

DOI: 10.12264/JFSC2020-0557

大菱鲂生长性状表型正交化混合模型关联分析

蒋仪^{1,3}, 宋禹昕², 高进³, 陈鹤丽³, 杨润清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. 南京农业大学无锡渔业学院, 江苏 无锡 214081;

3. 中国水产科学研究院 生物技术研究中心, 北京 100141

摘要: 多变量线性混合模型是检验单个核苷酸多态性位点与多个相关表型之间是否存在关联的强有力工具。为提高该复杂关联分析的计算效率, 本研究对多个性状的表型协方差矩阵进行谱分解, 将这些相关的性状转换为相互独立的“超性状”, 利用单性状的混合模型关联分析方法计算每个超性状的检验统计量, 通过这些统计量之和推断控制多个相关表型的一因多效核苷酸遗传位点。运用这种表型正交化的多性状混合模型关联分析方法, 检测大菱鲂(*Scophthalmus maximus*)的体重-体长-尾柄宽 3 个性状一因多效核苷酸遗传位点, 结果表明, GEMMA 方法在 1、5、6、8、12、16、20、22 号染色体上共检测到 11 个 QTNs, 而表型正交化方法则在 1、3、5、6、8、12、16、20、21、22 号染色体上共检测到 14 个 QTNs, 其中 2 种方法共同检测到的 QTNs 有 9 个。与 GEMMA 相比, 表型正交化方法具有更高的多性状检测效率, 且运算时间也远低于 GEMMA, 可高效定位大菱鲂生长性状相关 QTNs, 这为其他水产动物的多性状 GWAS 研究以及遗传育种提供了一个便捷高效的方法。

关键词: 大菱鲂; 多变量线性混合模型; 谱分解; 生长性状; 表型正交化

中图分类号: S931

文献标志码: A

文章编号: 1005-8737-(2021)07-0845-07

全基因组关联分析(genome-wide association study, GWAS)为全基因组范围内单个核苷酸变异所引起的 DNA 序列多态性即单核苷酸多态性(single nucleotide polymorphism, SNP)的检测提供了强有力的工具, 它能鉴别与复杂性状相关联的数量性状核苷酸(quantitative trait nucleotides, QTNs), 实现目标性状的基因定位。目前, 大多数的 GWAS 模型都是针对单个性状进行遗传位点和表型间的关联分析。当对个体进行多尺度表型测量时, 共同的环境效应和连锁不平衡使得表型间具有潜在相关性, 存在一个基因同时影响多个性状的情况, 即一因多效性。因此当不同性状间存在遗传相关时, 可以对多个潜在相关性状进行联合分析, 由于考虑了性状表型间的协方差, 与

单性状分析相比会显著提升统计检验的效力, 且具有一因多效性的 QTN 能够解释多个性状间的遗传亲缘相关, 结果更具有生物学意义^[1-3]。

线性混合模型(linear mixed model, LMM)由于能够综合考虑群体分层, 家系结构和复杂的亲缘关系等混杂因子已成为 GWAS 中广泛使用的分析模型。在多性状 GWAS 中, 多变量线性混合模型(multivariate linear mixed model, mvLMM)可高效定位一因多效 QTN, 提高多个相关性状中单个性状的检测效率^[4]。此外, mvLMM 在估计组织间基因表达的遗传力^[5], 理解进化模型和辅助动物遗传育种中也发挥着重要作用^[1,6-9]。对于多个相关表型性状, 在使用约束最大似然法估计多性状表型间的方差-协方差矩阵时, mvLMM 的计算

收稿日期: 2020-12-01; 修订日期: 2021-01-29.

基金项目: 国家自然科学基金面上项目(32072726); 中央级公益性科研院所基本科研业务费专项资金项目(2019ZY09).

