

基于转录组数据的大菱鲂(*Scophthalmus maximus*) SNP 标记开发及多态性分析*

王 婷^{1,2} 黄智慧¹ 马爱军^{1①} 马得友¹ 王新安¹ 夏丹丹¹ 马本贺¹

(1. 中国水产科学研究院黄海水产研究所 青岛市海水鱼类种子工程与生物技术重点实验室 农业部海洋渔业可持续发展重点实验室 青岛 266071; 2. 大连海洋大学 大连 116023)

摘要 大菱鲂(*Scophthalmus maximus*)是最具商业价值和养殖前途的海水鱼类之一,雌雄间生长有一定的差异。其高通量雌雄转录组测序的完成成为大规模鉴定和开发 SNP 标记提供了参考序列。本研究基于大菱鲂雌雄转录组测序数据,选择其中 45 个 SNP 位点,设计引物 63 组,其中 21 个位点(46.7%)应用小片段高分辨率溶解曲线(HRM)技术分型成功。对其进行多态性检测,21 个位点均具有二个单倍型。观测杂合度 H_o 的分布范围为 0.256—1.000,期望杂合度 H_e 的范围为 0.276—0.518,19 个位点符合 Hardy-Weinberg 平衡。14 个 SNP 位点位于基因编码区,其中 3 个属于非同义突变。含有这些 SNP 位点的基因大多与信号转导和转录翻译相关。本研究结果表明,高通量转录组测序和小片段 HRM 适合大规模 SNP 标记开发,为大菱鲂的分子遗传育种提供了候选标记资源。

关键词 大菱鲂;单核苷酸多态性(SNP);转录组;高分辨率溶解曲线(HRM)

中图分类号 Q953 **doi:** 10.11693/hyhz20140400101

大菱鲂(*Scophthalmus maximus*)属于鲽形目、鲽科、菱鲂属,在我国俗称“多宝鱼”。原产于欧洲大西洋东北部,20 世纪 70 年代,英国首度成功开发了大菱鲂人工养殖技术,于 1992 年引入我国,目前已成为我国北方沿海工厂化养殖业的主导品种之一(雷霖霖等, 2002)。随着养殖规模的不断扩大,集约化程度的不断提高,大菱鲂疾病频发、养殖周期长、养殖成本上升等严重制约了我国大菱鲂的养殖业的可持续性发展(雷霖霖等, 2012)。因此, 对大菱鲂进行遗传改良、选育新品种已成为提高养殖效益和健康化养殖水平的当务之急。

大规模发掘大菱鲂分子标记,是储备种质资源、构建性状相关遗传解析、开发分子遗传育种技术的基础工作。单核苷酸多态性标记(single nucleotide polymorphism, SNP)是美国学者 Lander 于 1996 年提出的第三代 DNA 遗传标记(Lander, 1996)。由于 SNP

数目多、覆盖密度大(Sachidanandam *et al*, 2001),具有遗传稳定性和代表性(Collins *et al*, 1997),便于高通量自动化检测等优点,已经广泛应用在构建高密度遗传连锁图谱构建、关联分析、分子辅助育种、群体遗传系统、品种鉴定等方面,并表现出良好的应用前景(Liu *et al*, 2004)。在水产动物中,SNP 的开发和应用已经有若干报道,如鲶鱼(Kucuktas *et al*, 2009)、栉孔扇贝(Jiang *et al*, 2011)、蟹(Ma *et al*, 2011)、虾夷扇贝(Liu *et al*, 2011)、对虾(吴莹莹等, 2013)等。在大菱鲂中也有部分开发和应用(刘庆明等, 2012; Navajas *et al*, 2012; Vera *et al*, 2013),但数量还未达到构建高密度遗传连锁图谱的要求。

目前大规模 SNP 位点信息的获得主要是通过基因组测序,伴随序列拼接比对而产生,但是这种方法成本较高。对于没有基因组信息的非模式生物,多利用已知的 EST 序列自行筛选,但是 EST 序列与实际

*国家高技术研究发展计划(863)项目资助, 2012AA10A408-8 号; 现代农业产业技术体系建设专项资金资助, CARS-50-01 号。
王婷, 硕士研究生, E-mail: cecily.333@163.com

