

基于高效混合模型的全基因组关联分析方法对 鲢鱼头尺性状的遗传解析

陈 澳^{1,2,3}, 常中宇^{1,2,3}, 赵兴宁⁴, 李 宁⁵, 蒋 丽^{2,3*}

1. 上海海洋大学, 水产科学国家级实验教学示范中心, 上海 201306
2. 中国水产科学研究院, 生物技术研究中心, 北京 100141
3. 中国水产科学研究院, 农业农村部水生动物基因组学重点实验室, 北京市渔业生物技术重点实验室, 北京 100141
4. 青铜峡市天源渔业专业合作社, 宁夏 吴忠 751600
5. 四川渔鲲水产科技有限公司, 四川 达州 635000

摘要: 头骨形态是生物进化和物种适应环境的基础, 对于水产养殖鱼类来说, 鱼头大小尺寸也是重要的经济性状, 它与鱼片产量和观赏性有关。因此, 挖掘控制鲢鱼头部大小的基因位点, 可为遗传选育提供分子标记及可利用的功能基因。本研究定位了与鲢鱼头部大小相关的数量性状核苷酸(QTN)和基因, 利用鲢鱼10个家系的杂交群体进行基因分型, 使用高效、稳健的基于混合模型的Optim-GRAMMAR法进行GWAS分析, 并将结果与EMMAX进行了比较。在得到206 763个高质量SNP的基础上, 同种条件下EMMAX无显著关联的SNP, 而Optim-GRAMMAR在头长、头深、头宽方面则分别发现了1、6和1个SNP, 这些基因大多与生长发育、细胞增殖、促进代谢、骨骼形态、摄食调节有关, 其中*slc12a5a*, *slc7a10a*, *Rab2a*等基因在斑马鱼等物种中对生长发育有正向影响。本研究报告的QTNs及基因将有助于了解鱼头大小性状的遗传结构, 并促进水产养殖品种鲢鱼的标记辅助选择育种。

关键词: 全基因组关联分析; Optim-GRAMMAR; EMMAX; 鲢鱼; 头部大小

中图分类号: S917.4

文献标识码: A

文章编号: 1000-2324(2025)06-0929-09

Genetic Analysis of Catfish Head-size Traits Using a Genome-Wide Association Analysis Method Based on an Efficient Mixed Model

CHEN Ao^{1,2,3}, CHANG Zhong-yu^{1,2,3}, ZHAO Xing-ning⁴, LI Ning⁵, JIANG Li^{2,3*}

1. College of Fisheries and Life Science, Shanghai Ocean University, Shanghai 201306, China
2. Research Centre for Aquatic biotechnology, Chinese Academy of Fishery Sciences, Beijing 100141, China
3. Beijing Key Laboratory of Fishery Biotechnology, Chinese Academy of Fishery Sciences, Beijing 100141, China
4. Tianyuan Fishery Professional Cooperative/Qingtongxia City, Wuzhong 751600, China
5. Sichuan Yukun Aquatic Products Technology Co., Ltd., Dazhou 635000, China

Abstract: Skull morphology is fundamental to biological evolution and species adaptation to their environment. For aquaculture fish species, head size is also a critical economic trait linked to fillet yield and ornamental value. Therefore, identifying the genetic loci controlling the catfish head size can provide target molecular markers and functional genes for genetic selection and breeding. This study identifies quantitative trait nucleotides (QTNs) and genes associated with catfish head size. It performs genotyping using a hybrid population derived from 10 catfish families, conducts genome-wide association study (GWAS) with the efficient and robust mixed model-based Optim-GRAMMAR method, and compares the results with those from the EMMAX method. Based on 206,763 high-quality single nucleotide polymorphisms (SNPs) obtained, the EMMAX method detects no significantly associated SNPs under the same conditions. In contrast, Optim-GRAMMAR identifies 1, 6, and 1 SNPs associated with head length, head depth, and head width, respectively. Most of these genes are involved in growth and development, cell proliferation, metabolic promotion, skeletal morphology, and feeding regulation. Among these genes, *slc12a5a*, *slc7a10a*, and *Rab2a* exert positive effects on growth and development in species such as zebrafish. The QTNs and genes reported in this study enhance our understanding of the genetic architecture underlying head size traits in catfish and will facilitate marker-assisted selection breeding in aquaculture of catfish.

收稿日期: 2025-05-19

修回日期: 2025-11-13

基金项目: 农业部黄河渔业资源与环境调查项目(HHDC-2022-06); 中国水产科学研究院水产生物遗传大数据研究与应用创新团队项目(2023TD25)

第1作者简介: 陈 澳(2002-), 男, 硕士研究生, 研究方向: 遗传育种。E-mail: m19503860141@163.com

*通讯作者: Author for correspondence. E-mail: jiangl@cafs.ac.cn

Keywords: Genome-wide association study; Optim-GRAMMAR; EMMAX; catfish; head size

