚种特有。经过注释分析, 15 M 的 SNP 位点中有 23364701 个位点在功能基因区产生影响。仅有 2.82% (656477 个) 的位点分布于外显子区, 而位于基因间区和内含子区的 SNPs 超过 36.0% (8384921 个) 和 32.3% (7522832 个)(表 1)。位于外显子内的 SNPs 中, 沉默突变占 55.1% (358166 个), 错义突变占 43.7% (284415 个), 由于产生终止密码子而引起的无义突变占 1.20% (7777 个)。所有 SNP 中包含了 97243196 个转换和 82018037 个颠换, 转换/颠换比为 1.19 (表 2)。

表 1 SNP 分布区域、影响程度和功能类别
Tab. 1 SNP annotation by genomic region, impact and function class

类别 category	类型(按字母顺序排列) type (alphabetical order)	数量 count	占比/% percentage
分布 区域 region	基因下游 downstream	3142002	13.5
	外显子 exon	656477	2.82
	基因 gene	10	0
	基因间区 intergenic	8384921	36.0
	内含子 intron	7522832	32.3
	剪接位点受体 splice site acceptor	1374	0.006
	剪接位点供体 splice site donor	1346	0.006
	剪接位点区域 splice site region	67256	0.289
	转录本 transcript	280	0.001
	基因上游 upstream	2968984	12.7
影响 程度 impact	3'端非翻译区 3' UTR prime	465302	1.99
	5'端非翻译区 5' UTR prime	76660	0.329
	高 high	23166	0.100
	低 low	427974	1.84
	中等 moderate	284767	1.22
功能 类别 function class	几乎无影响 modifier	22551537	96.8
	错义突变 missense	284415	43.7
	无义突变 nonsense	7777	1.20
	沉默突变 silent	358166	55.1

2.2 遗传多态性和连锁规律解析

如图 1 所示, SNP 在基因组中呈现不均匀分布, 红色部分为 SNP 热点区域(图 1a)。图 1b 分别展示了 $K=2, 3, 4$ 时的群体遗传结构, 其中南部亚种所有个体具有相同的遗传组分, 而北部亚种存在

表 2 核苷酸替换类型统计
Tab. 2 Summary of nucleotide substitutions

	A	C	G	T
A	—	725199	1859355	1459506
C	947290	—	344207	2190268
G	2193514	344160	—	945916
T	1468676	1855446	726424	—

一定的杂交特征。基于最大似然法的系统发生分析发现, 30 个样本明显分成了两簇(支持度为 100)(图 1c), 与其样本来源相对应。主成分分析(PCA)结果与系统发生分析相一致, 在 PC1 方向上显著分离(PC1 解释的变异度为 14.7%, PC2 解释的变异度为 4.30%)(图 1d)。可见海湾扇贝南北亚种基因组存在显著的遗传差异。

利用相邻 SNPs 的基因型频率相关性系数 r^2 , 本研究进一步分析了两个亚种的连锁不平衡(LD)衰减规律(图 2a)。在距离小于 300 kb 时, 南部亚种的连锁不平衡程度明显高于北部亚种。南部亚种的 r^2 在 300 kb 衰减至 0.197, 北部亚种的 r^2 衰减至 0.055。全基因组连续纯合片段(ROH)分析表明南部亚种比北部亚种保留更多 ROH 片段, 近交水平更高(图 2b)。

本研究对两个亚种的杂合度统计分析结果显示(表 3), 南部亚种的亚种内 F 统计值(F_{is})为 0.037, 北部亚种为 0.085。南部亚种平均杂合度为 0.230, 北部亚种为 0.256。