① 通讯作者: 马爱军, 博士, 研究员, 博士生导师, E-mail: maaj@ysfri.ac.cn

收稿日期: 2014-04-03, 收修改稿日期: 2014-06-26

序列有一定差异,所以这种方法前期准备工作比较艰巨,开发效率比较低,覆盖率不够广,目的性不够强(Liu *et al*, 2006; 刘庆明等, 2012)。目前大菱鲂已知的 EST 序列只有 15641 个,使得该物种 SNP 筛选工作难度更高;随着第二代测序技术的快速发展,高通量转录组测序成本大幅降低,为解决这些难题提供了契机。通过新一代高通量测序技术,能够全面快速地获得某一物种特定组织或器官在某一状态下的几乎所有转录本及基因序列,并且可以准确检测出高度复杂的序列,通过一系列对比识别大量的 SNP 位点(Zhou *et al*, 2010)。本研究首次利用大菱鲂雌雄高通量转录组测序信息,去冗余后得到 71107 contig,为大菱鲂的 SNP 标记大规模开发提供了丰富的序列资源和位点信息,而且为进一步的 QTL 定位、性别控制等研究带来了极大便利。

迄今已经有很多种检测开发 SNP 多态性的技术方法,如基于杂交的基因芯片技术(Gene chips)(Wang *et al*, 1998)、探针技术(TaqMan)(Martinson *et al*, 2009)、基于核酸构象的单链构象多态性(Single-strand conformational polymorphism, SSCP)(He *et al*, 2008)等。虽然方法很多,但是存在着开发量低、消耗时间多、操作步骤冗长、实验费用昂贵、适用范围有限等问题,限制了 SNP 标记的利用和推广。高分辨率熔解曲线分析技术(High Resolution Melting, HRM),是近年来兴起的一种 SNP 检测的新手段,可以迅速的检测出 DNA 片段中碱基的突变(Li *et al*, 2012);因其操作简便快速,结果准确效率高,费用成本低,并且实现了真正的闭管操作,已受到广泛的关注(罗文龙等, 2011)。另外,HRM 的小片段法特异性好,避免了探针设计的误差和不对称 PCR 的复杂操作,降低了开发成本,提高了开发效率(Liew *et al*, 2004)。本研究在大菱鲂转录组 Solexa 高通量测序数据基础上,根据预测的 SNP 位点设计引物,通过小片段 HRM 分型技术,成功开发了大菱鲂 EST-SNP 标记 21 个,为大菱鲂遗传图谱构建、性状相关 QTL 定位和分子遗传育种提供了候选标记资源。

1 材料与方法

1.1 实验材料

实验用大菱鲂来自本实验室在烟台天源水产国家级良种场所构建的家系(马爱军等, 2012),共 96 尾鱼,样品详细资料见表 1。DNA 取样品尾鳍,使用海洋动物组织基因组 DNA 提取试剂盒提取,用超微量紫外

分光光度计定量,稀释至 40ng/μL, -20℃ 保存备用。

表 1 大菱鲂 SNP 标记开发样品信息
Tab.1 Sample information of turbot for SNP in *S. maximus*

序号	性状	数量	取样方式
1	耐高温	12	随机
2	不耐高温	12	随机
3	免疫好	12	随机
4	免疫差	12	随机
5	体色有星点	12	随机
6	体色无星点	12	随机
7	生长快	12	随机
8	生长慢	12	随机

1.2 SNP 位点的查找和小片段法引物设计

大菱鲂转录组 Solexa 高通量测序、contig 的拼装和 SNP 位点的预测由国家人类基因组南方研究中心完成。根据高通量转录组测序数据结果,查找覆盖度大于 500 的候选 SNP 位点,共得到 147 个位点。

利用 Primer5.0 对每一个候选位点设计一组引物,要求引物长度 20bp 左右,GC 含量为 45%—65%,产物长度小于 80bp;利用 Oligo7.0 分析引物参数,不能有能值高的错配、二聚体和发卡结构。高低温内标为两段长度 50bp 的 DNA 片段,提供独立于 PCR 产物的大约 68℃ 和 88℃ 的熔解曲线。分别为 5-ATCGTG ATTTCTATAGTTATCTAAGTAGTTGGCATTAAATAA TTTCATTTT-3'和 5'-GCGGTCAGTCGGCCTAGCGG TAGCCAGCTGCGGCACTGCGTGACGCTCAG-3'及各自的反向互补序列(Seipp *et al*, 2007)。引物和内标由上海生工生物工程技术服务有限公司合成。

1.3 PCR 引物验证和 HRM 小片段法分型

以 8 个群体,每个群体 6 个个体 DNA 等量混合作为模板 DNA,进行引物检测。PCR 反应体系为 10μL: 40 ng/μLDNA 模板 1μL, 2×ES Taq Master Mix 4.5μL, 上游引物(10μmol/L) 0.5μL; 下游引物(10μmol/L) 0.5μL, ddH₂O 3.5μL。PCR 反应程序为: 95℃ 预变性 5min; 94℃ 变性 30 s, 退火温度(视具体引物 *T_m* 值而定)退火 30s, 72℃ 延伸 30s, 共 35 个循环;最后 72℃ 延伸 7min, 4℃ 保存。用含 Genefinder 染料的 2%琼脂糖凝胶电泳检测 PCR 扩增产物。

