

DOI: 10.19663/j.issn2095-9869.20200202001

http://www.yykxjz.cn/

高进, 杨润清. 大菱鲆体重和体尺性状联合 GWAS 分析. 渔业科学进展, 2021, 42(2): 63–70

Gao J, Yang RQ. Joint genome-wide association study of body mass and morphological traits in turbot (*Scophthalmus maximus*). Progress in Fishery Sciences, 2021, 42(2): 63–70

大菱鲆体重和体尺性状联合 GWAS 分析*

高进^{1,2} 杨润清^{1,2①}

(1. 南京农业大学无锡渔业学院 无锡 214081; 2. 中国水产科学研究院生物技术研究中心 北京 100141)

摘要 为了揭示大菱鲆(*Scophthalmus maximus*)体重和体尺性状的分子遗传机制,探寻用于改良目标性状的分子标记及候选基因,本研究以大菱鲆育种群体为研究对象,分别测量其体重、体长、体宽和尾柄宽性状的表型值,利用简化基因组测序技术(2b-RAD)获得相应基因型数据,进行全基因组关联研究(Genome-wide association study, GWAS),筛选与大菱鲆体重和体尺性状显著关联的数量性状核苷酸(Quantitative trait nucleotides, QTNs)遗传位点。结果显示,以多性状线性混合模型(mvLMM)对体重-体长和体长-体宽-尾柄宽 2 个性状组合进行多性状 GWAS 分析,分别检测到 9 个和 2 个一因多效 QTNs;以单一性状线性混合模型(LMM)对各个性状进行 GWAS 分析,在体重性状中检测到 4 个与之显著关联的 QTNs,在体长和体宽性状中各检测到 1 个 QTN,而在尾柄宽性状中则没有检测到显著的遗传位点。比较 2 种模型的结果,发现 mvLMM 相较于 LMM 能够检测到更多 QTNs,且检测到的 QTNs 为更具生物学意义的一因多效 QTNs。本研究首次利用 mvLMM 和 LMM 对大菱鲆体重和体尺性状进行联合 GWAS 分析,共筛选到 17 个显著的 QTNs,其中有 4 个 QTNs 被重复检测到。以这些检测到的 QTNs 为探针,在大菱鲆全基因组上找到了距离其最近的 12 个候选基因,它们可能是影响大菱鲆体重和体尺性状的重要候选标记和功能基因,本研究为大菱鲆体重和体尺性状的分子标记辅助选育提供了理论素材和参考。

关键词 大菱鲆; 体重和体尺性状; 全基因组关联分析; 多性状线性混合模型

中图分类号 S917.4 **文献标识码** A **文章编号** 2095-9869(2021)02-0063-08

大菱鲆(*Scophthalmus maximus*), 俗称“欧洲比目鱼”或“多宝鱼”, 是中国北部沿海地区重要的海水养殖经济鱼类。大菱鲆具有生长速度快、肉质鲜美以及抗逆性强等诸多优点, 其养殖区域迅速扩展、产量产值不断攀升(雷霖霖等, 2002、2005)。然而, 随着工厂化养殖规模和集约化程度的扩大, 特别是种质资源的退化, 成活率低、病害严重和生产缓慢等问题逐渐

显现(李杰等, 2019; 王岚等, 2017)。因此, 开展大菱鲆相关经济性状的选育研究, 培育具有强抗逆性和高生长性能的新品种对大菱鲆产业发展具有重要意义。传统的育种方法是基于目标性状的表型值来开展选育工作(Hulata, 2001), 对于高生长性能品种选育, 通常挑选符合标准的个体进行群体混养或是建立多个家系并选择优良的家系作为亲本用于进一步选育。然

*中国水产科学研究院基本科研业务费项目“鲆鲽鱼分子标记辅助配套选育技术”(2017A001)资助 [This work was supported by the Special Scientific Research Funds for Central Non-Profit Institutes, Chinese Academy of Fishery Sciences (2017A001)]. 高进, E-mail: gaojin427@126.com

① 通讯作者: 杨润清, 研究员, E-mail: runqingyang@cafs.ac.cn

收稿日期: 2020-02-02, 收修改稿日期: 2020-02-19

而,分子生物学和基因组学的快速发展则促使鱼类育种研究由传统选育和杂交育种向基于基因组信息的分子育种转变(Xu *et al.*, 2015),如以基因定位和数量性状位点(Quantitative trait locus, QTL)作图为基础的分子标记辅助选择(Marker-assisted selection, MAS)育种(Wang *et al.*, 2015; 刘晓菲等, 2019)。利用目标性状与分子标记间的显著关联性,依据标记对后代个体进行全基因组选择或是综合标记与表型信息构建选择指数筛选个体,加速育种过程,提高选育效率。