作者简介: 蒋仪(1996-), 女, 硕士研究生, 研究方向为数量遗传学. E-mail: 1586997696@qq.com

通信作者: 杨润清, 研究员, 研究方向为数量遗传学. E-mail: runqingyang@cafs.ac.cn

复杂度将比 LMM 增加性状个数倍。基于 mvLMM 的精确关联分析需要逐个 SNP 重复求解多性状混合模型,这对具有大量 SNP 和中等规模群体而言是不切实际的。GEMMA 使用无效模型中相同方差-协方差矩阵替换每个 SNP 间不同的协方差矩阵,并在计算复杂度相当低的情况下估计 mvLMM 中的协方差矩阵,这大大降低了计算时间的消耗和计算机内存运行的负担。然而在更大规模群体的多性状 GWAS 分析中,对更多相关联表型性状进行更为复杂的矩阵运算时, GEMMA 将表现出相对较低的关联统计检验效率。表型正交化方法是对这些相关表型进行降维的一种策略,通过对所分析的相关性状的表型协方差矩阵执行谱变换,对原始表型性状加权可以获得相互独立的“超表型”,然后逐个性状进行关联分析^[10],加和单个性状检验的卡方统计量,最后检验多个关联性状的一因多效 QTN。

大菱鲂(*Scophthalmus maximus*, Linnaeus, 1758)是一种广泛分布在包括波罗的海、黑海和地中海的欧洲大西洋沿岸的经济鱼类^[11]。因其生长速度快,耐冷水能力强,味道鲜美,是世界上养殖最广泛的商业鱼类之一。我国自 1992 年引入以来,大菱鲂已经发展成为我国北方海水养殖产业的主要养殖种类^[12]。因此,探索大菱鲂的生长和耐低温等重要经济数量性状的关联基因或 QTNs,对深入解析其数量性状遗传机制以及提高标记辅助育种效率有着重要的意义。本研究通过比较 GEMMA 和表型正交化方法对大菱鲂体质量和体尺性状多性状全基因组关联分析结果,来评价表型正交化方法的 QTN 检测效率,为水产动物育种在多性状 GWAS 的研究提供更为高效便捷的方法思路。

1 材料与方法

1.1 实验群体及表型测量

本研究以来自 29 个全同胞家系共 585 尾大菱鲂组成的育种群体作为验证表型正交化方法性能的实际资料。于 4 月龄时,对养殖的育种群体中所有大菱鲂个体注射电子标记,并提取鳍条组织 DNA 用于标记测序。随后,在自然光周期下,采

用循环水系统将已标记的个体混养在 6 m×6 m×1 m 的水泥池中,每天投喂两次商用饵料至饱食,水温控制在 5~24 ℃。自注射电子标记起,用电子秤定期称量每个个体体重,并在固定参考标尺下定期用数码相机从一定高度向下垂直拍摄个体的形态性状。到 36 月龄时,绝大多数个体均测量 9 次左右。每次测量前,利用 50×10⁻⁶ MS-222 鱼用安定剂将待测个体麻醉,避免处理压力。根据拍摄图像,利用 Image J 软件获得不同日龄体长和尾柄宽的大菱鲂体尺表型测量值。采用简化基因组 2b-RAD 高通量标记分型技术,共获得 30049 个多态性 SNP 标记,参考大菱鲂基因组(ASM318616v1)建立多态标记物理图谱。在其中挑选生长周期为 473 d 左右的大菱鲂个体的表型值,使用 PLINK v1.9 (<http://www.cog-genomics.org/plink2>)对与表型个体所对应的基因型数据进行严格的质量控制,去除低于 90%最小检出率的个体,以及最小检出频率低于 95%、最小哈代温伯格平衡为 1.0×10⁻⁶、最小等位基因频率小于 3%和方差变异大于 0.05 的 SNPs,最终得到 585 个样本的 21993 个 SNP 标记。

1.2 表型数据处理

为了去除性别、池子、测量日期的固定效应影响,利用 R 语言中的 GLM 模型对表型数据进行校正:

$$y = \mu + \text{Sex} + \text{Pool} + \text{Date} + e \quad (1)$$

式中, y 为个体表型测量值, Sex 为性别, Pool 为养殖池, Date 为测量日期, e 为剩余效应,本研究把剩余效应作为校正后的表型值 Y 用于后续的全基因组关联分析。