根据电泳结果,选择目的条带清晰单一的位点进行高低温内标杂交并加入 LC Green Plus 染料。每个 PCR 反应体系中加入染料 1.0μL, 高温内标和低温内标的正负向引物各加 0.25μL, 95℃ 变性 5min 后,降温至 4℃ 保存。

使用 LightScanner 仪器进行 HRM 分型, 要求以 $0.1^{\circ}\text{C}/\text{s}$ 的速度, 从 56°C 快速升温到 97°C , 以 $1/\text{s}$ 的密度采集荧光信号, 得到高分辨率溶解曲线。结束后用 LightScanner 配套软件对采集曲线进行基因分型。

1.4 遗传多态性检测

SNP 位点的分型结果采用 POPGEN32 (version 1.32) 软件处理, 纯合子分别记为 AA 和 BB, 杂合子记为 AB, 统计各位点的有效等位基因数 (Effective number of alleles, N_e), 期望杂合度 (Expected heterozygosity, H_e), 观测杂合度 (Observed heterozygosity, H_o), 统计各位点基因型分布及其频率和最小等位基因频率 (Minor allele frequency, MAF)。用 PIC_CALC(0.6) 计算各位点多态信息含量 (Polymorphism information content, PIC)。

2 实验结果

2.1 引物验证和小片段 HRM 分型

本课题组高通量测序系统大规模测序的结果共获得 8642 个 SNP 位点, 转换颠换比为 1.3。在覆盖度大于 500 的 SNP 位点中, 根据两侧序列特点, 选取 45 个位点设计引物 63 组, 经过 PCR 体系条件优化, 设置最佳的 T_m 值, 利用 2% 的琼脂糖凝胶电泳进行检测, 有 21 对引物产生大小符合预期的单一条带。其中部分 SNP 位点的详细信息见表 2。对扩增出目的条带的位点, 用 8 个群体 96 个个体进行重新 PCR, 变性, 经过小片段 HRM (LightScanner) 验证, 21 个位点均能明显区分不同基因型。一共有 11 个颠换, 10 个转换, 其中 C/T 和 G/A 最常见 (图 1)。

表 2 部分 SNP 位点信息
Tab.2 Detailed information of part of the SNP loci in *S. maximus*

位点名称	SNP 类型	SNP 序列(5'→3')	产物长度(bp)
CS6	C/T	TTTGTAGCATTAGGTA GTTTC	53
S12	G/A	GTTGGGGGGCGGATA ACTCAC	67
S14	G/C	AAGCTATGCTGGCTTT GCTGC	61
S15	G/T	TTTGCAGTGCTTCTAT TTGTC	76
S16	T/A	TTGGGATGTTTGCAA CTCAGT	75
S3	C/T	AAGGACCGTTTGGCA TACAAA	80
S42	T/A	TTTGTCTATCTCGTGT TAACG	68
S62	G/A	CGTCAGAGGAGCAGA AGTTCT	84
S63	G/C	GCTTGAAGGAGACAA ATACTG	57

黑斜体字母表示 SNP 位点

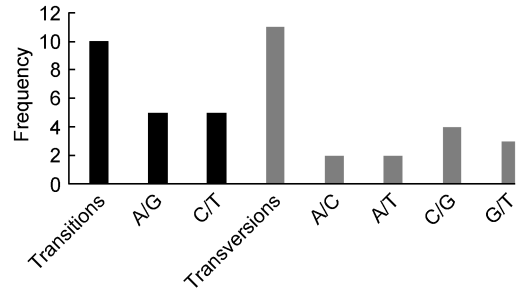


图 1 SNP 突变类型分布
Fig.1 Distribution of SNP variants

图 2a 所示为位点 S3 的小片段 HRM 分型图, 是转换, 其中红色曲线为纯合峰, 峰值较低, T_m 约 78.7°C , 是 T/T 型; 蓝色曲线为纯合峰, 峰值较高, T_m 值 79.4°C , 是 C/C 型; 灰色为杂合峰, 是 T/C 型。图 2b 所示为位点 S16 的小片段 HRM 分型图, 是颠换, 其中红色曲线为纯合峰, 峰值较高, T_m 约 80.7°C , 是 T/T 型; 蓝色曲线为纯合峰, 峰值较低, T_m 值 80.1°C , 是 A/A 型; 灰色为杂合峰, 是 A/T 型。

