

单核苷酸多态性及其在水产动物遗传育种中的应用

刘福平^{1,2}, 白俊杰¹

(1. 中国水产科学研究院 珠江水产研究所 中国水产科学研究院热带亚热带鱼类选育与养殖重点开放室, 广东 广州 510380;
2. 上海海洋大学 生命科学与技术学院, 上海 200090)

摘要: 单核苷酸多态性 (SNPs) 是广泛存在于基因组中的一类由单个碱基转换或颠换引起的 DNA 序列多态性。它是继限制性片段长度多态性 (RFLP)、微卫星标记 (SSR) 之后的新一代分子标记, 已广泛应用于构建遗传连锁图谱、QTL 定位、关联分析及研究群体遗传结构与亲缘关系等方面。本文主要介绍了 SNPs 的概念、特点及检测 SNPs 的主要方法, 综述了 SNPs 在水产动物遗传育种中的研究进展, 以期将 SNPs 更广泛地应用于水产动物群体遗传、分子标记辅助育种和生物进化等研究领域。[中国水产科学, 2008, 15 (4): 704-712]

关键词: 单核苷酸多态性; 分子标记; 水产动物; 遗传育种

中图分类号: Q953 **文献标识码:** A **文章编号:** 1005-8737-(2008)04-0704-09

遗传标记主要分为形态学标记、细胞学标记、生化标记和 DNA 分子标记 4 种类型。近年来, DNA 分子标记辅助育种作为一种新的育种方法在全世界兴起, 它是通过利用与目标性状紧密连锁的 DNA 分子标记对目标性状进行间接选择的现代育种技术。与常规育种相比, 该技术不仅可在早期进行准确、稳定的选择, 而且解决了再度利用隐性基因识别难的问题, 有效地克服了常规育种过程中所采用的形态学标记存在的数目少、受环境影响较大, 不易直接选择的不足, 从而加速育种进程, 可提高育种效率 2~3 倍。目前常用的 DNA 分子标记技术主要包括: 限制性内切酶酶切片段的长度多态性 (RFLP)、随机扩增 DNA 多态性 (RAPD)、DNA 指纹印迹 (DAF)、扩增片段长度多态性 (AFLP)、微卫星 DNA (SSR)、单核苷酸多态性 (SNPs) 等。单核苷酸多态性 (Single Nucleotide Polymorphisms, 简称 SNPs) 这个术语最早于 1994 年出现在人类分子遗传杂志上, 随后美国麻省理工学院学者 Lander^[1] 在 Science 上第一次正式提出 SNPs 为新一代分子标记。Wang D G 等^[2] 利用 DNA 芯片技术检测到的 3 241 个 SNPs 构建了遗传密度为 2 cM 的第一张人类 SNPs 遗传图谱。SNPs 是继限制性片段长

度多态性标记和微卫星标记之后的第 3 代遗传标记, 它分布广泛、数目多且相对稳定地存在于各类生物的染色体上, 近年来正被越来越多的应用于人类遗传性疾病诊治和动物遗传结构分析等研究领域^[3]。本文将主要论述 SNPs 的特点、SNPs 主要检测方法和 SNPs 在水产动物育种中的应用, 旨在将 SNPs 更广泛的应用于水产动物群体遗传、分子标记辅助育种和生物进化等研究领域。

1 SNPs的概念和特点

1.1 SNPs的概念

SNPs 主要是指在基因组水平上由于单个核苷酸的变异所引起的 DNA 序列多态性, 即同一物种不同个体间染色体上遗传密码单个碱基的变化。常表现为单碱基的转换、颠换、插入和缺失, 但通常所指的 SNPs 不包括后两种情况, 即插入和缺失; 满足 SNPs 条件的另一规定是其中一种等位基因在群体中出现的频率应不小于 1%^[4]。转换是指嘌呤与嘌呤之间或者嘧啶与嘧啶之间的相互转换 ($A \leftrightarrow G$ 或 $C \leftrightarrow T$); 颠换则是指嘌呤与嘧啶之间的相互替代 ($C \leftrightarrow A$, $C \leftrightarrow G$, $A \leftrightarrow T$)。通常发生转换与颠换的 SNPs 之比为 2:1, 这是由于 CpG 二核苷酸的胞嘧

收稿日期: 2008-01-25; 修订日期: 2008-05-04.