2.3 海湾扇贝亚种间遗传分化和选择特征

海湾扇贝基因组上的大部分区域没有表现出亚种间差异, 但在特定区域呈现出显著的遗传分化和多态性差异(图 3a)。通过选择清除分析, 本研究从南部亚种基因组的 79003 个区域中共鉴定到 349 个选择候选基因。筛选出了在生长、抗逆性和免疫中起重要作用的基因, 包括与生长相关的 *GHSR*、与耐热性相关的 *HSF1* 以及与免疫相关的 *HABP2* 和 *DnaJ* 等。功能富集分析显示, 这些位于强选择清除区域内的基因主要与细胞信号传导、维持细胞形态以及调节代谢过程等有关(图 3b)。

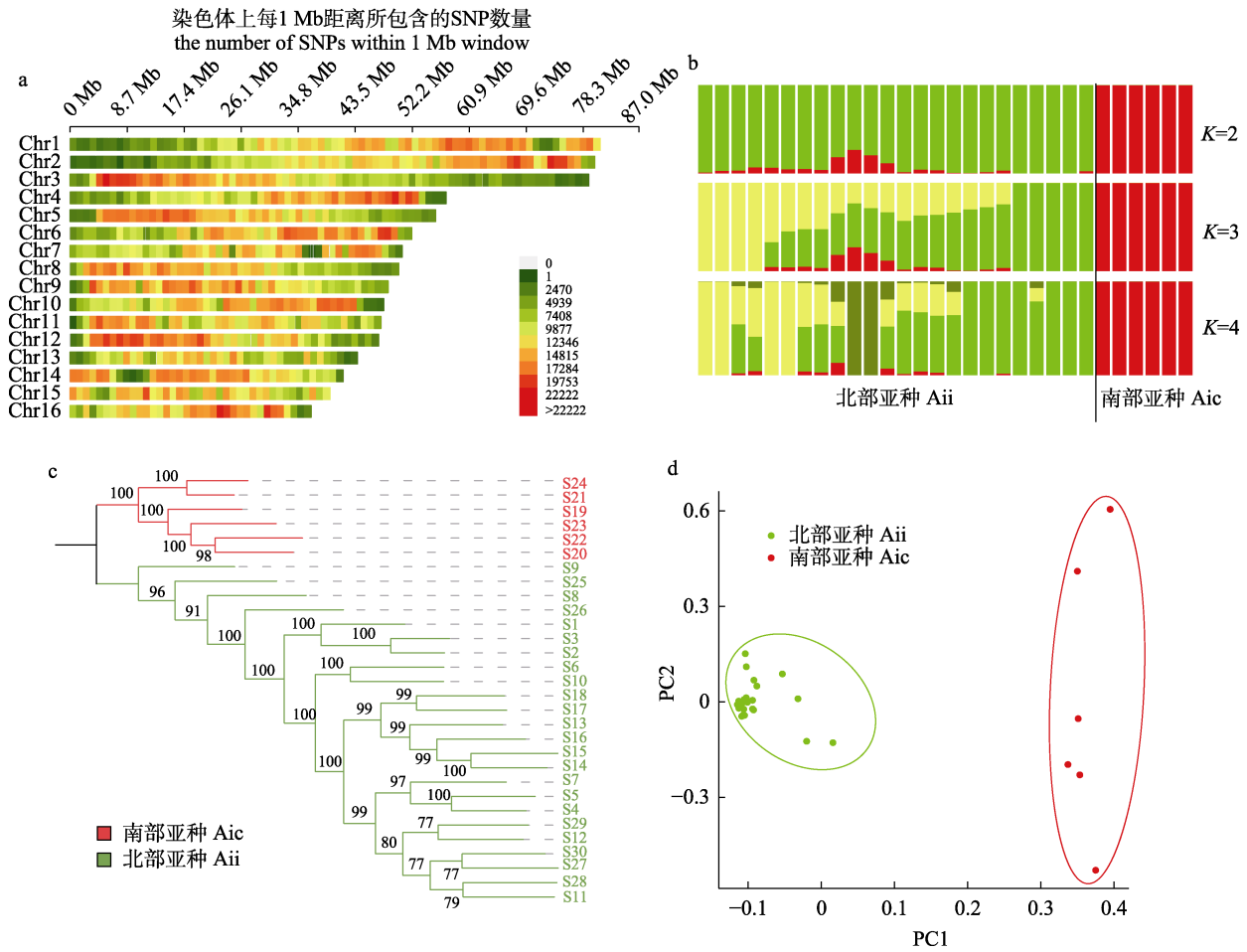


图 1 海湾扇贝群体结构和遗传分化分析

a. 基因组内 SNP 分布; b. 基于祖先成分的遗传结构推断; c. 基于全基因组 SNPs 的最大似然系统发育树; d. 基于全基因组 SNPs 的主成分分析.