2.2 SNP 位点多态性分析

所得的基因型数据利用 POPGEN32(version 1.32) 统计分析, 结果显示 21 个 SNP 位点均含有两个单倍体 (表 3); 观测杂合度 H_o 的分布范围为 $0.256-1.000$, 期望杂合度 H_e 的范围为 $0.276-0.518$ 。有效等位基因数 N_e 分布范围 $1.352-2.051$ 。最小等位基因频率 MAF 分布范围 $0.0426-0.5000$ 。多态信息含量分析显示 21 个位点的 PIC 值范围为 $0.226-0.446$, 其中 2 个位点 (CS5, S7) 属于低度多态 ($\text{PIC} < 0.25$), 其余 19 个位点属于中度多态 ($0.25 < \text{PIC} < 0.5$)。通过哈代-温伯格平衡、连锁不平衡检验, 结果显示只有 S11、S42 两个位点的 $P < 0.05$, 不符合 H-W 平衡。通过以上分析证明, 得到的大多数 SNP 位点能够用于大菱鲆的遗传学分析。

2.3 SNP 位点序列的功能分析

利用 NCBI 的 BlastP, 对成功筛查到的 21 个多态 SNP 位点所在的 contig 序列进行功能注释, 21 个 SNP 位点所在的 20 个 contig 均有明确有效功能注释, 涉及细胞周期调控、细胞骨架、能量转换及 RNA 的加工修饰等 (表 4)。

利用 ORF finder 推测出开发阅读框, 在 NCBI 数据库中对, 确保 ORF 选择的正确性, 并确定 SNP 位点在遗传密码子的位置, 与标准密码子比对发现 CS1 位于 5'UTR 区, S11、S15、S16、S42、S3、CS10 位于 3'UTR 区; 而 CS5、S7 位于密码子的首位, CS9

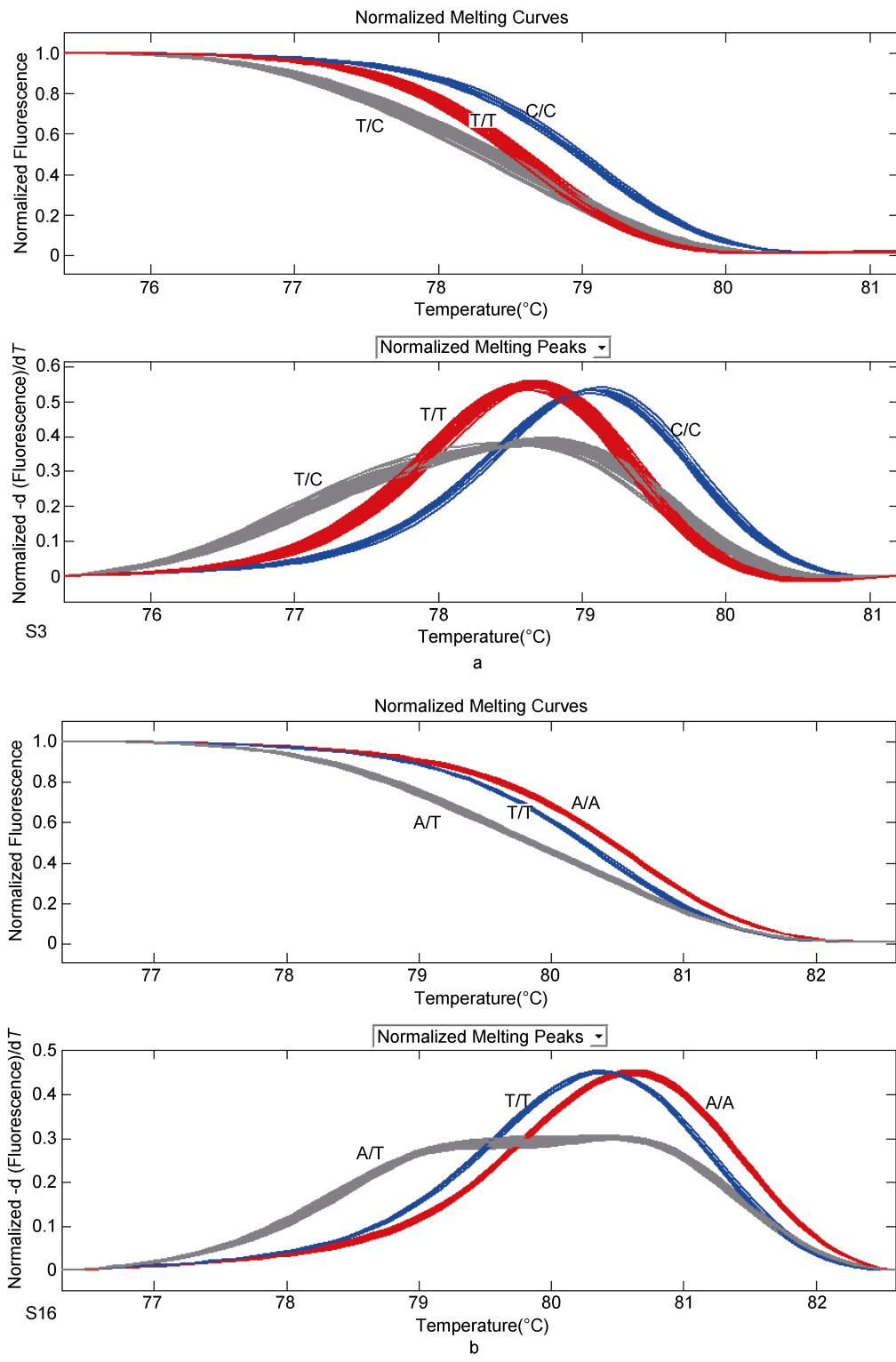


图 2 SNP 位点小片段 HRM 分型结果
Fig.2 Genotyping result using HRM with small amplicon
a: S3 分型情况, b: S16 分型情况