随着测序成本的下降与高通量基因分型技术的迅猛发展,基于海量单核苷酸多态性(Single-nucleotide polymorphisms, SNP)标记的全基因组关联分析(Genome-wide association studies, GWAS)已取代利用稀疏遗传标记的 QTL 定位,成为鉴定与目标性状显著关联的遗传变异的主要方法。早期 GWAS 研究主要针对单一性状表型(Yu *et al.*, 2006; Burton *et al.*, 2007),而在进行复杂性状的遗传解析且记录有多个相关性状表型值时,考虑多个性状联合信息的多性状 GWAS 策略通常要优于逐个性状与遗传位点间的关联分析(Kim *et al.*, 2009; Bolormaa *et al.*, 2010)。相较单一性状 GWAS 分析,表型具有潜在相关性的多性状 GWAS 分析更具优势,主要是因为:当不同性状间存在遗传相关时,因考虑了单一性状分析中忽略的性状间协方差,多性状分析提高了检验功效(Stephens, 2013)和参数估计的精度(Zhu *et al.*, 2009);在多性状 GWAS 分析过程中,数量性状核苷酸(Quantitative trait nucleotides, QTNs)的统计推断只进行一次统计检验,与逐个性状单独分析相比,降低了多重检验造成的误差(Klei *et al.*, 2008);存在基因多效性时,单个遗传变异与多个性状相关联,使得多性状 GWAS 分析更具生物学意义(Chavali *et al.*, 2010);Porter 等(2017)综合比较了当下流行的多性状 GWAS 分析方法,发现大多数现存的多性状 GWAS 分析方法具有明显相似的统计检验功效,与单一性状 GWAS 相比,能够大幅增加显著遗传变异的检测效率。

本研究采用实验成本较低的简化基因组(2b-RAD)测序技术进行 SNP 分型,基于多性状线性混合模型对体重-体长和体长-体宽-尾柄宽共 2 种多性状相关表型组合进行全基因组关联分析,关联定位控制这些性状的一因多效 QTNs,为大菱鲂的体重和体尺性状改良提供理论基础。此外,对逐个性状进行了单一性状线性混合模型 GWAS 分析,并将其结果与多性状 GWAS 分析结果进行比较。

1 材料与方法

1.1 实验群体及表型测量

实验用鱼共 585 条,取自 29 个全同胞家系组成的大菱鲂育种群体,4 月龄时,对群体中所有大菱鲂个体注射电子标记,并提取鳍条组织 DNA 用于后续的 2b-RAD 测序。随后,将标记后的大菱鲂随机混养在 2 个 6 m×6 m×1.5 m 具有循环水养殖系统的水泥池中,控制水温在 5℃~24℃,每天定时投喂商用饵料 2 次至饱食。实验用鱼按孵化日期至表型测量时的生长天数为 275~1001 d 不等。对实验用鱼进行 11 次不定期表型测量,每次测量前利用 50 mg/ml MS-222 鱼用安定剂麻醉待测大菱鲂个体,避免应激反应造成的鱼体损伤。利用电子秤称量每个个体的体重(Body weight, g, BM)性状,同时在统一的参考标尺下,用数码相机由固定高度向下垂直拍摄相应个体的体尺性状。根据拍摄图形,利用 ImageJ 软件标定每条大菱鲂的体长(Body length, cm, BL)、体宽(Body width, cm, BW)和尾柄宽(Caudal peduncle width, cm, CPW)共 3 个体尺性状的表型值。采用简化基因组 2b-RAD 高通量标记分型技术对具有表型测量记录的个体进行 SNP 分型,共获得 30049 个多态 SNP 分子标记,参考大菱鲂基因组(ASM318616v1)建立多态标记物理图谱。挑选生长周期为 473 d 左右(前后皆不超过 5 d)大菱鲂个体的表型观测值作为多性状 GWAS 分析的表型值,使用 PLINK v1.9 (<http://www.cog-genomics.org/plink2>)对相应样本个体的基因型数据进行质量控制。剔除低于 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 441 个样本的 23988 个 SNP 标记用于全基因组关联分析。

1.2 关联统计分析

本研究使用 GEMMA 软件(Zhou *et al.*, 2014)中的多性状线性混合模型(Multivariate linear mixed model, mvLMM)进行多个性状表型值和 SNPs 标记的全基因组关联分析,所用模型为:

$$Y = WA + x\beta + G + E$$

式中, Y 为 $n \times t$ 维表型矩阵, n 是样本个数, t 是分析性状个数; W 为 $n \times c$ 维协变量(非遗传固定效应,如性别、年龄等)矩阵, A 为 c 维相应系数行向量, c 是包含截距项在内的协变量个数; x 为 n 维当前检验

SNP 的基因型指示变量列向量, β 为 t 维当前检验 SNP 的加性遗传效应行向量; G 为 $n \times t$ 维不包括当前检验 SNP 的剩余多基因效应矩阵, $G \sim MN(0, V_g \otimes K)$, K 是由全基因组 SNP 标记构建的基因组亲缘关系矩阵, V_g 是 $t \times t$ 维剩余多基因方差-协方差矩阵; E 为 $n \times t$ 维误差项矩阵, $E \sim MN(0, V_e \otimes I)$, V_e 是 $t \times t$ 维误差方差-协方差矩阵, I 是单位矩阵。

最终, 全基因组上每个 SNP 都能得到 1 个一因多效 Wald 统计量和相对应的统计概率 P 值。此外, 还利用因子谱分解线性混合模型(Factored spectrally transformed linear mixed model, FaST-LMM) (Lippert *et al.*, 2011) 逐个性状进行单一性状的 GWAS 分析, 检测与各个性状相关联的 QTNs。使用 R 语言绘制曼哈顿和 Quantile-Quantile (QQ) 图, 同时, 统计用于判别群体分层影响大小的膨胀系数或称基因组控制 (Genomic control, GC) 值, 在实际研究中 GC 值被定义为所有 SNP 标记的卡方统计量均值 (Price *et al.*, 2010)。