头骨形态和身体构造是生物进化和物种适应环境的基础。物种进化出不同的头部形状和大小,以适应环境,并与其行为和生存方式有关。因此,以前曾在鱼类、青蛙(*Pelophylax nigromaculatus*)、狗(*Canis lupus familiaris*)、小白鼠(*Mus musculus*)和人类(*Homo sapiens*)上研究过几种控制头骨形状的基因^[1-6]。对于水产养殖鱼类来说,头部大小也是重要的经济性状,因为它与鱼片产量直接相关。较小的头部和均匀的体型能提供更大比例的鱼片,具有较高的水产养殖价值。了解头部形状和大小的基因及发育信号通路将有助于我们解析其遗传调控机制。此外,连锁的标记将可以用于标记辅助选择(Marker Assisted selection, MAS)。全基因组关联分析(GWAS)是一种基于高密度基因分型标记(一般为单核苷酸多态性),通过病例对照或相关性分析来确定与复杂性状相关的遗传变异的策略。近年来,这一策略被越来越多地应用于水产养殖物种,以探索相关性状与标记之间的连锁关系,如生长^[7]、鱼片产量^[8]、鱼片营养^[9]、性别决定^[10]和鱼头大小^[2,11]等。目前,在经济性状相关的分子遗传解析方法的效率上也在不断取得进步。在有关鲢鱼头部大小的 GWAS 研究中^[11],将雌性沟鲢(*Ictalurus punctatus*)与雄性蓝鲢(*I. furcatus*)交配产生的杂交鲢鱼作为实验样本,以鉴定与鲢鱼头部大小遗传相关的 SNP (Single Nucleotide Polymorphism) 位点。一般来说,沟鲢(*Ictalurus punctatus*)的头部比蓝鲢(*Ictalurus furcatus*)大,它们的 F1 代杂交种因异质性而表现出许多优良性状,包括生长速度快、抗病性强、鱼片产量高^[12,13]。因此,蓝鲢杂交系统为研究鱼头形状和大小提供了一个很好的模型。EMMAX^[14]通过固定方差分量来简化计算,大大提高了线性混合模型的计算速度,被广泛应用于全基因组关联分析。但 EMMAX 是近似方法,可能会产生不准确的检验统计量,特别是当数据集中的个体亲缘关系密切,且强相关的 SNPs 在表型变异中占很大比例时,使用方差分量的估计值进行检验可能会得到保守的 p 值,从而导致潜在的检测效率损失^[15]。Optim-GRAMMAR^[16]是一种高效、稳健的全基因组关联方法。Optim-GRAMMAR 法不是对多基因遗传力进行精准的

估计,而是通过对多基因遗传力进行适当的低估,这一做法解决了 GRAMMAR^[17]法可能出现的假阴性问题。Optim-GRAMMAR 具有以下四个独特的优点:(1)在大规模复杂群体中极大地简化了 GWAS,计算复杂度近似于简单回归法;(2)不需要预先估计方差分量或基因组遗传力;(3)通过联合关联分析,统计效力比现有方法显著提高;(4)显示出稳定和好的统计特性,不依赖于数量性状的模拟 QTN (Quantitative trait Nucleotides) 的数量和群体结构的复杂程度。我们认为 Optim-GRAMMAR 是非常适合鲢鱼头型的家系样本分析的统计方法。在本研究中,我们利用高效的 Optim-GRAMMAR 对鲢鱼头部大小(头尺)遗传性状进行全基因组关联分析,汇总了所发现的 SNP 位点以及与头尺性状决定显著相关的基因。

1 材料和方法

1.1 实验鱼和样本采集

研究群体为 1 龄鲢鱼,由雄性 F1 杂交鲢鱼(雌性沟鲢,雄性蓝鲢)与雌性沟鲢回交而成。种群由 10 个家系组成,后代进行混合共养。共随机获得 614 尾鱼,并采集了血液样本。在鱼生长 1 年又两个月时测量了目标鱼群鱼头的大小,包括头长、头宽和头深,以此作为研究对象的性状。头长是指上颌骨骨骺与腭后骨缘之间的水平距离。头宽是腭后骨质边缘两侧之间的距离。头部深度是颅骨顶部到底部横跨腭后骨性边缘的垂直距离。

1.2 DNA 分离、基因分型和质量控制

使用 TIANGEN 血液 DNA (Deoxyribo Nucleic Acid) 提取试剂盒 (DP304) 从鲢鱼血液样本中分离 DNA,利用华大智造的 DNBSEQ-T7 平台进行重测序基因分型,并与在 ncbi 上下载的鲢鱼 (*Ictalurus punctatus*) 基因组 ASM400665v2 对比。获得分型数据后,对其进行进一步的填充过滤,剔除不符合分析标准的 SNP 位点,以减少结果出现假阳性的可能。首先使用 BEAGLE v4.1 对缺失的基因型数据进行填充,随后用 PLINK v1.90 软件筛选出基因分型错误、等位基因频率 (MAF) < 5% 或检出率 < 95% 的 SNP 后,保留

206 763 个高质量 SNP 位点用于后续全基因组关联分析(GWAS)。

1.3 分析方法

我们使用了两种方法来分析体尺性状,即 EMMAX 和 Optim-GRAMMAR。在进行相关分析之前,我们将表型值缺失的行列(3 个个体)去除,以减少误差。此外,EMMAX 会以每个样本的前四个主成分得分作为协变量,而 Optim-GRAMMAR 则是剔除了家系的固定效应,这是两个统计方法在进行运行时的不同点。一个 QTN 所解释的表型方差的比率为 $R^2 = \sigma_{QTN}^2 / \sigma_p^2$, (σ_{QTN}^2 为该 QTN 的效应方差, σ_p^2 为表型方差)。基于 Bonferroni 校正,全基因组显著性水平设定为 $0.05/184\ 507 = 2.71e-7 [-\log_{10}(P\text{-value}) = 6.57]$ 。