位于密码子第三位, 造成了编码氨基酸的改变, 为错义突变。其余 11 个位点都在密码子第三位, 为同义突变(表 5)。

3 讨论

SNP 标记的兴起, 极大地促进了遗传学和育种繁殖

表 3 大菱鲃 21 个 SNP 位点的遗传多态性
Tab.3 Genetic variability at 21 SNPs in *S. maximus*

SNP 位点	PIC	N_e	H_o	H_e	MAF
8S82	0.370	1.958	0.854	0.495	0.4271
CS1	0.373	1.986	0.915	0.502	0.4574
CS10	0.373	1.983	0.907	0.502	0.4535
CS3	0.374	1.958	0.505	0.495	0.4271
CS4	0.374	1.992	0.723	0.503	0.4681
CS5	0.235	1.374	0.325	0.276	0.1625
CS6	0.259	1.441	0.333	0.310	0.1889
CS9	0.374	1.991	0.841	0.503	0.4659
S10	0.371	1.968	0.447	0.497	0.4362
S11	0.446	2.051	0.638	0.518	0.0426
S12	0.375	2.000	1.000	0.505	0.5000
S14	0.308	1.614	0.511	0.384	0.2553
S15	0.358	1.878	0.575	0.472	0.3723
S16	0.374	1.990	0.605	0.503	0.4651
S22	0.344	1.792	0.447	0.447	0.3298
S3	0.305	1.600	0.417	0.379	0.2500
S42	0.332	1.724	0.333	0.425	0.3000
S62	0.366	1.932	0.479	0.488	0.4062
S63	0.375	1.999	0.896	0.505	0.4896
S7	0.226	1.352	0.256	0.264	0.1538
S9	0.374	1.992	0.648	0.503	0.4688

表 4 20 个序列预测蛋白的基因注释信息
Tab.4 Gene annotation of 20 predicted proteins

SNP 位点	Query Name	Annotation	Length of contig
8S82	comp30956	carbonic anhydrase 4-like (碳酸酐酶 4)	1645
CS1	comp38245	ribosomal protein L7a (核糖体蛋白 L7a)	1089
CS10	comp31070	uridine-cytidine kinase 2-B-like (尿苷胞苷激酶-2)	1108
CS3	comp28103	cyclin-A2-like (细胞周期蛋白-A2)	2668
CS4	comp35299	ribonucleoside-diphosphate reductase subunit M2-like isoform X2 (核糖核苷二磷酸还原酶)	1638
CS5	comp38118	tubulin alpha-1C chain isoform X1 (微管蛋白 α -1C 链)	2207
CS6	comp17868	mid1-interacting protein 1-B-like (相互作用蛋白 1)	702
CS9	comp18331	gem-associated protein 6-like (gem 相关蛋白 6)	1064
S10	comp29672	RNA-binding protein PNO1 (RNA 结合蛋白 PNO1)	1298
S11	comp32333	makorin 4 protein (假基因 4 蛋白)	2310
S12	comp32257	microfibril-associated glycoprotein 4-like (微纤维联合糖蛋白 4)	1547
S14	comp33063	eautophagy-related protein 101-like (与自噬性溶酶体相关的蛋白)	1444
S15	comp31115	ATPase, Na ⁺ /K ⁺ transporting, beta 2a polypeptide (Na ⁺ /K ⁺ 运输 ATP 合成酶 β -2a 多肽)	1149
S16	comp28490	tumor suppressor protein p53 (肿瘤抑制蛋白 53)	3138
S22	comp28329	ATP synthase subunit alpha, mitochondrial-like (线粒体中 ATP 合成酶 α 亚基)	2306
S3	comp17656	caprin-2-like (甘油酯-2)	3546
S42	comp31217	ubiquitin carboxyl-terminal hydrolase isozyme L5-like isoform 1 (泛素羧基末端水解酶的同功酶 L5)	1527
S62, S63	comp28258	putative deoxyribonuclease TATDN2-like (假定脱氧核糖核酸 TATDN2)	2760
S7	comp29640	calphotin-like (钙感光蛋白)	2824
S9	comp27994	nucleolar RNA helicase 2-like (核仁 RNA 解旋酶 2)	3633