1.3 遗传参数估计和基因注释

将质量控制后的基因型数据用 PLINK 软件 (Chang *et al.*, 2015) 处理为分析所需的格式, 再利用单性状约束最大法估计法 (Restricted maximum likelihood,

REML) 逐个估计各性状的遗传参数。查找关联分析得到的显著 SNP 位点在全基因组上所处的物理位置, 于大菱鲂全基因组 (GCA_003186165.1) 上选择距离其最近的候选基因进行注释分析。

2 结果

2.1 表型值的描述性统计

选取生长天数为 473 d 左右的 441 条大菱鲂的体重、体长、体宽和尾柄宽 4 个性状进行联合 GWAS 分析。4 个性状的表型频数分布见图 1。从图 1 可以看出, 各个性状的表型值基本都符合正态分布, 具有一定的可靠性, 适合后续的 GWAS 分析。各性状的原始表型的统计分析和遗传力估计值见表 1。此外, 还对原始以及校正了性别、池子和测定日期等非遗传固定效应后的表型值进行了性状间的相关分析, 发现各性状表型值间具有较强的相关性 (表 2)。

2.2 GWAS 分析结果

根据质量控制结果, 441 条大菱鲂的 23988 个 SNPs 被用于体重和体尺性状的联合 GWAS 分析, 分别进行了 BM-BL 和 BL-BW-CPW 两个性状组合的多

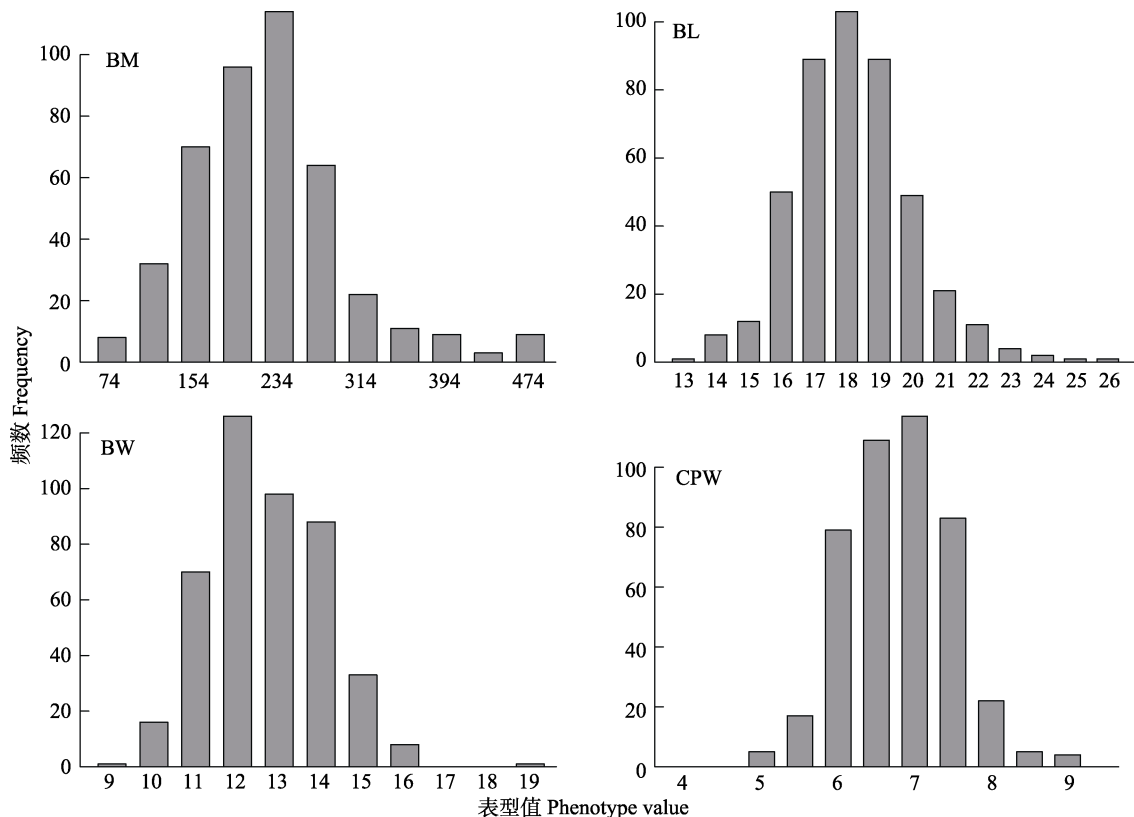


图 1 4 个性状的频数分布

Fig.1 Frequency distribution histograms of four traits

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

参数 Parameter	性状 Trait			
	体重 Body weight (g)	体长 Body length (cm)	体宽 Body width (cm)	尾柄宽 Caudal peduncle width (cm)
最大值 Maximum	541.60	26.71	9.16	3.62
最小值 Minimum	74.40	13.69	4.45	1.58
平均值 Mean	245.72	18.69	6.53	2.53
标准误差 SE	1.77×10^{-1}	1.04×10^{-3}	1.51×10^{-3}	7.78×10^{-4}
变异系数 VC	0.32	0.10	0.10	0.14
遗传力 Heritability	0.72	0.58	0.59	0.48