1.3 理论方法

1.3.1 多变量线性混合模型 按下列公式计算:

$$Y = Z \otimes b + G + E \quad (2)$$

其中, Y 为 $n \times d$ 维的多性状表型值矩阵, n 为分析的个体数, d 为分析的性状数; b 为 d 维当前检验 SNP 的加性遗传效应行向量; z 为 n 维当前检验 SNP 的基因型指示变量列向量; G 和 E 分别是 $n \times d$ 维的随机多基因效应和剩余误差矩阵。通常, G 和 E 服从多变量正态分布: $G \sim N(0, K \otimes V_g)$, $E \sim N(0, I_n \otimes V_e)$ 。 V_g 和 V_e 分别是 $d \times d$ 维的多基因和

剩余协方差矩阵, $\mathbf{I}_n$ 是单位矩阵, $\mathbf{K}$ 和 $\otimes$ 分别表示基因组亲缘关系矩阵和克罗内克积, 满足 $\mathbf{E}(\mathbf{Y}) = \mathbf{z} \otimes \mathbf{b}$, $\mathbf{V}(\mathbf{Y}) = \mathbf{K} \otimes \mathbf{V}_g + \mathbf{I}_n \otimes \mathbf{V}_e$ 。

1.3.2 多性状表型正交化模型 用 $\mathbf{R}$ 表示 d 个性状的表型协方差矩阵。将 $\mathbf{R}$ 谱分解为 $\mathbf{R} = \mathbf{U}_R \mathbf{S}_R \mathbf{U}_R^T$, 令 $\mathbf{Y}^* = \mathbf{Y} \mathbf{U}_R$, $\mathbf{b}^* = \mathbf{b} \mathbf{U}_R$, $\mathbf{G}^* = \mathbf{G} \mathbf{U}_R$ 和 $\mathbf{E}^* = \mathbf{E} \mathbf{U}_R$ 则模型(1)可被转化为下式:

$$\mathbf{Y}^* = \mathbf{z} \otimes \mathbf{b}^* + \mathbf{G}^* + \mathbf{E}^*$$

(3)

由 $\mathbf{R} = \mathbf{V}_g + \mathbf{V}_e$ 可知, $\mathbf{V}(\mathbf{G}^* + \mathbf{E}^*) = \mathbf{U}_R^T (\mathbf{V}_g + \mathbf{V}_e) \mathbf{U}_R = \mathbf{U}_R^T \mathbf{R} \mathbf{U}_R = \mathbf{S}_R$ 。令 $\mathbf{V}(\mathbf{G}^* + \mathbf{E}^*) = (\mathbf{S}_R - \mathbf{I}_d) + \mathbf{I}_d$, 则模型(2)变成相互独立的:

$$y_l^* = \mathbf{z} b_l^* + g_l^* + e_l^*$$

(4)

式中, $l = 1, 2, \dots, d$, $g_l^* \sim N(0, s_R - 1)$ 与 $e_l^* \sim N(0, 1)$ 。

1.3.3 多性状关联检验 为了消除个体间的相关, 把基因组亲缘关系矩阵谱分解 $\mathbf{K} = \mathbf{U}_k \mathbf{S}_k \mathbf{U}_k^T$, $\mathbf{S}_k$ 和 $\mathbf{U}_k$ 是相对应的特征值和特征根矩阵。令 $y_l' = (\mathbf{S}_K + \mathbf{I}_n)^{\frac{1}{2}} \mathbf{U}_k^T y_l^*$, $\mathbf{z}' = (\mathbf{S}_K + \mathbf{I}_n)^{\frac{1}{2}} \mathbf{U}_k^T \mathbf{z}$, $g_l' = (\mathbf{S}_K + \mathbf{I}_n)^{\frac{1}{2}} \mathbf{U}_k^T g_l^*$ 以及 $e_l' = (\mathbf{S}_K + \mathbf{I}_n)^{\frac{1}{2}} \mathbf{U}_k^T e_l^*$, 然后把模型(4)转换为一般线性回归模型:

$$y_l' = \mathbf{z}' b_l' + e_l'$$

(5)

经最小二乘法求解, 我们使用检验统计量推断标记的效应:

$$\chi_l^2 = \hat{b}_l'^2 [\mathbf{V}(\hat{b}_l')]^{-1}$$

(6)

式中, $\hat{b}_l'$ 是估计的遗传效应, $\mathbf{V}(\hat{b}_l')$ 是当前检验 SNP 的方差。

对 d 个独立的模型(3), 我们加和服从 d 自由度卡方分布的 d 个统计量(5)去检验一因多效 QTN:

$$\chi^2 = \sum_{l=1}^d \chi_l^2$$

(7)