表 5 SNP 位点的氨基酸突变
Tab.5 Amino acid mutations of SNPs

SNP	ORF	匹配序列 ID	氨基酸	ORF 长度(aa)	SNP 位置	E 值
8S82	1571-618	emb CAH69488.1	Ser TCT→TCG	317	1197	1.00E-157
CS1	239-1039	dbj BAF98653.1	5'UTR	266	230	0
CS10	160-948	ref XP_003453410.1	3'UTR	262	970	0
CS3	2248-899	ref XP_003443464.1	Ser TCA→TCC	449	1463	0
CS4	1587-430	ref XP_005808744.1	Gly GGT→GGC	385	499	0
CS5	293-1648	ref XP_005532035.1	Ala GCC→Ser TCC	451	434	0
CS6	547-140	ref XP_005455118.1	Leu CTA→CTG	135	359	2.00E-70
CS9	319-816	ref XP_003459829.1	Ser AGC→Arg AGG	165	630	2.00E-64
S10	1127-378	gb ACQ58604.1	Thr ATC→ATT	249	840	4.00E-158
S11	2177-1047	ref XP_005732714.1	3'UTR	376	1043	8.00E-136
S12	151-864	ref XP_004565306.1	Ala GCG→GCA	237	762	2.00E-109
S14	558-1214	ref XP_003459021.1	Leu CTG→CTC	218	623	3.00E-155
S15	903-37	ref XP_004576377.1	3'UTR	288	35	0
S16	438-1586	gb AGO20700.1	3'UTR	381	1804	0
S22	2192-537	ref XP_003974858.1	Val GTC→GTG	551	1926	0
S3	3367-653	ref XP_006805618.1	3'UTR	904	639	0
S42	213-1202	ref XP_003453569.1	3'UTR	329	1319	0
S62	2504-393	ref XP_003441345.1	Cys TGC→TGT	703	1434	0
S63	2504-393	ref XP_003441345.1	Val GTC→GTG	703	1506	0
S7	2519-819	ref XP_005738185.1	Thr ACG→Ala GCG	563	1052	1.00E-45
S9	231-2495	ref XP_005467227.1	Thr ACA→ACG	754	623	0

学的微观发展和深度发掘。然而, 开发和检测 SNP 是最关键的两个技术难点, 也限制了 SNP 标记在非模式生物(特别是水产动物)中的开发和利用。目前, SNP 标记的大规模开发, 一般是利用基因组测序获得的序列结果进行对比推测, 这种途径比较昂贵(庄伟, 2010); 或者利用 EST 序列, 但是由于序列误差的存在, 增加了验证的难度(刘庆明等, 2012)。高通量转录组测序的推广很好地弥补了这两种途径的缺陷(周华等, 2012), 得到的大量 EST 序列既可以提供相对更准确的序列信息, 又降低了成本。本研究利用 Solexa 高通量测序技术获得大量 EST-SNP 位点, 经过生物信息学分析, 筛选出 4314 个覆盖度大于 10 的候选 SNP 位点, 明确了其基因功能。数量低于其它水产动物, 除物种差别, 另外可能的原因是筛查条件较为严格(Ma *et al.*, 2011; Vera *et al.*, 2011)。

由于转录组测序的样本是大菱鲂 mRNA, 有一定的剪切和编辑, 本研究选取的 45 个 SNP 位点中有 5 个位点(11%)的 PCR 产物大于预期, 推测可能是由于内含子的存在而造成的。但是此数据远小于使用其它方法的比例(李纪勤等, 2013), 原因应是本研究使用

的是小片段法 HRM, 设计引物时, 使目的片段尽可能地小, 有效避免了内含子的出现。另外有 4 个位点, 有可观的多态性, 但是没有准确的分型结果, 因为产物溶解峰与高温内标溶解峰有部分重叠。要解决这个问题, 需要溶解温度更高的高温内标, 或者重新设计引物以期得到溶解温度更低的目的片段。

本研究采用高分辨率溶解曲线(HRM)技术, 并探索了小片段法分型技术, 通过该方法可以有效避免在同一个目标产物中存在 2 个以上 SNP 位点的可能性, 减少了内含子的影响, 从而提高分型效率, 大幅度降低了 SNP 标记开发的成本。本研究对 45 对 SNP 位点进行引物设计、验证、基因分型、多态性分析最终获得 21 个 SNP 标记可用于大菱鲂的遗传和育种研究, 成功率达到 46.7%, 远远高于非标记探针法(刘庆明等, 2012; 吴莹莹等, 2013)。