性状线性混合模型以及逐个性状的单一性状线性混合模型 GWAS 分析。对 BM-BL 两个性状的多性状 GWAS 分析后(图 2 和表 3)发现, 超过基因组显著水平(5% Bonferroni 校正阈值, 即 2.084×10^{-6})的 QTN 位点共 9 个, 其中, 各有 1 个显著 QTN 位于 1、10 和 20 号染色体上, 即 SNP_1_31495825、SNP_10_6893888 和 SNP_20_18773114, 分别有 2 个显著 QTNs 位于 3 (SNP_3_21601589 与 SNP_3_24007357)、5(SNP_5_25635891 与 SNP_5_26888833)和 22(SNP_22_6190492 与 SNP_22_6380029)号染色体上。而在相同显著水准下, BL-BW-CPW 3 个体尺性状的 GWAS 分

表 2 大菱鲆体重和体尺性状相关程度
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	体宽 Body width	尾柄宽 Caudal peduncle width
体重 Body weight	1	0.759	0.801	0.662
体长 Body length	0.834	1	0.825	0.772
体宽 Body width	0.807	0.832	1	0.729
尾柄宽 Caudal peduncle width	0.784	0.842	0.787	1

注: 下三角中的值为性状原始表型间相关程度, 上三角中的值为校正了非遗传固定效应后表型间的相关程度

Note: Values in lower-triangular matrix represent degrees of correlation among original phenotypes, and values in up-triangular matrix mean degrees of correlation among phenotypes which were adjusted by non-genetic fixed effects

析则在 10 号染色体上检测到 2 个显著的 QTN 位点 SNP_10_2724296 和 SNP_10_6893888。2 组多性状 GWAS 分析的 GC 值分别为 BM-BL(1.035)和 BL-BW-CPW(1.044)。

对体重和体尺的逐个性状单一性状 GWAS 分析结果(图 3 和表 3)显示, 在 BM 性状中, 于 3 和 5 号染色体上各检测到 1 个显著的 QTN 位点, 即 SNP_3_21601589 和 SNP_5_26888833, 10 号染色体上检测到 2 个显著的 QTN 位点 SNP_10_2724296 和 SNP_10_6893888。BL 和 BW 性状分别在 62 和 8 号

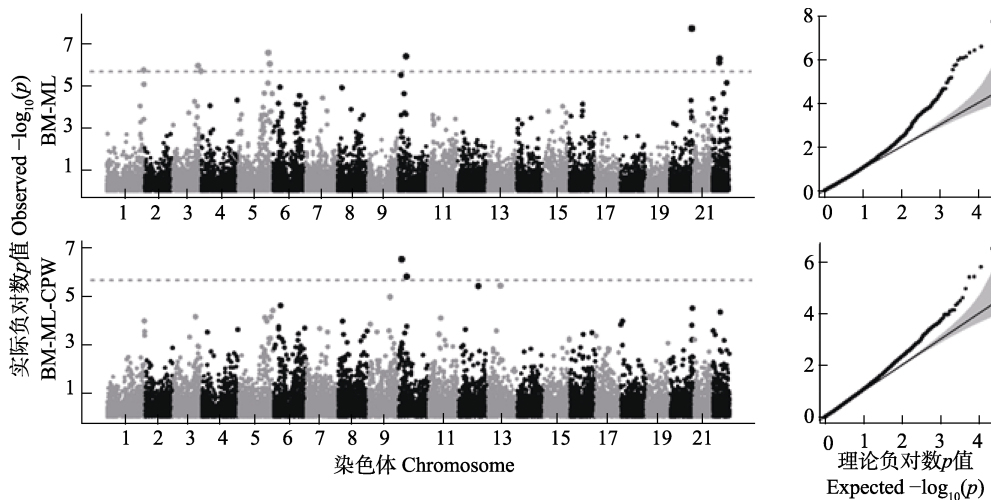


图 2 大菱鲆体重与体尺多性状 GWAS 曼哈顿和 QQ 图
Fig.2 Manhattan and Q-Q plots of multi-trait GWAS on body mass and morphological traits in turbot