1.4 Optim-GRAMMAR 原理

Optim-GRAMMAR 软件可从 <https://github.com/RunKingProgram/Optim-GRAMMAR> 免费获取。

模型如下:

$$y = X\beta + Za + g + e$$

其中 y 为矫正固定效应后的表型值向量; β 是如家系、性别和年龄等固定效应组成的向量, X 是其所对应的指示变量矩阵, a 是被检验单核苷酸多态性对表型的加性遗传效应; Z 为被检验 SNP 标记所组成的基因型指示变量,它常常用 0, 1 和 2 分别编码三种不同的基因型; $g: N(0, K\sigma_g^2)$ 和 σ_g^2 为多基因方差; e 是残差,分布遵循 $e \sim N(0, I\sigma_e^2)$, 其中 σ_e^2 为残差方差, I 为单位矩阵。

采用 Optim-GRAMMAR 进行统计分析,优化步骤如下: 1. 在开区间 0-1 设定初始遗传力。2. 根据 BLUP 方程求解出基因组育种值 g 。3. 统计学上推断每个 SNP 的遗传效应。4. 计算每个 SNP 的全基因组卡方统计量的均值。5. 绘制全基因组统计概率的 Q-Q 图。6. 使用 Brent 法更新遗传力。7. 重复第 2 到第 6 步,直至基因组控制值 λ 接近 1,基因组控制值越接近 1,则代表研究群体中的分层现象得到了较好的控制,不同亚群体之间的遗传差异对结果的干扰较小,使得关联分析结果更能真实地反映基因与性状之间的关系,降低了因群体结构导致的虚假关联的可能性,或者得到满意的 Q-Q 图。

1.5 序列分析

根据候选基因的位置和功能,对每个 QTN 中最显著 SNP 周围的 1Mb 区域进行了研究。利用 FGENESH^[18] 预测基因组区域内的基因,因为参考基因组的注释信息对于我们所检测出来的 SNP 位点旁边并无注释信息,所以需要通过 BLAST 分析与非冗余蛋白质数据库^[19,20]进行注释。

2 结果与分析

2.1 头部性状的 GWAS 分析结果

分别使用 EMMAX 和 Optim-GRAMMAR 法对头深性状进行 GWAS 分析结果如图 1 所示,EMMAX 法没有检测到 SNP 与头深性状连锁,还出现了严重的假阴性(Q-Q 图可以看出)。相比之下,使用 Optim-GRAMMAR 法检测到了 6 个与头深性状连锁的 SNP, GWAS 结果不存在明显的假阳性或者假阴性,拟合度较好。

分别使用 EMMAX 和 Optim-GRAMMAR 法对头宽性状进行 GWAS 分析结果如图 2 所示,EMMAX 法没有检测到 SNP 与头宽性状连锁,拟合度稍差,有轻微假阴性。相比之下,使用 Optim-GRAMMAR 法检测到了 1 个与头宽性状连锁的 SNP, GWAS 结果不存在明显的假阳性或者假阴性,拟合度较好。

分别使用 EMMAX 和 Optim-GRAMMAR 法对头长性状进行 GWAS 分析结果如图 3,EMMAX 法没有检测到 SNP 与头长性状连锁,拟合度稍差,有轻微假阴性。相比之下,使用 Optim-GRAMMAR 法检测到了 1 个与头长性状连锁的 SNP, GWAS 结果不存在明显的假阳性或者假阴性,拟合度较好。

2.2 EMMAX 与 Optim-GRAMMAR 的结果比较

由 EMMAX 和 Optim-GRAMMAR 结果生成的曼哈顿图和 Q-Q 图分别见图 1~2 和图 3。基于 EMMAX 计算出的头深、头长和头宽的基因组控制值分别为 1.068 1、1.011 2 和 1.081 3,而使用 Optim-GRAMMAR 计算出的头深、头长和头宽的基因组控制值分别为 1.002 3、1.105 0 和 1.080 7。如图 1~3 所示,EMMAX 的大多数检验统计量超过理论线,与 EMMAX 相比,Optim-GRAMMAR 前段 Q-Q 图完美的贴合理论线,并且在 SNP 位