基金项目: 农业部 948 项目 (2006-G55); 国家科技支撑计划 (2006BAD01A1209); 国家科技基础条件平台工作项目 (2005DKA21103).

作者简介: 刘福平 (1983-), 男, 在读硕士研究生, 从事水产动物遗传育种研究.

通讯作者: 白俊杰. Tel: 020-81616129; E-mail: jjbai@163.net

啶是最易发生突变的位点,胞嘧啶可发生甲基化脱去氨基而形成胸腺嘧啶。理论上,在一个二倍体生物群体中,SNPs可能是由2个、3个或4个等位基因构成,但实际上3个或4个等位基因的SNPs罕见,所以SNPs通常被认为只有2种等位基因,即双等位基因分子标记^[5]。根据SNPs在基因组中的位置可分为编码区SNPs(coding region SNPs)、基因周边SNPs(perigenic SNPs)和基因间SNPs(intergenic SNPs)3类^[4]。大多数SNPs位于基因组的非编码区,对蛋白质无直接影响,这一类SNPs作为遗传标记在群体遗传和生物进化地研究中有着很重要的作用^[6]。少数分布在基因编码区的SNPs称为cSNPs(coding SNPs)。Halushka等^[7]根据对75个基因的检测推测人类基因组中约存在100万个SNPs位点,其中约50万个分布在非编码区,20万~40万个SNPs在编码区,而分布在编码区的非同义突变SNPs只有2.4万~4万个。根据对遗传性状的影响,cSNPs还可分为:(1)同义突变:编码序列的改变并不影响所翻译的氨基酸序列,蛋白质的功能不变;(2)非同义突变:可分为错义突变和无义突变,前者指改变编码序列导致翻译的氨基酸序列改变,从而改变蛋白质的生物学功能;无义突变是指突变成终止密码子,翻译终止。Henikoff^[8]估计SNP数据库中3084条非同义SNPs中,25%的SNPs影响蛋白质的功能。

1.2 SNPs的特点

1.2.1 数量多且分布广泛 据估计,在人类基因组中大约平均每1900bp就会出现1个SNPs,整个基因组中大约有142万个SNPs,其发生频率超过1%^[9]。正是由于SNPs数量巨大,从而弥补了其多态性不足的缺点。Kruglyak^[10]认为使用700~900个SNPs进行基因组扫描构建遗传图谱的能力相当于目前使用300~400个微卫星标记,而如果使用1500~3000个SNPs作基因组扫描,其结果明显高于目前普遍采用的微卫星标记。

1.2.3 富有代表性 某些位于基因内部的SNPs有可能直接影响蛋白质结构或表达水平,因此它们可能代表某些性状的遗传基础。

1.2.4 遗传稳定性 SNPs是基因组中分布最广泛且稳定的点突变,突变率低,与微卫星等重复序列多态标记相比,SNPs具有更高的遗传稳定性,尤其是处于编码区的SNPs^[11]。

1.2.5 检测易于实现自动化 SNPs通常是一种双

等位基因的遗传变异,在检测时无需检测微卫星标记那样对片段的长度做出测量,只需一个“+/-”或“全或无”分析的方式,有利于发展自动化筛选或检测SNPs^[12]。

2 SNPs的检测方法

从技术上来讲,凡是能够检测出点突变的方法都可以用来鉴定SNPs。随着分子生物学技术的飞跃发展,SNPs基因分型技术也不断涌现。在一些经典的SNPs检测技术,如单链构象多态性(SSCP)和限制性片段长度多态性(RFLP)等仍在实践中广泛使用的同时,近几年又出现了一系列高灵敏度、高通量的基因分型方法,如温控高效液相色谱法、TaqMan探针法等,可以满足大样本及多SNPs位点的基因分型要求。在实际应用时要根据研究目的、实验条件以及经费情况等来进行有目的的选择。在这些众多的检测方法中,根据其检测原理可以大致分为以下4种类型^[13]。