Fig. 1 Population structure and genetic differentiation analysis of *Argopecten irradians*

a. Genome-wide SNP distribution; b. Genetic structure inferred by ancestral components; c. Maximum likelihood phylogenetic tree based on whole genomic SNPs. d. Principal component analysis based on genome-wide SNPs.

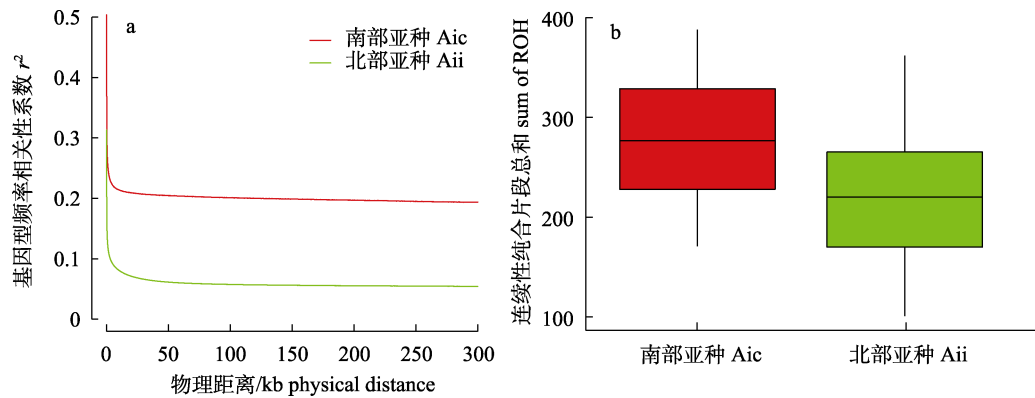


图 2 连锁不平衡衰减程度和近交水平

a. 0–300 kb 长度内的连锁不平衡衰减; b. 在每个亚种中检测到的连续性纯合片段总和的箱式图.

Fig. 2 Selection signatures of southern subspecies of *Argopecten irradians*.

a. Linkage disequilibrium decay at distance of 0–300 kb. b. Boxplot of sum of runs of homozygosity detected in each population.

表 3 两个海湾扇贝亚种的杂合度统计

Tab. 3 Heterozygosity statistics for two *Argopecten irradians* subspecies

亚种 subspecies	样本数 sample size	平均杂合度 expected heterozygosity	亚种内 F 统计值 F_{is}
南部亚种 <i>Argopecten irradians concentricus</i>	6	0.230	0.037
北部亚种 <i>Argopecten irradians irradians</i>	24	0.256	0.085