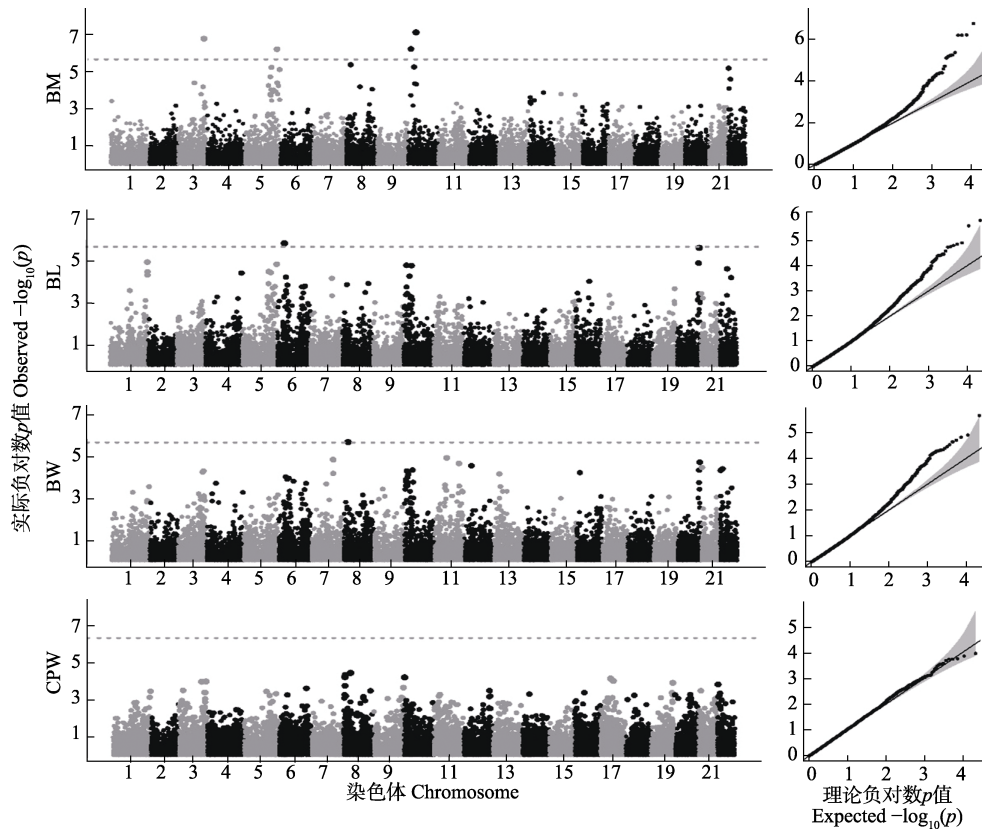


图 3 大菱鲂体重与体尺单一性状 GWAS 曼哈顿和 QQ 图

Fig.3 Manhattan and Q-Q plots of single-trait GWAS on body mass and morphological traits in turbot

表 3 大菱鲂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 Chr	位置 Position	P 值 P value	附近基因 Nearby gene	区域 Region (bp)
多性状线性混合模型 Multivariate Linear Mixed Models							
BM-BL	SNP_1_31495825	G/A	1	31,495,825	1.78×10^{-6}	GRIK2	31268951~31431331
	SNP_3_21601589	C/G	3	21,601,589	1.09×10^{-6}	cyp21a2	21601235~21604716
	SNP_3_24007357	G/T	3	24,007,357	1.96×10^{-6}	novel gene*	23968869~23970083
	SNP_5_25635891	A/G	5	25,635,891	2.64×10^{-7}	tshz3b	25624917~25661383
	SNP_5_26888833	C/T	5	26,888,833	8.85×10^{-7}	adamts17	26823608~26965578
	SNP_10_6893888	A/G	10	6,893,888	3.91×10^{-7}	SYT8	6878968~6899153
	SNP_20_18773114	G/A	20	18,773,114	1.88×10^{-8}	myt11a	18751768~18792541
	SNP_22_6190492	T/C	22	6,190,492	7.73×10^{-7}	psmd4a	6191068~6196105
	SNP_22_6380029	T/C	22	6,380,029	4.98×10^{-7}	novel gene*	6365881~6382399
BL-BW-CPW	SNP_10_2724296	A/C	10	2,724,296	2.83×10^{-7}	novel gene*	2717495~2721226
	SNP_10_6893888	A/G	10	6,893,888	1.48×10^{-6}	SYT8	6878968~6899153
单性状线性混合模型 Single-trait Linear Mixed Model							
BM	SNP_3_21601589	C/G	3	21,601,589	1.61×10^{-7}	cyp21a2	21540626~21547626
	SNP_5_26888833	C/T	5	26,888,833	5.94×10^{-7}	adamts17	26823608~26965578
	SNP_10_2724296	A/C	10	2,724,296	5.75×10^{-7}	novel gene*	2717495~2721226
	SNP_10_6893888	A/G	10	6,893,888	7.32×10^{-8}	SYT8	6878968~6899153
BL	SNP_6_6047375	G/A	6	6,047,375	1.42×10^{-6}	grm4	5928125~6045591
BW	SNP_8_3870447	G/A	8	3,870,447	1.98×10^{-6}	BFAR	3863424~3871475
CPW	-	-	-	-	-	-	-

注: *表示未在大菱鲂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 genome of *S. maximus*

染色体上检测到 1 个显著的 QTN 位点 SNP_6_6047375 和 SNP_8_3870447。而在 CPW 性状中则没有检测到显著的遗传变异位点。各性状单一性状 GWAS 分析的 GC 值分别为 1.026 (BM)、1.021 (BL)、1.034 (BW) 和 1.031 (CPW)。

2.3 候选基因

以大菱鲃多性状与单一性状 GWAS 获得的显著 SNP 位点为探针, 根据其在大菱鲃全基因组的位置, 向上下游寻找距离最近的候选基因。在本研究检测到的 17 个 QTNs 附近共找到 12 个候选基因(表 3), 分别为 GRIK2 (Glutamate receptor, ionotropic, kainate 2)、cyp21a2 (Cytochrome P450, family 21, subfamily A, polypeptide 2)、tshz3b (Teashirt zinc finger homeobox 3b)、adamts17 (ADAM metalloproteinase with thrombospondin type 1 motif, 17)、SYT8 (Synaptotagmin 8)、myt11a (Myelin transcription factor 1-like, a)、psmd4a (Proteasome 26S subunit, non-ATPase 4a)、grm4 (Glutamate receptor, metabotropic 4)、BFAR (Bifunctional apoptosis regulator) 和 3 个新基因。对这些候选基因进行 GO 注释分析发现, 细胞组分主要涉及核膜、细胞质膜及细胞核等; 分子功能主要涉及离子通道活性、激素活性、氧化还原酶活性、肽酶活性、核苷酸结合、蛋白结合及 G 蛋白偶联受体活性等; 而生物学进程则主要涉及离子转运、糖皮质激素生物合成过程、氧化还原过程、细胞信号转导、基因表达调控、蛋白酶解、胞外分泌、转录调控和 G 蛋白偶联受体信号通路等。其中, 3 个新基因在大菱鲃基因组中未有相关注释信息。