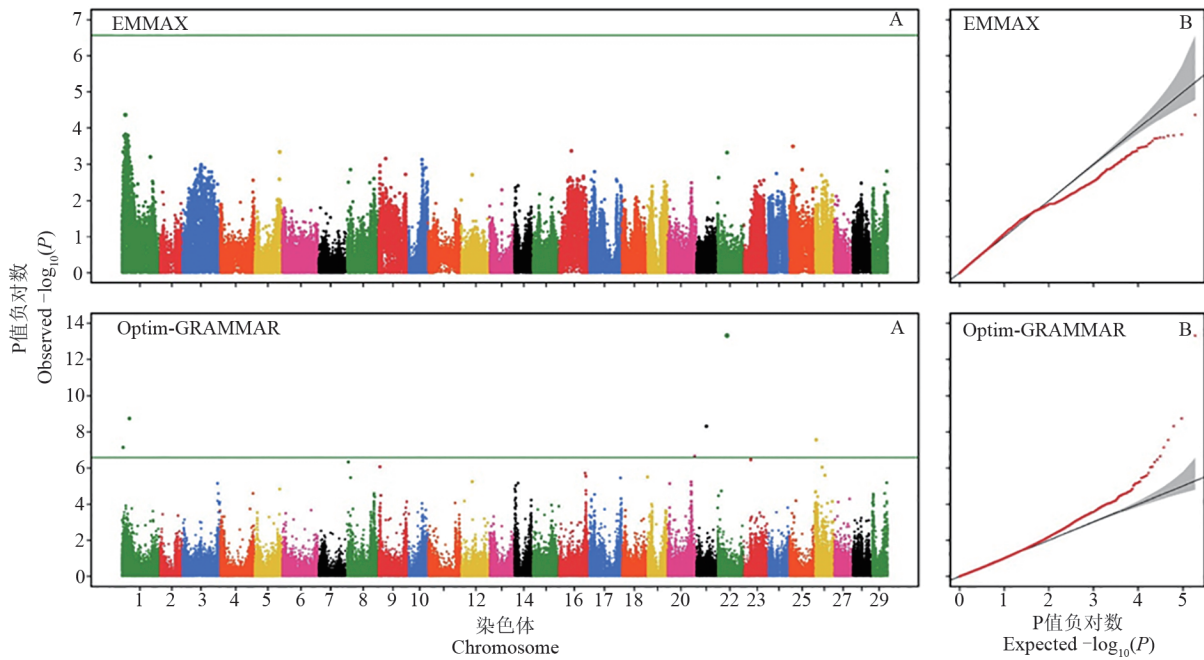


图 1 使用 EMMAX 和 Optim-GRAMMAR 得出的头部深度的 Manhattan 图(A)和 Q-Q 图(B)

Fig. 1 Manhattan plots(A) and Q-Q plots(B) for head depth obtained with EMMAX and Optim-GRAMMAR

注:水平线是基于 Bonferroni 校正的全基因组显著性。

Note: The horizontal line is genome-wide significance based on Bonferroni correction.

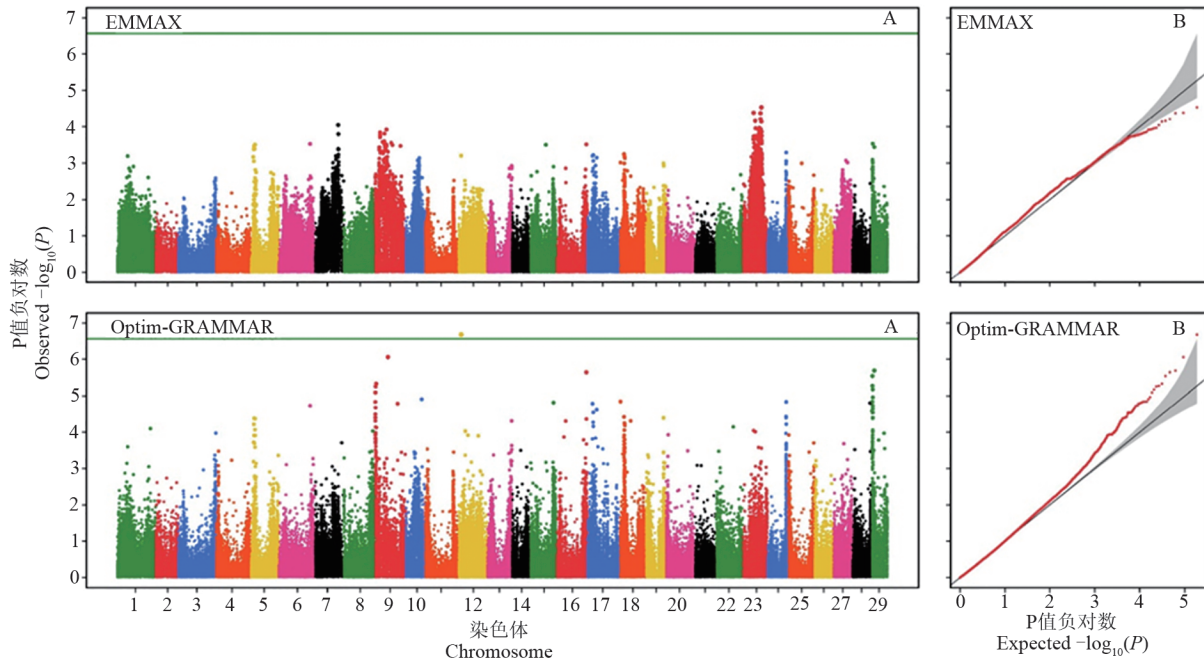


图 2 利用 EMMAX 和 Optim-GRAMMAR 得出的头宽曼哈顿图(A)和 Q-Q 图(B)。

Fig. 2 Manhattan (A) and Q-Q(B) plots for head width obtained with EMMAX and Optim-GRAMMAR.

注:水平线是基于 Bonferroni 校正的全基因组显著性。

Note: The horizontal line is genome-wide significance based on Bonferroni correction.