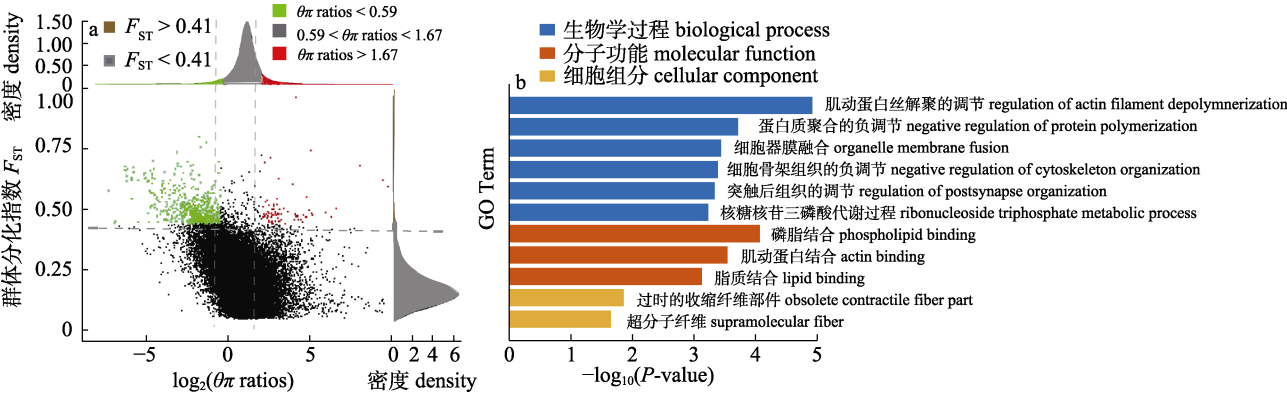


图 3 海湾扇贝南部亚种的选择性清除和正选择基因的功能富集分析

a. $\theta\pi$ 比值(北部亚种/南部亚种)和 F_{ST} 值的分布; 红色区域(对应于 $\theta\pi$ 比值经验分布的 5% 右尾部和 F_{ST} 经验分布的 1% 右尾部)为海湾扇贝南部亚种的选择性清除区域; b. 南部亚种中正选择基因的功能富集。

Fig. 3 Selective sweep and GO enrichment analysis of *Argopecten irradians concentricus*

a. Distribution of $\theta\pi$ ratios (*Argopecten irradians irradians*/*Argopecten irradians concentricus*) and F_{ST} values. Data points in red (corresponding to the 5% right tails of the empirical $\theta\pi$ distribution, and the 1% right tail of the empirical F_{ST} distribution) represented regions under selective sweep in *Argopecten irradians concentricus*. b. Functional enrichment of genes under positive selection in *Argopecten irradians concentricus*.

3 讨论

本研究利用全基因组重测序的方法, 获取了海湾扇贝全基因组尺度的 SNP 标记集合并对其进行注释, 探究了南北亚种的群体分化和全基因组尺度的选择特征, 对海湾扇贝种质资源的遗传改良具有潜在的重要意义^[29]。在本研究中, 群体结构、PCA 以及进化树分析结果都一致证明海湾扇贝南北亚种已经出现了明显的遗传分化。本研究对当前中国养殖的海湾扇贝南北亚种进行连锁不平衡分析, 结果显示, 南部亚种保留了更高水平的 LD, 与 Yang 等^[30]发现南北亚种在耐热性状相关基因 *HSP90* 启动子区的连锁不平衡状态存在显著差异的结果相一致。

群体遗传多样性分析显示, 在基因组特定区域, 南北亚种之间存在明显分化, 且南部亚种遗传多样性有所降低。Wang 等^[10]利用微卫星标记计算杂合度和 F_{is} 指数后发现, 相比于美国大西洋沿岸自然群体(杂合度: 0.447; F_{is} 指数: 0.117), 引

种后的中国养殖群体(杂合度: 0.364; F_{is} 指数: 0.182)发生等位基因多态性丢失。本研究基于全基因组 SNP 数据进行基因组杂合度分析后发现, 与 Wang 等^[10]2007 年前后收集的美洲自然群体以及国内南部亚种养殖群体相比, 现有国内南部亚种养殖群体表现出最低的杂合度 0.230 和 F_{is} 指数 0.037。同时, 本研究南北亚种遗传分化指数(F_{ST})为 0.08, 相比于 2007 年前后采集的南北亚种遗传分化指数(0.06), 发生进一步遗传分化^[10]。因此本研究认为我国南部亚种现有养殖群体遗传多样性丢失可能与海湾扇贝的耐热性、生长速度等持续人工选育相关^[31-32]。陈静等^[33]对海湾扇贝南部亚种进行连续 5 代闭锁选育后与对照组普通群体的遗传分化指数为 0.046, 选育群体显示出明显的生长及高成活率优势的同时, 遗传型也出现了一定的分化, 并且选育群体的平均杂合度有所下降。在动物育种中, 为了子代育种值的提高, 亲缘关系较近的个体可能会被留种, 从而导致近交的产生, 进而影响动物的生长、存活和繁殖能力