3 讨论

GWAS 作为一种定位影响重要经济性状分子标记的有效方法, 是以全基因组 SNPs 直接关联复杂性状表型, 广泛应用于动植物分子标记开发和辅助选择育种等方面(Santana *et al.*, 2015; He *et al.*, 2017)。随着诸多水产动物基因组测序工作相继完成以及测序成本的下降, 水产动物生长和抗病等复杂性状的 GWAS 分析多有研究报道。Gutierrez 等(2015)对大西洋鲑(*Salmo salar*) 在性成熟时的生长速率性状进行 GWAS 分析, 检测到多个在大西洋鲑代谢和生长发育过程中发挥重要作用的候选基因。Zhou 等(2019)利用 GWAS 揭示了半滑舌鳎(*Cynoglossus semilaevis*)抗病的遗传机制, 共定位了与抗病性状显著关联的 33 个 SNP 位点, 结合基因表达和甲基化分析, 检测到数个影响半滑舌鳎抗病力的候选基因。Jiang 等(2019)则对罗非鱼

(*Tilapia*)的耐盐性状进行 GWAS 和 QTL 定位研究, 为罗非鱼耐盐性遗传机理揭示及进一步的功能分析奠定了基础。目前, 大菱鲃生长和体尺相关性状的 GWAS 分析鲜有应用, 但 QTL 定位研究已在大菱鲃生长(Enrique *et al.*, 2011)和气单胞菌(*Aeromonas salmonicida*)耐药性(Silvia *et al.*, 2011)等相关性状中开展。

现有的绝大多数 GWAS 研究都是关联全基因组上 SNP 位点与单个性状表型值, 即使当多个相关表型性状存在时, 往往也会选择逐个性状分别进行单一性状 GWAS 分析(Stephens, 2013)。在多个相关表型性状条件下, 单一性状关联定位方法忽略了 QTNs 对多个表型性状的共同影响。相较于单一性状关联分析方法, 多性状 GWAS 方法具有更高的统计功效和参数估计及定位精确度。通过比较大菱鲃体重和体尺多性状 GWAS 与各个性状逐个进行单一性状 GWAS 的定位结果发现, 在相同的显著水准下, 前者检测到的 QTN 数多于后者(表 3)。在 BM-BL 性状组合的多性状 GWAS 分析中, 共检测到 9 个 QTNs, 而在相对应的 BM 和 BL 单一性状 GWAS 分析中, BM 中检测到 4 个 QTNs, BL 中仅定位到 1 个 QTN, 且 BM 中检测到的 4 个 QTNs 中有 3 个与多性状 GWAS 分析得到的 QTNs 重复。同样, BL-BW-CPW 性状组合多性状 GWAS 分析检测到的 QTNs 也多于 3 个性状各自进行单一性状 GWAS 分析得到的 QTNs。除了能够检测到更多的 QTNs, 由于多性状 GWAS 分析检测到的 QTN 是一因多效 QTN, 相较于单一性状 GWAS 分析检测到的 QTN 在生物学和实际应用层面更有意义。

在大菱鲃全基因组上寻找距离每个 QTN 最近的基因, 共找到 12 个候选基因, 其中, 有 9 个已知的功能基因及 3 个新基因。这些影响大菱鲃体重和体尺性状的候选基因在全基因组分布较为分散, 没有明显集中区域。本研究为首次对大菱鲃体重和体尺性状进行全基因组关联分析, 检测到在大菱鲃相关研究中尚未见报道的显著的已知功能基因。相关研究表明, GRIK2 为离子型受体, 它偶联离子通道并形成受体通道复合物, 且与细胞缺血缺氧损伤有关(张冬梅等, 2010)。Eachus 等(2017)和 Weger 等(2018)的研究均表明, cyp21a2 在斑马鱼(*Brachydanio rerio var*)幼鱼的糖皮质激素生物合成中发挥着重要作用。Erickson 等(2011)以斑马鱼为实验动物模型, 发现 tshz3b 可能调控后脑中的 Hox 功能, 而 Hox 功能则被认为与形态发生和器官形成有关联(Gair *et al.*, 2003)。由此可见, 该基因可能参与大菱鲃体形态发育过程。Myt11a 可能在有丝分裂结束阶段发挥重要作用(Nakajima *et al.*,

2008), BFAR 则能与 p75NTR 的蛋白相互作用并抑制 p75NTR 信号转导, 而 p75NTR 可与高亲和力受体 TrkA 协同作用促进细胞增殖或与细胞内配体结合诱导细胞凋亡 (李红梅等, 2011)。本研究找到的候选基因可能都是大菱鲆体重和体尺性状的重要候选功能基因, 它们在大菱鲆生长发育过程中的影响有待进一步的功能实验验证。