置提供了更高的检验统计量,同时拥有完美的基因组控制值。EMMAX 没有检测到 SNP,而使用 Optim-GRAMMAR 对头长、头深和头宽进行

SNP 检测,分别检测到 1、6、1 个 SNP。表 1 列出了与头部大小性状(头深、头长、头宽)显著相关的 SNP 信息。

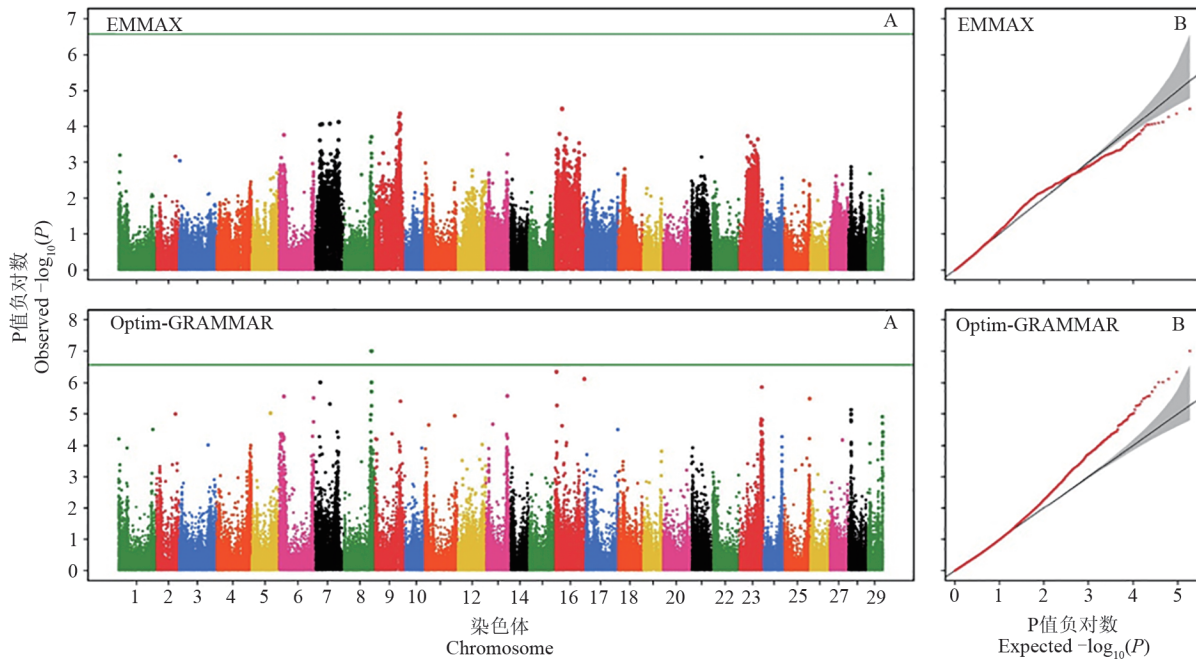


图 3 使用 EMMAX 和 Optim-GRAMMAR 得出的头长的曼哈顿图(A)和 Q-Q 图(B)。

Fig. 3 Manhattan (A) and Q-Q (B) plots for head length obtained with EMMAX and Optim-GRAMMAR.

注:水平线是基于 Bonferroni 校正的全基因组显著性。

Note: The horizontal line is genome-wide significance based on Bonferroni correction.

表 1 从 Optim-GRAMMAR 中获得的与头部大小特征相关的 SNP

Table 1 The SNPs associated with head size traits obtained from Optim-GRAMMAR

性状 Trait	染色体 Chromosome	单核苷酸多态性标识符 SNP ID	单核苷酸多态性 位置 SNP position	负对数值 $-\log_{10}(P)$	效应值 Effect
头深 Head depth	1	85396595	1398786	7.135 0	0.128 3
	1	85213638	7667386	8.742 3	0.174 5
	20	85384972	27289727	6.644 2	0.027 6
	21	85260459	10081917	8.312 2	0.072 1
	22	85285966	9288869	13.325 8	0.244 5
	26	85385060	1133528	7.539 5	-0.129 9
头宽 Head width	12	86130377	2636929	6.685 5	0.136 8
头长 Head length	8	86041121	27882954	7.007 9	0.186 2

表 2 与头部长度相关的 SNP

Table 2 SNPs related to head length

性状 Trait	染色体 Chromosome	单核苷酸多 态性标识符 SNP ID	单核苷酸多 态性位置 SNP position	负对 数值 $-\log_{10}(P)$	变异占比/% Variance	基因位置/kb Gene position	基因名称 Gene name	基因功能 Gene function
长度 Length	8	86041121	27882954	7.007 9		27,597,776..27, 849,043	<i>LOC124390429</i>	离子型谷氨酸受体 活性
						27,889,781..27, 934,957	<i>plb1</i>	溶血磷脂酶活性
						28,116,801..28, 147,219	<i>clic4</i>	氯离子通道活性
						28,311,351..28, 344,791	<i>kif20ba</i>	激活 ATP 水解活 性、微管结合活性 和微管马达活性