2.1 直接测序法

直接测序法是检测SNPs最准确的方法,其检出率可达100%。主要是指对不同个体同一基因或基因片段的PCR产物进行测序和序列比较,或对已定位的序列标签位点(STS)和表达序列标签(EST)进行再测序。其流程为:PCR扩增目的片段→纯化、回收→测序。随着测序自动化程度的提高和测序成本的降低,直接测序将会越来越多地用于SNPs的检测与分型,但目前直接测序仍然是一种费用较高的方法,而且对于杂合体不易分型。另外常用的一种方法为简化代表性散弹(Reduced Representation Shotgun, RRS)测序。其主要步骤为:首先用限制性内切酶消化DNA成许多片段,然后将这些片段连接到载体,载入受体细胞,最后进行测序。Altshuler等^[14]用该方法在人类基因组中发现了47172个SNPs,为建立人类高密度SNPs图谱奠定了基础。

2.2 以构象为基础的检测法

2.2.1 单链构象多态性(Single-strand conformational polymorphism, SSCP) 日本学者Orita等^[15]研究发现,单链DNA片段呈复杂的空间折叠构象,这种立体结构主要是由其内部碱基配对等分子内相互作用力来维持的,当有一个碱基发生改变时,会或多或少地影响其空间构象,使构象发生改变,空间构象有差异的单链DNA分子在聚丙烯酰胺凝胶中受

排阻大小不同。因此,通过非变性聚丙烯酰胺凝胶电泳(PAGE),可以非常敏锐地将构象上有差异的分子分离开。该方法称为单链构象多态性。在随后的研究中,Orita等^[16]又将SSCP用于检查PCR扩增产物的基因突变,从而建立了PCR-SSCP技术,进一步提高了检测突变方法的简便性和灵敏性。PCR-SSCP的基本过程包括:①PCR扩增靶DNA;②将特异的PCR扩增产物变性,而后快速复性,使之成为具有一定空间结构的单链DNA分子;③将适量的单链DNA进行非变性聚丙烯酰胺凝胶电泳;④最后通过放射性自显影、银染或溴化乙锭显色分析结果。若发现单链DNA带迁移率与正常对照的相比发生改变,就可以判定该链构象发生改变,进而推断该DNA片段中有碱基突变。经实验证明,小于300 bp的DNA片段中的单碱基突变,90%可被SSCP发现。该方法简便、快速、灵敏,不需要特殊的仪器,但它的不足之处是只能作为一种粗略地突变检测方法,要最后确定突变的位置和类型,还需进一步测序;此外电泳条件也要求较严格;另外,由于SSCP是依据点突变引起单链DNA分子立体构象的改变来实现电泳分离的,这样就可能会出现当某些位置的点突变对单链DNA分子立体构象的改变不起作用或作用很小时,再加上其他条件的影响,使聚丙烯酰胺凝胶电泳无法分辨造成漏检。尽管如此该方法还是在各领域得到了广泛的应用,乃至今天仍然不失为一种检测SNPs的经典方法^[17-18]。

SSCP技术自创立以来,经历了自身发展和完善的过程,刚建立时是将同位素掺入PCR扩增物中,通过放射自显影来显示结果。这给该技术的推广造成一定的困难,随着DNA银染方法与PCR-SSCP结合,使得该方法大大简化。SSCP值得注意的改进是将DNA-SSCP分析改为RNA-SSCP分析,其基本原理是:RNA有着更多精细的二级和三级构象,这些构象对单个碱基的突变很敏感,从而提高了检出率,其突变检出率可达90%以上。另外,RNA不易结合成双链,因此可以较大量的进行电泳,有利于用溴化乙锭染色。但该方法增加了一个反转录过程,还需要一个较长的引物,内含有启动RNA聚合酶的启动序列,从而相对地增加了该方法的难度。Sarkar等^[19]用RNA-SSCP法检测了28名B型血友病患者的凝血因子IX的基因序列,对于全长2.6 kb的凝血因子IX基因组用直接测序法检测到的20处碱基点突变,用RNA-SSCP可检测