参 考 文 献

- Allison DB, Thiel B, St Jean P, *et al.* Multiple phenotype modeling in gene-mapping studies of quantitative traits: Power advantages. *American Journal of Human Genetics*, 1998, 63(4): 1190–1201
- Bolormaa S, Pryce JE, Hayes BJ, *et al.* Multivariate analysis of a genome-wide association study in dairy cattle. *Journal of Dairy Science*, 2010, 93(8): 3818–3833
- Burton PR, Clayton DG, Cardon LR, *et al.* Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 2007, 447(7145): 661–678
- Chang CC, Chow CC, Tellier LC, *et al.* Second-generation PLINK: Rising to the challenge of larger and richer datasets. *GigaScience*, 2015, 4(7): 1–16
- Chavali S, Barrenas F, Kanduri K, *et al.* Network properties of human disease genes with pleiotropic effects. *BMC Systems Biology*, 2010, 4(78): 1–11
- Eachus H, Zaucker A, Oakes JA, *et al.* Genetic disruption of 21-hydroxylase in zebrafish causes interrenal hyperplasia. *Endocrinology*, 2017, 158(12): 4165–4173
- Enrique SM, Alex C, Miguel AT, *et al.* Detection of growth-related QTL in turbot (*Scophthalmus maximus*). *BMC Genomics*, 2011, 12(1): 473
- Erickson T, Pillay LM, Waskiewicz AJ. Zebrafish Tshz3b negatively regulates hox function in the developing hindbrain. *Genesis*, 2011, 49(9): 725–742
- Gair HJC, Lovegrove B. Beyond homeosis-HOX function in morphogenesis and organogenesis. *Differentiation*, 2003, 71(8): 461–476
- Gutierrez AP, José MY, Fukui S, *et al.* Genome-wide association study (GWAS) for growth rate and age at sexual maturation in Atlantic salmon (*Salmo salar*). *PLoS One*, 2015, 10(3): e0119730
- He JB, Meng S, Zhao TJ, *et al.* An innovative procedure of genome-wide association analysis fits studies on germplasm population and plant breeding. *Theoretical and Applied Genetics*, 2017, 130(11): 2327–2343
- Hulata G. Genetic manipulations in aquaculture: A review of stock improvement by classical and modern technologies. *Genetica*, 2001, 111(1–3): 155–173
- Jiang DL, Gu XH, Li BJ, *et al.* Identifying a long QTL cluster across chrLG18 associated with salt tolerance in *Tilapia* using GWAS and QTL-seq. *Marine Biotechnology*, 2019, 21(2): 250–261
- Kim S, Sohn KA, Xing EP. A multivariate regression approach to association analysis of a quantitative trait network. *Bioinformatics*, 2009, 25(12): 204–212
- Klei L, Luca D, Devlin B, *et al.* Pleiotropy and principal components of heritability combine to increase power for association analysis. *Genetic Epidemiology*, 2008, 32(1): 9–19
- Lei JL, Ma AJ, Chen C, *et al.* The present status and sustainable development of turbot (*Scophthalmus maximus* L.) culture in China. *Engineering Sciences*, 2005(5): 34–38 [雷霖霖, 马爱军, 陈超, 等. 大菱鲆(*Scophthalmus maximus* L.)养殖现状与可持续发展. *中国工程科学*, 2005(5): 34–38]
- Lei JL, Men Q, Wang YG, *et al.* The factory farming mode of Dalingyu “greenhouse shed + deep well seawater”. *Marine Fisheries Research*, 2002, 23(4): 1–7 [雷霖霖, 门强, 王印庚, 等. 大菱鲆“温室大棚+深井海水”工厂化养殖模式. *海洋水产研究*, 2002, 23(4): 1–7]
- Li HM, Shi HL, Huo KK. p75NTR signal transduction suppressed by BFAR and p75NTR interactions. *Scientia Sinica Vitae*, 2011, 41(12): 1140–1147 [李红梅, 施慧莉, 霍克克. BFAR 与 p75NTR 的蛋白相互作用抑制 p75NTR 信号转导. *中国科学: 生命科学*, 2011, 41(12): 1140–1147]
- Li J, Liu YK, Bai L, *et al.* Isolation and identification of *Mycobacterium marinum* associated with splenic and renal granuloma disease of cultured turbot (*Scophthalmus maximus*). *Progress in Fishery Sciences*, 2019, 40(5): 195–199 [李杰, 刘耀宽, 白露, 等. 大菱鲆脾肾结节病病原菌的分离和鉴定. *渔业科学进展*, 2019, 40(5): 195–199]
- Lippert C, Listgarten J, Liu Y, *et al.* FaST linear mixed models for genome-wide association studies. *Nature Methods*, 2011, 8(10): 833–835
- Liu XF, Ma AJ, Huang ZH, *et al.* Expression characteristics analysis of major QTL candidate genes in response to high temperature stress in turbot (*Scophthalmus maximus*). *Journal of Fisheries of China*, 2019, 43(6): 1407–1409 [刘晓菲, 马爱军, 黄智慧, 等. 大菱鲆高温胁迫应答主效 QTL 候选基因的表达特性分析. *水产学报*, 2019, 43(6): 1407–1409]
- Nakajima H, Yonemura S, Murata M, *et al.* Myt1 protein kinase is essential for Golgi and ER assembly during mitotic exit. *Journal of Cell Biology*, 2008, 181(1): 89–103
- Porter HF, O'Reilly PF. Multivariate simulation framework reveals performance of multi-trait GWAS methods. *Scientific Reports*, 2017, 7: 38837
- Price AL, Zaitlen NA, Reich D, *et al.* New approaches to population stratification in genome-wide association studies. *Nature Reviews Genetics*, 2010, 11(7): 459–463
- Santana MHA, Ventura RV, Utsunomiya YT, *et al.* A genomewide association mapping study using ultrasound-scanned information identifies potential genomic regions and candidate genes affecting carcass traits in Nelore cattle. *Journal of Animal Breeding and Genetics*, 2015, 132(6): 420–427
- Silvia TRR, Toro MA, Bouza C, *et al.* QTL detection for *Aeromonas salmonicida* resistance related traits in turbot (*Scophthalmus maximus*). *BMC Genomics*, 2011, 12(1): 541
- Stephens M. A unified framework for association analysis with multiple related phenotypes. *PLoS One*, 2013, 8(7): e65245.
- Wang L, Wang YG, Zhang Z, *et al.* Diversity and drug resistance of bacterial pathogens isolated from bacterial ascetic disease in cultured turbot *Scophthalmus maximus*. *Progress in*