2.3 头部尺寸各性状关联基因的遗传信息

表 3 与头部深度相关的 SNP
Table 3 SNPs related to head depth

性状 Trait	染色体 Chromosome	单核苷酸 多态性标 识符 SNP ID	单核苷酸 多态性位 置 SNP position	负对数 值 $-\log_{10}$ (<i>P</i>)	变异占 比/% Variance	基因位置/kb Gene position	基因名称 Gene name	基因功能 Gene function
深度 Head depth	1	85396595	1398786	7.135 0	1.5	12,967,088...12, 998,695	<i>anks1ab</i>	生长因子受体结合 活性
						13,228,845...13, 452,480	<i>cacna2d2a</i>	电压门控钙通道活性
						13,619,782..13, 650,620	<i>tead3b</i>	DNA 结合转录因子 活性
						13,723,688..13, 807,671	<i>nav1b</i>	神经胶质细胞发育
						13,991,985..14, 132,347	<i>slc12a5a</i>	钾离子转运
						7,422,487..7,463,653	<i>nck2b</i>	蛋白质结合
	2	85213638	7667386	8.742 3	2.1	7,505,146..7,556,216	<i>lrrfip1b</i>	DNA 结合转录因子 活性
						7,627,604..7,678,268	<i>iqca1</i>	ATP酶活性
						7,698,761..7,841,620	<i>agap1</i>	膜运输和细胞骨架动 力学
	20	85384972	85384972	6.644 2	0.2	9,866,192..9,907,543	<i>LOC124375480</i>	B细胞增殖的正调控
	21	85260459	85260459	8.312 2	0.8	9,930,658..9,965,103	<i>mppel</i>	GPI 锚定结合活性
						9,965,641..10, 040,092	<i>gnal</i>	嗅觉上皮中气味信号 传导
						10,263,952..10, 298,250	<i>sycp2l</i>	细胞减数分裂核分裂
						8,976,557..9,054,962	<i>cacna1ha</i>	低电压门控钙通道活 性和电压门控钠通道 活性
	22	85285966	9288869	13.325 8	1.65	9,151,154..9,167,475	<i>rhot1a</i>	GTP结合活性和GTP 酶活性
						9,081,048..9,100,922	<i>epn2</i>	网格蛋白结合活性和 磷脂结合活性
9,204,099..9,220,393						<i>ago4</i>	mRNA 介导的翻译 抑制	
11,000,203..11, 071,727						<i>igdcc3</i>	细胞间粘附	
26	85385060	1133528	7.539 5	2.6	11,082,031..11, 139,796	<i>Igdcc4</i>	细胞间粘附	
					11,160,934..11, 184,964	<i>rfx7a</i>	DNA 结合活性和 DNA 结合转录因子 活性	
					11,225,355..11, 242,759	<i>lrp3</i>	低密度脂蛋白受体蛋 白活性	
					11,246,546..11, 276,828	<i>slc7a10a</i>	L 氨基酸跨膜转运蛋 白活性和中性氨基酸 跨膜转运蛋白活性	

表 4 与头部宽度相关的 SNP
Table 4 SNPs related to head width

性状 Trait	染色体 Chromosome	单核苷酸多态性标识符 SNP ID	单核苷酸多态性位置 SNP position	负对数值 $-\log_{10}(P)$	变异占比/% Variance	基因位置/kb Gene position	基因名称 Gene name	基因功能 Gene function
宽度 Head Width	12	86130377	2636929	6.685 5	3.1	25,928,649..25,992,457	<i>grin3ba</i>	配体门控离子通道活性和信号受体活性
						26,089,042..26,160,564	<i>efna2a</i>	肝配蛋白受体结合活性
						26,372,852..26,393,116	<i>rab2a</i>	GTP 结合活性和 GTP 酶活性
						26,405,299..26,469,755	<i>chd7</i>	ATP 结合活性、DNA 结合活性和 DNA 螺旋酶活性。
						26,474,765..26,491,666	<i>asph</i>	肽基-天冬氨酸 3-双加氧酶活性

3 讨论

基于大型家系的资源群体适用于大多数水产养殖物种开展 GWAS, 因为大多数水产养殖的鱼类每次产卵的后代数量较多, 可以确保以较低的成本获得足够的样本, 本研究使用了十个家系的样本。与非亲缘样本相比, 基于大型家系的 GWAS 在鉴定 SNP 方面具有一些优势^[21]。首先, 致病变异和相关标记物之间缺乏重组, 从而提高了检测能力^[22]。其次, 基于家系的群体设计具有明确的血统信息, 因此更容易控制群体分层造成的混杂影响。然而, 家系的缺陷也不容忽视, 家系或群体规格是造成表型差异的主要原因之一。基于家系的样本中个体的数量有限, 可能会降低检出 SNP 的能力。在样本中加入更多的家系可以进行相互验证, 提高图谱分辨率, 提高统计效力, 避免假阳性。

在之前的研究中, 可以发现在对鲢鱼头部生长发育相关的基因中, 也有许多基因也同时直接或间接作用于另外的鱼类头部性状, 例如 *slc12a5a* 在斑马鱼 (*Danio rerio*) 中通过钾离子转运和跨膜转运的上游或内部起作用, 间接对斑马鱼头部大小起作用; *slc7a10a* 在斑马鱼基因组中是非重复的, 对其的研究有助于阐明二元氨基酸、胱氨酸和(大)中性氨基酸的吸收与此类营养物质对生物体生长的影响之间的复杂相互作用^[23]; *Rab2a* 在斑马鱼体内启用 GTP 结合活性和 GTP 酶活性, 在轴向中胚层中表达, 对其生长发