对获得的 21 个 SNP 标记, 在群体中进行遗传多态性验证, 多态信息(PIC)含量分析显示所有位点均具有多态性, 19 个为中等多态。观测杂合度的分布范围为 0.256 到 1.000, 期望杂合度的平均值为 0.451, 由于 SNP 标记为双等位形式, 每个位点所携带的多

态信息较少, 因此略小于大菱鲂的其它分子标记(Hou *et al*, 2011; Vera *et al*, 2013)。另外有两个偏离了 Hardy-Weinberg 平衡, 说明样本群体可能有过迁徙漂变等。

利用 EST 数据库开发 SNP 标记, 有一个显著的优点, 即标记位点来自编码序列(Coding Sequences, CDS), 与功能基因直接相关, 为进一步的功能基因研究提供依据。本研究获得 3 个非同义 cSNP (non-synonymous cSNP), 即碱基序列的改变可使以其为蓝本翻译的蛋白质序列发生改变, 从而影响了蛋白质的功能(刘越等, 2008)。另外, 有 7 个位于基因的 UTR 区, 可能与基因的表达调控相关(Vera *et al*, 2011, 2013), 后续基因功能表达途径等跟进研究中。

总之, 对于大菱鲂这种没有基因组数据的非模式生物, 应用高通量转录组测序结果, 对比 EST 序列, 使用小片段 HRM 技术, 可以实现高效率大规模 SNP 标记的开发, 而且开发位点与功能基因表达途径直接相关, 为进一步的遗传连锁图谱构建、经济性状的 QTL 定位以及分子标记辅助育种提供更多的标记基础。

致谢 中国科学院海洋研究所张国范研究员课题组在实验仪器方面提供了帮助, 谨致谢忱。

参 考 文 献

- 马爱军, 黄智慧, 王新安等, 2012. 大菱鲂(*Scophthalmus maximus*)耐高温品系选育及耐温性能评估. 海洋与湖沼, 43(4): 797—804
- 庄伟, 2010. 半滑舌鳎(*Cynoglossus semilaevis*)基因组 SNP 标记的开发与检测. 青岛: 中国海洋大学硕士学位论文, 1—57
- 刘越, 吕社民, 2008. 单核苷酸多态性影响基因功能的机制. 生命的化学, 28(2): 214—216
- 刘庆明, 李猛, 马爱军等, 2012. 应用高分辨率熔解曲线(HRM)开发大菱鲂(*Scophthalmus maximus*)SNP 标记的研究. 海洋与湖沼, 43(6): 1141—1148
- 李纪勤, 包振民, 李玲等, 2013. 栉孔扇贝 EST-SNP 标记开发及多态性分析. 中国海洋大学学报(自然科学版), 43(1): 56—63
- 吴莹莹, 孟宪红, 孔杰等, 2013. 非标记探针 HRM 法在中国对虾 EST-SNP 筛选中的应用. 渔业科学进展, 34(1): 111—118
- 罗文龙, 郭涛, 王慧等, 2011. HRM 及其在植物育种中的应用展望. 中国农学通报, 27(3): 10—14
- 周华, 张新, 刘腾云等, 2012. 高通量转录组测序的数据分析与基因发掘. 江西科学, 30(5): 607—611
- 雷霖, 王秉新, 2002. 大菱鲂“温室大棚+深井海水”工厂化养殖模式. 海洋水产研究, 23(4): 1—7
- 雷霖, 刘新富, 关长涛, 2012. 中国大菱鲂养殖 20 年成就和展望——庆祝大菱鲂引进中国 20 周年. 渔业科学进展, 33(4): 123—130
- Collins F S, Guyer M S, Chakravarti A, 1997. Variations on a theme: cataloging human DNA sequence variation. Science, 278(5343): 1580—1581
- He F, Wen H S, Dong S L *et al*, 2008. Identification of single nucleotide polymorphism cytochrome P450-c19a and its relation to reproductive traits in Japanese flounder (*Paralichthys olivaceus*). Aquaculture, 279(1): 177—181
- Hou S, Ma A, Wang X A *et al*, 2011. Isolation and characterization of 45 Polymorphic microsatellite loci of turbot (*Scophthalmus maximus*) and cross-species amplification. Chinese Journal of Oceanology and Limnology, 29: 311—316
- Jiang G, Li J, Li L *et al*, 2011. Development of 44 gene-based SNP markers in Zhikong scallop, *Chlamys farreri*. Conservation Genetics Resources, 3(4): 659—663
- Kucuktas H, Wang S, Li P *et al*, 2009. Construction of genetic linkage maps and comparative genome analysis of catfish using gene-associated markers. Genetics, 181(4): 1649—1660
- Lander E S, 1996. The new genomics: global views of biology. Science, 274(5287): 536—539
- Li F, Niu B, Huang Y *et al*, 2012. Application of high-resolution DNA melting for genotyping in lepidopteran non-model species: *Ostrinia furnacalis* (Crambidae). PloS One, 7(1): e29664
- Liew M, Pryor R, Palais R *et al*, 2004. Genotyping of single-nucleotide polymorphisms by high-resolution melting of small amplicons. Clinical Chemistry, 50(7): 1156—1164
- Liu W, Li H, Bao X *et al*, 2011. The first set of EST-derived single nucleotide polymorphism markers for Japanese scallop, *Patinopecten yessoensis*. Journal of the World Aquaculture Society, 42(3): 456—461
- Liu Y G, Liu L X, Lei Z W *et al*, 2006. Identification of polymorphic microsatellite markers from RAPD product in turbot (*Scophthalmus maximus*) and a test of cross-species amplification. Molecular Ecology Notes, 6(3): 867—869
- Liu Z J, Cordes J F, 2004. DNA marker technologies and their applications in aquaculture genetics. Aquaculture, 238(1): 1—37
- Ma H, Ma Q, Ma C *et al*, 2011. Isolation and characterization of gene-derived single nucleotide polymorphism (SNP) markers in *Scylla paramamosain*. Biochemical Systematics and Ecology, 39(4): 419—424
- Martinsohn J T, Ogden R, 2009. FishPopTrace-Developing SNP-based population genetic assignment methods to investigate illegal fishing. Forensic Science International: Genetics Supplement Series, 2(1): 294—296
- Navajas Pérez R, Robles F, Molina Luzón M J *et al*, 2012. Exploitation of a turbot (*Scophthalmus maximus* L.) immune-related expressed sequence tag (EST) database for microsatellite screening and validation. Molecular Ecology Resources, 12(4): 706—716