2 结果与分析

2.1 表型值的统计性质

本研究选取生长天数为 473 d 左右的 585 尾大菱鲆的体重、体长、尾柄宽 3 个性状, 分别采用 GEMMA 和表型正交化两种方法进行多性状 GWAS 分析。图 1 为大菱鲆 3 个生长性状的表型

值正态分布图, 从中可以看出, 各个性状的表型值都符合正态分布, 适合进行后续的多性状 GWAS 分析。表 1 为各性状的原始表型值最大值、

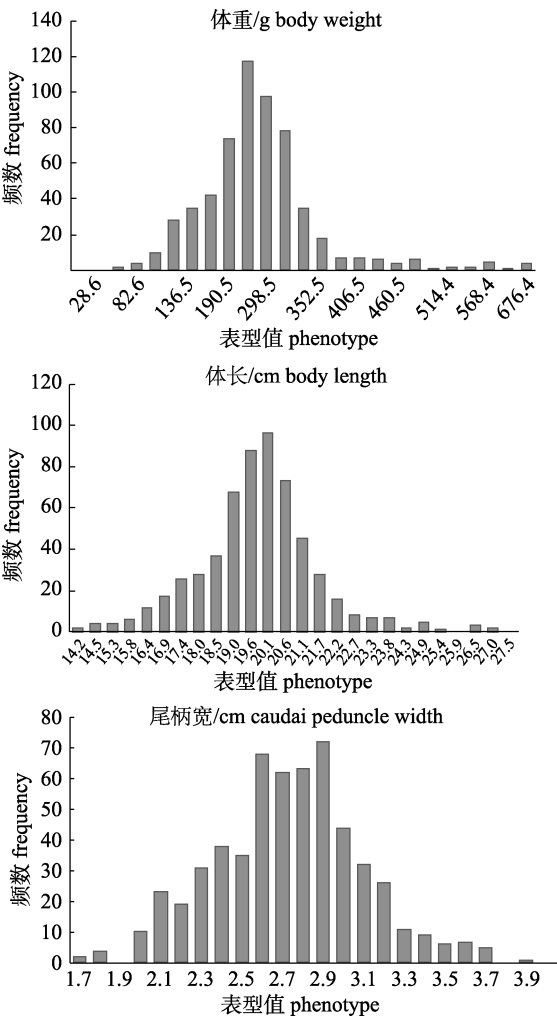


图 1 大菱鲆体重和体尺性状的的正态分布图
Fig. 1 Normal distribution of body mass and morphological traits of turbot

表 1 大菱鲆体重和体尺性状的描述性统计
Tab. 1 Descriptive statistics of body mass and morphological traits in turbot

参数 parameter	体重/g body mass	体长/cm body length	尾柄宽/cm caudal peduncle width
最大值 maximum	676.40	27.01	3.90
最小值 minimum	74.40	13.69	1.58
平均值 mean	274.94	19.51	2.67
标准误 SE	88.69	1.85	0.37
变异系数 VC	0.32	0.10	0.14
遗传力 heritability	0.48	0.81	0.18

最小值、平均值、标准误差、变异系数的统计分析和遗传力估计参数。表 2 为去除性别、池子、测量日期等固定效应后的表型值进行了性状间的协方差分析,发现所研究的 3 个大菱鲂生长性状具有较强的相关性。

表 2 大菱鲂体重和体尺性状间的相关
Tab. 2 Degree of correlation for body mass and morphological traits in turbot

相关系数 correlation coefficient	体重 body weight	体长 body length	尾柄宽 caudal peduncle width
体重 body weight	1	0.844	0.834
体长 body length	0.844	1	0.805
尾柄宽 caudal peduncle width	0.834	0.805	1

2.2 GWAS 结果分析

在全基因组关联分析中,通常使用曼哈顿图和 QQ 图来反映所分析的所有标记的显著与否以及假阳性和假阴性的控制,图 2 显示,585 个个体,21993 个 SNP 大菱鲂体重和体尺性状数据进行 GEMMA 和表型正交化分析得到的基因定位结果。