等性状。因此, 监测和管理种群近交水平在遗传改良中至关重要, 在未来的扇贝育种中, 应该注重保护种质资源, 避免对扇贝品种的过度近交, 以维持其遗传多样性^[34-35]。此外, 本研究缺少美洲自然群体基因数据、海湾扇贝引种产生奠基者效应(founder effect)以及本研究中所包含的南部亚种有限的样本数量也是连锁不平衡程度较高和遗传多样性偏低的因素^[8-9,36-37]。

本研究在海湾扇贝南部亚种中鉴定到 349 个正选择基因, 其中 *GHSR*、*HSF1*、*HABP2* 和 *DnaJ* 基因与生长、耐热以及免疫相关。有研究表明 *GHSR* 基因是调节生长激素的重要途径, 可以促进生长发育, 有利于马氏珠母贝(*Pinctada fucata martensii*)形成生长杂种优势^[38]。Liu 等^[39]研究表明, 在热胁迫下太平洋牡蛎(*Crassostrea gigas*)的 *HSP70* 基因可能受到 *HSF1* 的调节。因此可推测, 海湾扇贝南部亚种的耐热性也与 *HSF1* 和热休克蛋白之间的调控有关。Yoshida 等^[40]在红鲷(*Lutjanus erythropterus*)中确定了 *HABP2* 的核苷酸序列, 该基因编码循环丝氨酸蛋白酶, 在调节凝血和纤溶过程中发挥重要作用。有研究结果表明 *DnaJ* 可能参与介导菲律宾蛤蜊(*Venerupis philippinarum*)的免疫反应和环境胁迫^[41]。基因组选择育种与基于系谱和其他分子标记辅助选择的方法相比, 在选择准确性方面有实质提升, 已成为动物育种的热点技术, 但基因组选择的成本仍然较高, 全基因组遗传标记数据集的生成可推动全基因组选择的应用^[42]。

4 结论

本研究对两个海湾扇贝亚种进行了全基因组重测序, 获得了 15059961 个高质量的单核苷酸多态性位点。本研究中群体结构分析、主成分分析以及系统发生分析结果均显示海湾扇贝南北亚种基因组存在显著的遗传差异。南部亚种的连锁不平衡和近交程度均高于北部亚种。选择性清除分析显示, 南部亚种的 349 个基因处于选择清除区域内, 这些基因可能与生长、耐热及免疫等性状和环境适应特征相关。本研究获得的高密度单核苷酸多态性是扇贝基因组资源的重要补充, 将有助于扇贝育种计划的实际应用。

参考文献:

- [1] Béné C, Barange M, Subasinghe R, et al. Feeding 9 billion by 2050—Putting fish back on the menu[J]. Food Security, 2015, 7(2): 261-274.
- [2] Fiorella K J, Okronipa H, Baker K, et al. Contemporary aquaculture: Implications for human nutrition[J]. Current Opinion in Biotechnology, 2021, 70: 83-90.
- [3] Song L S, Wang L L, Zhang H, et al. The immune system and its modulation mechanism in scallop[J]. Fish & Shellfish Immunology, 2015, 46(1): 65-78.
- [4] Li C L, Song A H, Hu W, et al. Current situation analysis and development countermeasures of scallop culture industry in Shandong Province[J]. Marine Sciences, 2011, 35(3): 92-98. [李成林, 宋爱环, 胡炜, 等. 山东省扇贝养殖产业现状分析与发展对策[J]. 海洋科学, 2011, 35(3): 92-98.]
- [5] Dong X, Kim Y S, Kim E K, et al. Scallop extracts inhibited LPS-induced inflammation by suppressing MAPK and NF-κB activation in RAW264.7 macrophages[C]//Hu J, Piao F, Schaffer S, et al. Taurine 11. Singapore: Springer, 2019: 1069-1081.
- [6] Li C Y, Ding J, Chang Y Q, et al. Isolation of microsatellite markers in Japanese scallop and its application in genetic structure analysis[J]. Journal of Fishery Sciences of China, 2009, 16(1): 39-46. [李春艳, 丁君, 常亚青, 等. 虾夷扇贝微卫星标记的分离及其养殖群体的遗传结构分析[J]. 中国水产科学, 2009, 16(1): 39-46.]
- [7] Guo X M, Luo Y S. Scallops and scallop aquaculture in China[M]. Scallops-Biology, Ecology, Aquaculture, and Fisheries. Amsterdam: Elsevier, 2016: 937-952.
- [8] Zhang F S, He Y C, Liu X S, et al. A report on the introduction, spat-rearing and experimental culture of bay scallop, *Argopecten irradians* Lamarck[J]. Oceanologia Et Limnologia Sinica, 1986, 17(5): 367-374. [张福绥, 何义朝, 刘祥生, 等. 海湾扇贝(*Argopecten irradians*)引种、育苗及试养[J]. 海洋与湖沼, 1986, 17(5): 367-374.]
- [9] Zhang F S, He Y C, Qi L X, et al. Introduction and spat-rearing of F1 generation of *Argopecten irradians* concentricus Say[张福绥, 何义朝, 齐铃欣, 等. 墨西哥湾扇贝的引种和子一代苗种培育[J]. 海洋与湖沼, 1994, 25(4): 372-377.]
- [10] Wang L L, Zhang H, Song L S, et al. Loss of allele diversity in introduced populations of the hermaphroditic bay scallop *Argopecten irradians*[J]. Aquaculture, 2007, 271(1-4): 252-259.
- [11] Zheng H P, Zhang G F, Liu X, et al. Establishment of different shell color lines of bay scallop *Argopecten irradians* (Lamarck 1819) and their development[J]. Oceanologia Et Limnologia Sinica, 34(6): 632-639. [郑怀平, 张国范, 刘晓, 等. 不同贝壳颜色海湾扇贝(*Argopecten irradians*)家系的建立及生长发育研究[J]. 海洋与湖沼, 2003, 34(6): 632-639.]
- [12] Liu X, Li C, Chen M, et al. Draft genomes of two Atlantic