育有着很大的间接作用^[24]。 *kif20ba* 在斑马鱼骨骼肌发育过程中表现出不重叠的表达模式, 可进一步研究其横纹肌功能^[25]; *Rfx7a* 基因的表达和斑马鱼, 鲤 (*Cyprinus carpio*), 鲢鱼等鱼类中有着相同的作用^[26]; *CHD7* 不仅在斑马鱼, 人类, 甚至在长鳍领航鲸, 虎鲸等动物中也有着充分的表达^[27]; *agap1* 在斑马鱼体内可实现多种功能, 包括 GTP 结合活性, GTP 酶激活剂活性和 GTP 酶活性, 也可在信号转导的上游或内部起作用, GTP 酶作为能量运输的关键, 在能量运输、信号传导、蛋白质合成和细胞分化中起着关键的作用, 对斑马鱼头部大小性状有重要作用。 *Clic4* 能够启用氯离子通道活性, 参与氯离子转运, 预计在谷胱甘肽代谢过程的上游或内部起作用; 离子传输; 和离子跨膜转运的调节, 是氯离子通道复合物的一部分, 在细胞质和膜中具有活性。

在之前的研究中, 针对基于家系的样本评估了 EMMAX 和 QFAM 这两种方法, QFAM 将基因型分为家系间和家系内两个部分^[28,29]。家系内分型可以控制分层, 从而不受分层的影响, 确定真正的关联结果。与 QFAM 不同的是, EMMAX 首先根据高密度标记计算成对亲缘关系矩阵, 以表示样本结构。然后, EMMAX 可以估计样本结构对表型的贡献, 并在没有样本结构产生的混杂效应的情况下检测关联^[14]。然而, 有研究表明, 在计算成对亲缘关系矩阵时加入候选标记可能会导致模型中候选标记的双重拟合, 从而导致损失统计效力^[30]。基于家系的

关联检验似乎更适合于水产养殖中广泛使用的基于家系的样本。虽然 EMMAX 和 Optim-GRAMMAR 均基于之前的方法进行了调整,但更重要的是,尽管 EMMAX 采用前面几个显著关联的 PCs 作为协变量大大降低了模型的假阳性,但是又表现出检测结果的假阴性,但在 Optim-GRAMMAR 中显示了不同的统计结果。结合此前的研究结果^[15]和本研究所示,虽然 GRAMMAR 近似会导致所有 p 值被低估,但 Optim-GRAMMAR 可以完全纠正 GRAMMAR 法的假阴性。因此,Optim-GRAMMAR 避免了检验统计量的低估,从而得到一个稳健的 GWAS 检测结果。更重要的是,经过联合分析,Optim-GRAMMAR 提高了头部深度的 QTN 的检验效力。通过 Optim-GRAMMAR 的优化步骤,该方法可以实现快速的全基因组关联分析,并有效避免关联统计量的紧缩和小群体基因组遗传力的不准确估计问题。

4 结论

本研究中,分别在头深,头长,头宽等性状中得到的关联的 SNP 位点,他们对性状表达的百分比分别为 1.5、2.1、0.2、0.8、1.65、2.6、3.1,这表示我们虽然找到了很多显著的 SNP 位点,但是他们占比都比较小,并没有出现大比重,表示仍然存在很多 SNP 位点仍未找到。此次找到的 QTN 位点相邻相近的基因均直接或间接作用于生长发育,同时也发现这些性状是由微小多基因所共同决定的,他们之间位置相距不远,可以肯定他们有着类似的作用,为以后的育种选育工作提供了候选位点。

Optim-GRAMMA 和 EMMAX 的研究方法各有优劣,不能说因为我们这次没有利用 EMMAX 找到显著点位就说 EMMAX 这个方法不行,只能说其在此次 GWAS 研究中不如 Optim-GRAMMA 的检测效率高,Optim-GRAMMA 是在 GRAMMA 上加入了新的算法机制,使其更加适合于 GWAS 研究,并且在其中加入了二进制 C 语言语法,其运算效率在我们之前的测试中也是有着极高的优势。Optim-GRAMMA 和 EMMAX 相对比,能够体现研究结果的准确度以及 Optim-GRAMMA 方法的优势。我们在研究中发现的更多 SNPs 将有助于了解鲶鱼头部大小

性状的遗传基础。从长远目标来看,详细的 SNP 信息将用于通过标记辅助选择来改良种群,提高鲶鱼的育种效率。本研究找到的与头部大小性状显著相关的 SNP 位点,以及与其相邻或相近的基因,这些基因为以后的改良育种以及利用这些基因进行基因编辑育种提供了重要的候选基因。

参考文献

- [1] Albertson R C, Streelman J T, Kocher T D. Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes[J]. Proceedings of the National Academy of Sciences, 2003, 100(9): 5252-5257.
- [2] Chen L, Peng W, Kong S, et al. Genetic mapping of head size related traits in common carp (*Cyprinus carpio*)[J]. Front Genet, 2018, 9(9):448.
- [3] Wunnenberg-Stapleton K, Blitz I L, Hashimoto C, et al. Involvement of the small GTPases XRhoA and XRnd1 in cell adhesion and head formation in early *Xenopus* development[J]. Development, 1999, 126(23):5339 - 5351.
- [4] Schoenebeck J J, Ostrander E A. The genetics of canine skull shape variation[J]. Genetics, 2013, 193(2): 317 - 325.
- [5] Schoenebeck J J, Hutchinson S A, Byers A, et al. Variation of BMP3 contributes to dog breed skull diversity[J]. PLoS Genet, 2012, 8(8):e1002849.
- [6] Maga A M, Navarro N, Cunningham ML, et al. Quantitative trait loci affecting the 3D skull shape and size in mouse and prioritization of candidate genes in-silico[J]. Front Physiol, 2015, 26(6):92.
- [7] Su S, Raouf B, He X, et al. Genome wide analysis for growth at two growth stages in a new fast-growing common carp strain (*Cyprinus carpio* L.) [J]. Scientific Reports, 2020, 10(1):7259.
- [8] Gonzalez-Pena D, Gao G, Baranski M, et al. Genome-Wide association study for identifying loci that affect fillet yield, carcass, and body weight traits in rainbow trout (*Oncorhynchus mykiss*) [J]. Front Genet, 2016, 22(7):203.
- [9] Horn S, Ruyter B, Meuwissen E, et al. GWAS identifies genetic variants associated with omega-3 fatty acid composition of Atlantic salmon fillets[J]. Aquaculture, 2019:734494.
- [10] Cáceres G, López E, Cádiz I, et al. Fine mapping using whole-genome sequencing confirms anti-müllerian hormone as a major gene for sex determination in