由图 2 可知 GEMMA 方法超过基因组显著水平(5% Bonferroni 校正阈值, 2.273451×10^{-6})的 QTNs 共 11 个位于 1、5、6、8、12、16、20、21、22 号染色体上,而表型正交化方法共检测到 14 个 QTNs 位于 1、3、5、6、8、12、16、20、21、22 号染色体上。其中与 GEMMA 方法分别在 1(SNP_1_31688963), 5(SNP_1_19132118), 8(SNP_8_14134711), 12(SNP_12_14910921), 16(SNP_16_11311257), 20(SNP_20_18773114), 21(SNP_21_808333), 22(SNP_22_6190492)号染色体上共检测到相同的 QTNs 有 9 个(表 3)。表 4 中记录了 GEMMA 和表型正交化方法在分析不同性状个数时的运算时间。结果表明,在相同的内存条件下,表型正交化方法的运算时间远远低于 GEMMA,具有更高的关联检测效率。

3 讨论

动植物生长性状的研究分析能够直接有助于生产实践,因此这些生长性状是重要的经济性状,探索与大菱鲂重要经济性状关联的 QTNs、候选基因,对解析水产动物的遗传机制,提高基因辅助育种效率有着重要意义。随着二代测序技术的

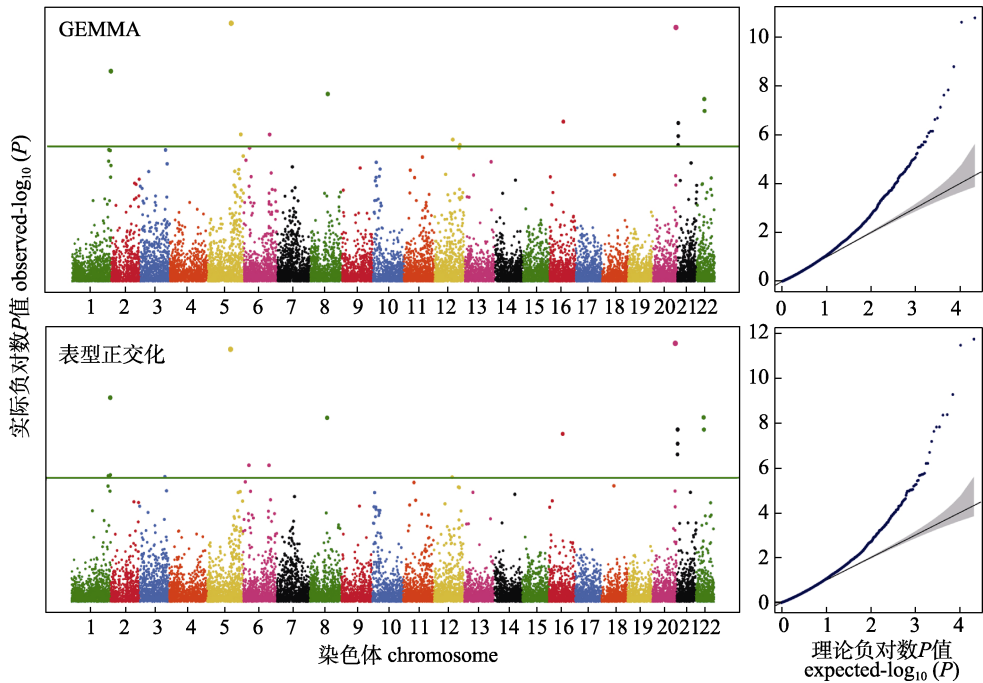


图 2 大菱鲂体重与体尺单一性状 GWAS 曼哈顿和 QQ 图
Fig. 2 Manhatta and QQ plots of single-trait GWAS on body mass and morphological traits in turbot

表 3 大菱鲆体重和体尺性状显著关联的 QTNs 信息

Tab. 3 Information of QTNs significantly associated with body mass and morphological traits in turbot