- Bay scallop subspecies *Argopecten irradians irradians* and *A. i. concentricus*[J]. Scientific Data, 2020, 7: 99.
- [13] Sambrook J, Russell D W. Molecular cloning: a laboratory manual[M]. 3rd ed. New York: Cold Spring Harbor Laboratory Press, 2001
- [14] Bolger A M, Lohse M, Usadel B. Trimmomatic: A flexible trimmer for Illumina sequence data[J]. Bioinformatics, 2014, 30(15): 2114-2120.
- [15] Langmead B, Salzberg S L. Fast gapped-read alignment with bowtie 2[J]. Nature Methods, 2012, 9(4): 357-359.
- [16] Li H, Handsaker B, Wysoker A, et al. The sequence alignment/map format and SAMtools[J]. Bioinformatics, 2009, 25(16): 2078-2079.
- [17] 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.
- [18] Danecek P, Auton A, Abecasis G, et al. The variant call format and VCFtools[J]. Bioinformatics, 2011, 27(15): 2156-2158.
- [19] Cingolani P, Platts A, Wang L L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3[J]. Fly, 2012, 6(2): 80-92.
- [20] Alexander D H, Novembre J, Lange K. Fast model-based estimation of ancestry in unrelated individuals[J]. Genome Research, 2009, 19(9): 1655-1664.
- [21] 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.
- [22] 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.
- [23] Catchen J, Hohenlohe P A, Bassham S, et al. Stacks: An analysis tool set for population genomics[J]. Molecular Ecology, 2013, 22(11): 3124-3140.
- [24] Zhang C, Dong S S, Xu J Y, et al. PopLDdecay: A fast and effective tool for linkage disequilibrium decay analysis based on variant call format files[J]. Bioinformatics, 2019, 35(10): 1786-1788.
- [25] Lei Z H, Sun W B, Guo T T, et al. Genome-wide selective signatures reveal candidate genes associated with hair follicle development and wool shedding in sheep[J]. Genes, 2021, 12(12): 1924.
- [26] Li Y D, Liu X, Li Z W, et al. A combination of genome-wide association study and selection signature analysis dissects the genetic architecture underlying bone traits in chickens[J]. Animal, 2021, 15(8): 100322.
- [27] Wang H, Lv J, Zeng Q F, et al. Genetic differentiation and selection signatures in two bay scallop (*Argopecten irradians*) breeds revealed by whole-genome resequencing analysis[J]. Aquaculture, 2021, 543: 736944.
- [28] Huang D W, Sherman B T, Lempicki R A. Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists[J]. Nucleic Acids Research, 2009, 37(1): 1-13.
- [29] Hosomi R, Yoshida M, Fukunaga K. Seafood consumption and components for health[J]. Global Journal of Health Science, 2012, 4(3): 72-86.
- [30] Yang C Y, Wang L L, Liu C H, et al. The polymorphisms in the promoter of HSP90 gene and their association with heat tolerance of bay scallop[J]. Cell Stress and Chaperones, 2015, 20(2): 297-308.
- [31] Xiao J, Ford S E, Yang H S, et al. Studies on mass summer mortality of cultured Zhikong scallops (*Chlamys farreri* Jones et Preston) in China[J]. Aquaculture, 2005, 250(3-4): 602-615.
- [32] Zhu X H, Liu P P, Hou X J, et al. Genome-wide association study reveals PC4 as the candidate gene for thermal tolerance in Bay scallop (*Argopecten irradians irradians*)[J]. Frontiers in Genetics, 2021, 12: 650045.
- [33] Chen J, Liu Z G, Wu S R, et al. Growth and microsatellite analysis of f5 selected line of *Argopecten irradians concentricus*[J]. Oceanologia et Limnologia Sinica, 2012, 43(4): 783-788. [陈静, 刘志刚, 邬思荣, 等. 海湾扇贝南部亚种(*Argopecten irradians concentricus*)选育系 F5 的生长及 SSR 分析[J]. 海洋与湖沼, 2012, 43(4): 783-788.]
- [34] Slatkin M. Linkage disequilibrium—Understanding the evolutionary past and mapping the medical future[J]. Nature Reviews Genetics, 2008, 9(6): 477-485.
- [35] Meuwissen T H E, Sonesson A K. Maximizing the response of selection with a predefined rate of inbreeding: Overlapping generations[J]. Journal of Animal Science, 1998, 76(10): 2575-2583.
- [36] Qin Y J, Liu X, Zhang H B, et al. Effect of parental stock size on F1 genetic structure in the bay scallop, *Argopecten irradians* (Lamarck, 1819)[J]. Aquaculture Research, 2007: 070116074017003.
- [37] Watterson G A. Allele frequencies after a bottleneck[J]. Theoretical Population Biology, 1984, 26(3): 387-407.
- [38] Yang J M, Luo S J, Li J H, et al. Transcriptome analysis of growth heterosis in pearl oyster *Pinctada fucata martensii*[J]. FEBS Open Bio, 2018, 8(11): 1794-1803.
- [39] Liu Y L, Li L, Huang B Y, et al. RNAi based transcriptome suggests genes potentially regulated by HSF1 in the Pacific oyster *Crassostrea gigas* under thermal stress[J]. BMC Genomics, 2019, 20(1): 1-15.
- [40] Yoshida A, Wang Y J, Bae I, et al. Molecular cloning and tissue distribution of hyaluronan binding protein 2 (*HABP2*)

- in red sea bream *Pagrus major*[J]. Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology, 2013, 165(4): 271-276.
- [41] Li C H, Li L Y, Liu F, et al. Alternation of *Venerupis philippinarum* *Hsp40* gene expression in response to pathogen challenge and heavy metal exposure[J]. Fish & Shellfish Immunology, 2011, 30(1): 447-450.
- [42] Houston R D. Future directions in breeding for disease resistance in aquaculture species[J]. Revista Brasileira De Zootecnia, 2017, 46(6): 545-551.