- farmed Nile tilapia (*Oreochromis niloticus* L.)[J].G3: Genes,Genome,Genetics,2019,9(10):3213-3223.
- [11] Geng X, Liu S, Yao J, et al. A genome wide association study identifies multiple regions associated with head size in catfish[J]. G3: Genes, Genome,Genetics,2016,6(10):3389-3398.
- [12] Argue B J,Liu Z ,Dunham R A.Dress-out and fillet yields of channel catfish, *Ictalurus punctatus*, blue catfish, *Ictalurus furcatus*, and their F1, F2 and backcross hybrids[J].Aquaculture,2003,228(1):81-90.
- [13] Dunham A,Umalı G, Beam R,et al. Comparison of production traits of NWAC103 channel catfish, NWAC103 channel catfish×blue catfish hybrids, Kansas Select 21 channel catfish, and blue catfish grown at commercial densities and exposed to natural bacterial epizootics[J]. North American Journal of Aquaculture,2008,70(1):98-106.
- [14] Kang H M, Sul J H, Service S K, et al. Variance component model to account for sample structure in genome-wide association studies[J].Nature Genetics, 2010,42(4):348-354.
- [15] Zhou X,Stephens M. Genome-wide efficient mixed-model analysis for association studies[J]. Nature Genetics,2012,44(7):821-824.
- [16] Yang R,Gao J, Song Y, et al. Grammar-Lambda: an extreme simplification for genome-wide mixed model association analysis[J]. BioRxiv, 2021: 2021.03.10.434574.
- [17] Aulchenko Y S, De Koning D J,Haley C. Genome wide rapid association using mixed model and regression: A fast and simple method for genomewide pedigree-based quantitative trait loci association analysis[J]. Genetics, 2007, 177(1): 577-585.
- [18] Froese R. Cube law, condition factor and weight - length relationships: History, meta - analysis and recommendations[J].Journal of Applied Ichthyology, 2006,22(4):241-253.
- [19] Solovyev V, Kosarev P, Seledsov I ,et al. Automatic annotation of eukaryotic genes, pseudogenes and promoters[J].Genome Biology,2006,7(Suppl 1):S10.
- [20] Pruitt K D,Tatusova T,Maglott D R . NCBI reference sequences (RefSeq): A curated non-redundant sequence database of genomes, transcripts and proteins[J]. Nucleic Acids Research, 2007, 35 (Database issue):D61- D65.
- [21] Altschul S F, Gish W, Miller W, et al. Basic local alignment search tool[J]. Journal of Molecular Biology, 1990, 215(3): 403-410.
- [22] Ott J,Kamatani Y,Lathrop M. Family-based designs for genome-wide association studies[J]. Nature Reviews Genetics,2011,12(7):465-474.
- [23] Ellingsen S,Narawane S,Fjose A,et al. The zebrafish cationic amino acid transporter/glycoprotein-associated family: sequence and spatiotemporal distribution during development of the transport system b 0, +(slc3a1/slc7a9)[J].Fish Physiology and Biochemistry,2021, 47:1507-1525.
- [24] Diotel N, Rodriguez Viales R, Armant O, et al. Comprehensive expression map of transcription regulators in the adult zebrafish telencephalon reveals distinct neurogenic niches[J]. Journal of Comparative Neurology,2015,523(8):1202-1221.
- [25] Boskovic S, Marin-Juez R, Jasic J, et al. Characterization of zebrafish (*Danio rerio*) muscle ankyrin repeat proteins reveals their conserved response to endurance exercise[J].PloS One,2018,13 (9):e0204312.
- [26] Pasquier J,Cabau C,Nguyen T,et al. Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database[J]. BMC Genomics, 2016,17:1-10.
- [27] Sanosaka T, Okuno H, Mizota N, et al. Chromatin remodeler CHD7 targets active enhancer region to regulate cell type-specific gene expression in human neural crest cells[J]. Scientific Reports, 2022, 12(1): 22648.
- [28] Purcell S,Neale M,Todd-Brown E,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] Fulker D W, Cherny S S, Sham P C,et al. Combined linkage and association sib-pair analysis for quantitative traits[J]. American Journal of Human Genetics,1999,64(1):259.
- [30] M Ac Kay I , Powell W . Methods for linkage disequilibrium mapping in crops[J]. Trends in Plant Science,2007, 12(2):57-63.