性状 trait	SNP 位点 SNP ID	等位基因 allele	染色体 Chromosome	位置 position	$-\log_{10}P$	附近基因 nearby gene	区域/bp region
GEMMA							
BM-BL-CPW	SNP_1_31688963	C/A	1	31688963	8.788	Zgc:77056	29724588-30027306
	SNP_5_19132118	G/A	5	19132118	10.787	novel gene*	19137312-19237777
	SNP_5_27008012	A/G	5	27008012	6.143	novel gene*	27022392-27082668
	SNP_6_21034563	T/G	6	21034563	6.141	novel gene*	20953184-21013608
	SNP_8_14134711	G/C	8	14134711	7.834	novel gene*	14121245-14133456
	SNP_12_14910921	G/T	12	14910921	5.930	Slc1a6	14903868-14913454
	SNP_12_20732502	G/A	12	20732502	5.710	lpp	20675041-20804711
	SNP_16_11311257	C/T	16	11311257	6.679	novel gene*	11336381-11348325
	SNP_20_18773114	G/A	20	18773114	10.610	Myt1la	18751768-18792541
	SNP_21_808333	T/C	21	808333	6.623	Zgc112356	804720-812315
	SNP_22_6190492	T/C	21	619492	7.620	novel gene*	617712-621986
Fast-LMM							
	SNP_1_30018909	G/A	1	31688963	5.737	VPS13B	29724588-30027306
	SNP_1_31036272	G/A	1	31688963	5.723	novel gene*	31063130-31066902
	SNP_1_31688963	C/A	1	31688963	9.300	Zgc:77056	31688749- 31690775
	SNP_1_31719031	C/A	1	31719031	5.781	novel gene*	31717594-31718666
	SNP_3_20633231	C/A	3	20633231	5.701	novel gene*	20633231
	SNP_5_19132118	G/A	5	19132118	11.502	novel gene*	19137312-19237777
	SNP_6_4721035	T/G	6	4721035	6.223	novel gene*	4808453-4971126
	SNP_6_21034563	A/G	6	21034563	6.220	novel gene*	20953184-21013608
	SNP_8_14134711	G/C	8	14134711	8.380	novel gene*	14121245-14133456
	SNP_12_14910921	G/T	12	14910921	5.679	Slc1a6	14903868-14913454
	SNP_16_11311257	C/T	16	11311257	7.653	novel gene*	11336381-11348325
	SNP_20_18773114	G/A	20	18773114	11.773	Myt1la	18751768-18792541
	SNP_21_808333	T/C	21	808333	7.850	Zgc112356	804720-812315
	SNP_22_6190492	T/C	22	619492	8.402	novel gene*	617712-621986

注: *表示未在大菱鲆全基因组中进行功能注释的新基因。
Note: * represents novel gene which has not found any gene function annotation in the genome of *Scophthalmus maximus*.

表 4 GEMMA 和表型正交化方法多性状
关联检验运算时间的比较

Tab. 4 Comparison in runtimes (min) between
GEMMA and phenotypic principal component
methods for association tests

方法 method	性状个数 number of traits				
	2	3	4	5	6
GEMMA	0.94	2.11	4.44	9.70	32.74
表型正交化 phenotypic	0.19	0.28	0.36	0.49	0.57

迅猛发展, 全基因组测序成本大大降低, 有关水产动物经济性状的 GWAS 研究逐渐增多。

Gutierrez 等^[13]使用混合模型方法发现了与大西洋鲑性成熟时生长速率、年龄相关的主效 QTN 位点。Gonzalez-Pena 等^[14]使用单变量模型对虹鳟 10、13 月龄的体重进行 GWAS 分析, 检测到虹鳟在 omy5 染色体上检测到一个 SNP 位点可分别解释表型 1.4%和 1%变异。Jiang 等^[15]使用 QTL-seq 的方法在距 LG18 染色体 23.0Mb 的位置上定位到与罗非鱼耐盐性相关的 QTNs, 揭示了罗非鱼的耐盐性渗透调节遗传机制。目前, 有关大菱鲆经济性状的定位分析大都采用 QTL 定位的方法, Sánchez-Molano 等^[16]通过 QTL 筛选和微卫星搜