Genomic analysis of the genetic diversity and signatures of selection in a bay scallop southern subspecies *Argopecten irradians concentricus*

LIN Lu¹, WANG Hao², ZENG Qifan^{1,2}, BAO Zhenmin^{1,2}

1. Key Laboratory of Tropical Aquatic Germplasm of Hainan Province; Sanya Ocean Institute, Ocean University of China, Sanya 572000, China;
2. Key Laboratory of Marine Genetics and Breeding, Ministry of Education; College of Marine Life Sciences, Ocean University of China, Qingdao 266003, China

Abstract: Bay scallop (*Argopecten irradians*) is an economically important bivalve species in China due to its fast growth rates and suitability for artificial breeding. There are five bay scallop subspecies distributed along the Atlantic coast of North America and the Gulf of Mexico that exhibit distinct growth and thermal stress tolerance characteristics, making them valuable resources for adaptive evolutionary studies. In this study, we used whole-genome resequencing analysis to uncover the genetic diversity and selective signatures of the bay scallop southern subspecies *A. i. concentricus* via comparison with the northern subspecies *A. i. irradians*. We identified a total of 15059961 high-quality single nucleotide polymorphisms (SNPs), of which 1616072 were unique to *A. i. concentricus* and 3319537 were unique to *A. i. irradians*. Following annotation analysis, a total of 23364701 effects were identified from the 15 million SNP sites. Only 2.819% of the loci were located in the exon regions; the majority of these SNPs were instead located in the intergenic (36.0%) and intron (32.3%) regions. Among the SNPs located within exons, silent mutations accounted for 55.1% (358166), missense mutations accounted for 43.7% (284415), and nonsense mutations accounted for 1.2% (7777). The total numbers of transitions and transversions across all SNPs were 97243196 and 82018037, respectively, resulting in a transitions/transversions ratio of 1.1856. Population structure, principal component analysis (PCA), and phylogenetic analysis based on genome-wide SNPs indicated significant genetic differentiation between *A. i. concentricus* and *A. i. irradians*. Linkage disequilibrium (LD) analysis showed that *A. i. concentricus* retained a higher level of LD compared to that of *A. i. irradians*, which may have been potentially influenced by domestication selection and inbreeding. Runs of homozygosity (ROH) analysis across the whole genome revealed that *A. i. concentricus* retained more ROH, indicating a higher level of inbreeding. Population genetic diversity analysis revealed a decrease in genetic diversity in certain regions of the *A. i. concentricus* genome compared to those of the northern subspecies population; this may be the result of continuous artificial selection for heat tolerance and inbreeding. Selective sweep analysis of *A. i. irradians* revealed 349 genes under positive selection, including *GHSR*, *HSFI*, *HABP2*, and *DnaJ*, which may be related to its corresponding rapid growth rate, strong heat tolerance, high immunity, and other environmental adaptation characteristics. Functional enrichment analysis revealed that the genes located in the selective sweep regions were primarily associated with cell signal transduction, maintenance of cell morphology, and regulation of metabolic processes. Overall, this study sheds light on the genetic differentiation and signatures of selection in bay scallop subspecies, thereby providing a valuable resource for scallop genomic research and molecular breeding.

Key words: *Argopecten irradians concentricus*; SNP; genetic differentiation; selective sweep

Corresponding author: WANG Hao, E-mail: wanghao5533@126.com; ZENG Qifan, E-mail: zengqifan@ouc.edu.cn