- Sachidanandam R, Weissman D, Schmidt S C *et al*, 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature*, 409(6822): 928—933
- Seipp M T, Durtschi J D, Liew M A *et al*, 2007. Unlabeled oligonucleotides as internal temperature controls for genotyping by amplicon melting. *The Journal of Molecular Diagnostics*, 9(3): 284—289
- Vera M, Álvarez-Dios J A, Millán A *et al*, 2011. Validation of single nucleotide polymorphism (SNP) markers from an immune Expressed Sequence Tag (EST) turbot, *Scophthalmus maximus*, database. *Aquaculture*, 313(1): 31—41
- Vera M, Álvarez-Dios J, Fernández C *et al*, 2013. Development and validation of single nucleotide polymorphisms (SNPs) markers from two transcriptome 454-runs of turbot (*Scophthalmus maximus*) using high-throughput genotyping. *International Journal of Molecular Sciences*, 14(3): 5694—5711
- Wang D G, Fan J, Siao C *et al*, 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*, 280(5366): 1077—1082
- Zhou X, Ren L, Li Y *et al*, 2010. The next-generation sequencing technology: a technology review and future perspective. *Science China Life Sciences*, 53(1): 44—57

DEVELOPMENT AND POLYMORPHIC ANALYSIS OF SNP MARKERS IN *SCOPHTHALMUS MAXIMUS* BASED ON TRANSCRIPTOME DATABASE

WANG Ting^{1,2}, HUANG Zhi-Hui¹, MA Ai-Jun¹, MA De-You¹,
WANG Xin-An¹, XIA Dan-Dan¹, MA Ben-He¹

(1. Yellow Sea Fisheries Research Institute, Chinese Academy of Fishery Sciences; Qingdao Key Laboratory for Marine Fish Breeding and Biotechnology; Key Laboratory of Sustainable Development of Marine Fisheries, Ministry of Agriculture, Qingdao 266071, China;
2. Dalian Ocean University, Dalian 116023, China)

Abstract Turbot *Scophthalmus maximus* is a commercially valuable fish and one of the most promising aquaculture species. Completion of turbot high-throughput transcriptome sequencing provides a reference sequence to identify and develop large-scale SNP markers. Based on *Scophthalmus maximus* transcriptome sequencing data, we selected 45 SNPs and designed 63 pairs of primers, of which 21 (46.7%) loci using high-resolution melting (HRM) technology with small amplicons genotyping success. All the 21 sites have two haplotypes, and proved polymorphic. The observed heterozygosity (H_o) estimates ranged 0.256—1.000; expected heterozygosity (H_e) estimates 0.276—0.518, and 19 loci were in Hardy-Weinberg equilibrium after Bonferroni correction. Fourteen SNPs were located in the coding region, 3 non-synonymous changes. Most SNP-containing genes were related to signal transduction mechanisms and transcription. This study shows that high throughput transcriptome sequencing and small amplicons of HRM is suitable for large-scale SNP marker development, which provides a number of candidate markers resources to molecular genetic breeding of turbot.

Key words turbot *Scophthalmus maximus*; single nucleotide polymorphisms (SNP); transcriptome; high resolution melting (HRM)