索发现了与生长性状相关的共 11 个 QTL。

大多数动植物 GWAS 研究主要针对单个性状进行 QTN 定位分析, 有关水产动物生长性状的多性状 GWAS 研究尚处在初步阶段。Bolormaa 等^[17]使用主成分分析和一系列双变量分析对奶牛状进行了多性状 GWAS, 发现在统计错误率没有增加的情况下, 多性状 GWAS 比单性状 GWAS 具有更高的统计效力。到目前为止, 基于表型正交化的大菱鲆多性状 GWAS 还未见报道, 我们的研究通过对相关表型正交化, 使得多变量混合模型关联检验变成独立的单变量检验, 与 GEMMA 相比, 在相同的实验群体下表型正交化方法通过对性状的单独关联检验不但增加了所分析性状个数, 且保证了 QTN 的检测效率, 甚至比 GEMMA 的检测效率更高。这为分析其他水产动物性状的多性状 GWAS 研究提供一个有价值的工具。

在大菱鲆全基因组上寻找距离每个 QTN 最近的基因, 结果找到 *VPS13B*、*Zgc:77056*、*Slc1a6*、*Myt1la*、*Plxdc2*、*Zgc112356* 共 6 个已知候选基因。相关研究表明, *Vps13B* 是顶体生物发生所必需的^[18]。*Slc1a6* 是转运体基, 因为代谢提供氨基酸, 保持细胞外氨基酸水平, 具有高的非耦合阴离子电导作用^[19]。*Myt1la* 髓磷脂转录因子是一种神经特异性的含锌指的 DNA 结合蛋白, 在神经系统的发育中发挥作用^[20]。*Plxdc2* 是 I 型跨膜蛋白, 与巢蛋白和丛状蛋白有一定的同源性, 它在发育中的神经系统中以高度离散和动态的模式表达, 在各种模式中心有显著的表达^[21]。这些候选基因可能在大菱鲆生长发育过程中发挥着重要的作用, 它们的筛选定位极大地提高了对大菱鲆遗传结构的认识。

参考文献:

- [1] Korte A, Vilhjálmsson B J, Segura V, et al. A mixed-model approach for genome-wide association studies of correlated traits in structured populations[J]. *Nature Genetics*, 2012, 44(9): 1066-1071.
- [2] Ferreira M A R, Purcell S M. A multivariate test of association[J]. *Bioinformatics (Oxford, England)*, 2009, 25(1): 132-133.
- [3] Korol A B, Ronin Y I, Itskovich A M, et al. Enhanced efficiency of quantitative trait loci mapping analysis based on multivariate complexes of quantitative traits[J]. *Genetics*, 2001, 157(4): 1789-1803.
- [4] Stephens M. A unified framework for association analysis with multiple related phenotypes[J]. *PLoS One*, 2013, 8(7): e65245.
- [5] Price A L, Helgason A, Thorleifsson G, et al. Single-tissue and cross-tissue heritability of gene expression via identity-by-descent in related or unrelated individuals[J]. *PLoS Genetics*, 2011, 7(2): e1001317.
- [6] Amos C I. Robust variance-components approach for assessing genetic linkage in pedigrees[J]. *American Journal of Human Genetics*, 1994, 54(3): 535-543.
- [7] Lee S H, Yang J, Goddard M E, et al. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood[J]. *Bioinformatics (Oxford, England)*, 2012, 28(19): 2540-2542.
- [8] Trzaskowski M, Yang J, Visscher P M, et al. DNA evidence for strong genetic stability and increasing heritability of intelligence from age 7 to 12[J]. *Molecular Psychiatry*, 2014, 19(3): 380-384.
- [9] Vattikuti S, Guo J E, Chow C C. Heritability and genetic correlations explained by common SNPs for metabolic syndrome traits[J]. *PLoS Genetics*, 2012, 8(3): e1002637.
- [10] Gao H, Wu Y, Zhang T, et al. Multiple-trait genome-wide association study based on principal component analysis for residual covariance matrix[J]. *Heredity*, 2015, 114(4): 428.
- [11] Blanquer A, Alayse J P, Berrada-Rkhami O, et al. Allozyme variation in turbot (*Psetta maxima*) and brill (*Scophthalmus rhombus*) (Osteichthyes, Pleuronectoformes, Scophthalmidae) throughout their range in Europe[J]. *Journal of Fish Biology*, 1992, 41(5): 725-736.
- [12] Lei J L, Liu X F, Guan C T. Turbot culture in China for two decades: Achievements and prospect[J]. *Progress in Fishery Sciences*, 2012, 33(4): 123-130. [雷霖霖, 刘新富, 关涛. 中国大菱鲆养殖 20 年成就和展望——庆祝大菱鲆引进中国 20 周年[J]. *渔业科学进展*, 2012, 33(4): 123-130.]
- [13] Gutierrez A P, Yáñez J M, Fukui S, et al. Genome-wide association study (GWAS) for growth rate and age at sexual maturation in Atlantic salmon (*Salmo salar*)[J]. *PLoS One*, 2015, 10(3): e0119730.
- [14] Gonzalez-Pena D, Gao G T, 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]. *Frontiers in Genetics*, 2016, 7: 203.
- [15] Jiang D L, Gu X H, Li B J, et al. Identifying a long QTL cluster across chrLG18 associated with salt tolerance in tilapia using GWAS and QTL-seq[J]. *Marine Biotechnology*,