- Fishery Sciences, 2017, 38(4): 17–24 [王岚, 王印庚, 张正, 等. 养殖大菱鲂(*Scophthalmus maximus*)腹水病的病原多样性及其耐药性分析. 渔业科学进展, 2017, 38(4): 17–24]
- Wang WJ, Hu YL, Ma Y, *et al.* High-density genetic linkage mapping in turbot (*Scophthalmus maximus* L.) based on SNP markers and major sex- and growth-related regions detection. PLoS One, 2015, 10(3): e0120410
- Weger M, Diotel N, Weger BD, *et al.* Expression and activity profiling of the steroidogenic enzymes of glucocorticoid biosynthesis and the FDX1 co-factors in zebrafish. Journal of Neuroendocrinology, 2018, 30(4): e12586
- Xu K, Duan W, Xiao J, *et al.* Development and application of biological technologies in fish genetic breeding. Science China, Life Sciences, 2015, 58(2): 187–201
- Yu JM, Pressoir G, Briggs WH, *et al.* A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature Genetics, 2006, 38(2): 203–208
- Zhang DM, Wu AM, Lou LX, *et al.* Alteration of ion channel gene expression profile in rat model of post-myocardial infarction heart failure. Chinese Journal of Integrated Traditional and West Medicine, 2010, 30(1): 53–57 [张冬梅, 吴爱明, 娄利霞, 等. 心肌梗死后心力衰竭模型大鼠离子通道基因表达谱的变化. 中国中西医结合杂志, 2010, 30(1): 53–57]
- Zhou Q, Su ZC, Li YZ, *et al.* Genome-wide association mapping and gene expression analyses reveal genetic mechanisms of disease resistance variations in *Cynoglossus semilaevis*. Frontiers in Genetics, 2019, 10: 1167
- Zhou X, Stephens M. Efficient multivariate linear mixed model algorithms for genome-wide association studies. Nature Methods, 2014, 11(4): 407–409
- Zhu W, Zhang H. Why do we test multiple traits in genetic association studies? Journal of Korean Statistics Society, 2009, 38(1): 25–27

(编辑 冯小花)

Joint Genome-Wide Association Study of Body Mass and Morphological Traits in Turbot (*Scophthalmus maximus*)

GAO Jin^{1,2}, YANG Runqing^{1,2①}

(1. Wuxi Fisheries College, Nanjing Agricultural University, Wuxi 214081;

2. Research Centre for Aquatic Biotechnology, Chinese Academy of Fishery Sciences, Beijing 100141)

Abstract To reveal the molecular genetic mechanisms of body mass and morphological traits in turbot (*Scophthalmus maximus*) and scan molecular markers and candidate genes, which can be used to improve the target traits, a genome-wide association study (GWAS) was carried out using specific-locus amplified fragment technology (restriction site-associated DNA, 2b-RAD). First, body mass (BM), body length (BL), body width (BW), and caudal peduncle width (CPW) of 441 individuals were measured at about 473 days of growth period in a turbot breeding population. Second, all individuals were genotyped using 2b-RAD, and 23,988 SNPs were obtained after strict quality control. Using a multivariate linear mixed model (mvLMM) for GWAS of traits of BM-BL and BL-BW-CPW, 9 and 2 pleiotropic QTNs were detected for each phenotypic combination, respectively. However, a single-trait linear mixed model (LMM) based on the FaST-LMM algorithm was used for the association analysis of each trait, and the results showed that 4 QTNs were detected in the BM trait, 1 QTN was associated with BL and BW traits, respectively, and no significant locus was found in the CPW trait. A comparison between results of mvLMM and LMM found that mvLMM could detect more QTNs than LMM in GWAS, and the pleiotropic QTNs detected by mvLMM were more biologically meaningful. This study applied mvLMM and LMM to the joint GWAS of body mass and morphological traits in turbot, 17 significant QTNs were detected both using mvLMM and LMM, and 4 of them were detected repeatedly. Furthermore, 12 candidate genes were found by searching the nearest gene of each detected QTN on the whole turbot genome. All of them might be important candidate markers and functional genes, which could influence turbot body mass and morphology. Our study also provided the theory and a reference for marker-assisted selection of body mass and morphological traits in turbot.

Key words Turbot; Body mass and morphological traits; Genome-wide Association Analysis; Multivariate linear mixed models

① Corresponding author: YANG Runqing, E-mail: runqingyang@cafs.ac.cn