- 2019, 21(2): 250-261.
- [16] Sánchez-Molano E, Cerna A, Toro M A, et al. Detection of growth-related QTL in turbot (*Scophthalmus maximus*)[J]. BMC Genomics, 2011, 12: 473.
- [17] Bolormaa S, Pryce J E, Hayes B J, et al. Multivariate analysis of a genome-wide association study in dairy cattle[J]. Journal of Dairy Science, 2010, 93(8): 3818-3833.
- [18] Da Costa R, Bordessoules M, Guilleman M, et al. Vps13b is required for acrosome biogenesis through functions in Golgi dynamic and membrane trafficking[J]. Cellular and Molecular Life Sciences, 2020, 77(3): 511-529.
- [19] Zhou Y, Danbolt N C. GABA and glutamate transporters in brain[J]. Frontiers in Endocrinology, 2013, 4: 165.
- [20] Ti W, Zhen Z, Tao L, et al. Common SNPs in myelin transcription factor 1-like (MYT1L): Association with major depressive disorder in the Chinese Han population[J]. PLoS One, 2010, 5(10): e13662.
- [21] Miller-Delaney S F C, Lieberam I, Murphy P, et al. Plxdc2 is a mitogen for neural progenitors[J]. PLoS One, 2011, 6(1): e14565.

Multivariate mixed model association analysis for growth trait orthogonalization in turbot (*Scophthalmus maximus*)

JIANG Yi^{1,3}, SONG Yuxin², GAO Jin³, CHEN Heli³, YANG Runqing³

1. National Demonstration Center for Experimental Fisheries Science Education, Shanghai Ocean University, Shanghai 201306, China;
2. Wuxi Fisheries College, Nanjing Agricultural University, Wuxi 214081, China;
3. Research Centre for Aquatic Biotechnology, Chinese Academy of Fishery Sciences, Beijing 100141, China

Abstract: The multivariate linear model is a powerful tool for examining associations between a single nucleotide polymorphism and multiple related phenotypes. To improve the computational efficiency of this complex correlation analysis, we spectrum-transformed multiple correlated phenotypes to mutually independent “super traits” by using the mixed model association analysis method of a single trait. The sum of these statistics was used to infer pleiotropic nucleotide genetic loci that control multiple related phenotypes. The multivariate linear mixing model of GEMMA software and the phenotypic orthogonalization method proposed in this study were used to analyze the genome-wide associations (GWASs) of three traits in turbot (*Scophthalmus maximus*): body length, body mass, and caudal peduncle width. The results showed that 11 QTNs were detected on chromosome 1, 5, 6, 8, 12, 16, 20, 21, and 22 by the multivariable linear mixed model method. However, our method detected a total of 14 QTNs on chromosome 1, 3, 5, 6, 8, 12, 16, 20, 21, and 22, among which nine QTNs were detected jointly by the two methods. Compared with that of GEMMA, our method exhibited higher power to detect QTNs of multi related traits, which provides a convenient and efficient strategy for multi-trait GWAS research and the genetic breeding of other aquatic species.

Key words: *Scophthalmus maximus*; multivariate mixed model; spectral factorization; growth traits; phenotypic orthogonalization

Corresponding author: YANG Runqing. E-mail: runqingyang@cafs.ac.cn