出其中的70%,而DNA-SSCP只能检测出35%,证明了RNA-SSCP比DNA-SSCP具有更高的灵敏性。另一方面,为了进一步提高SSCP的检出率,还可将SSCP分析与其他突变检测方法相结合。其中与杂交双链分析(Heteroduplex analysis, Het)法结合可以大大提高检出率。Het法是用探针与要检测的单链DNA或RNA进行杂交,含有一对碱基对错配的杂交链可以和完全互补的杂交链在非变性PAGE凝胶上通过电泳被分离开。对同一靶序列分别进行SSCP和Het分析可以使点突变的检出率100%,而且实验简便。Ravnik-Glavac等^[20]对膀胱纤维基因27个外显子中的134个突变点进行了检测,结果表明,单独用SSCP检出的效率为75%~98%,而结合Het法则能全部检测出来。

2.2.2 温度梯度凝胶电泳(Temperature gradient gel electrophoresis, TGGE) 该方法主要是根据点突变和正常DNA片段由于解链温度 T_m 值的不同所表现出的不同解链行为。DNA片段在聚丙烯酰胺凝胶电泳中通过设置温度梯度来进行分离,当温度达到最低熔解区即开始解链,使片段呈Y形结构,因而降低其迁移速率。一个碱基的不同足以导致迁移速率的不同,从而达到在温度梯度电泳中分离的效果。如果序列是已知的,则DNA双链片段的变性行为可以用Poland软件进行预测。该方法可用于200~900 bp内的DNA片段,检出率可达100%^[21]。

2.2.3 变性梯度凝胶电泳(Denaturing gradient gel electrophoresis, DGGE) 变性梯度凝胶电泳是利用双链DNA分子在一定浓度梯度变性剂的凝胶中电泳时,会在一定变性剂浓度下发生部分解链,导致电泳迁移率下降,两个DNA分子间即使只有一个碱基对的差异,也会在不同时间不同地点发生部分解链,从而被分离成两条带^[22]。DGGE与TGGE类似,只是DGGE依靠变性剂使 T_m 值不同的分子分离,而TGGE是依靠温度梯度。DGGE能检测的片段可长达1 kb,若SNPs发生在最先解链的DNA区域,其检出率也可达100%,尤其是100~500 bp的片段,所以该技术已被广泛应用于SNPs的检测。

2.2.4 变性高效液相色谱(Denaturing high-performance liquid chromatography, DHPLC) 变性高效液相色谱法(DHPLC)是近年来在DGGE和SSCP基础上发展起来的一种检测SNPs的技术,其基本原理是:含有突变位点的PCR扩增产物经变性、逐步降温退火后,将形成同源和异源双链2种DNA分

子。在部分变性条件下,由于异源双链的解链温度较低,发生错配的异源双链 DNA 更易于解链为单链 DNA,会先形成单螺旋 DNA 从色谱柱中流出,从而与同源双链 DNA 分离^[23]。DHPLC 的检出率为 90%~95%,有效片段大小为 70~1 000 bp。它的优点是成本低,操作简单,易自动化,与 DGGE 和 SSCP 相比,更适合于大样本筛选,且有较高的重复性;缺点是仍然不能准确指明 SNPs 在片段中的位置。

2.3 基于 PCR、酶切的检测法

2.3.1 聚合酶链式反应-限制性片段长度多态性 (PCR-RFLP) 检测 PCR-RFLP 是 SNPs 筛选中最经典的方法之一,已广泛应用于 SNPs 的检测中。限制性片段长度多态性 (Restricted Fragment Length Polymorphism, RFLP) 是指用限制性内切酶切割不同个体基因组 DNA 后,含有同源序列的酶切片段在长度上的差异。最初由 Botstein 等^[24]首先提出利用 RFLP 标记构建遗传图谱,直至 1987 年 Donis Keller 等^[25]才构建出第一张人的 RFLP 图谱。由于传统的 RFLP 检测需要 Southern 杂交等过程,使得分析成本偏高、速度较慢,其应用受到了很大的限制。现在的 RFLP 标记一般与 PCR 技术结合起来应用,其原理是根据基因的突变会产生或消除某些酶切位点,从而利用限制性内切酶的特异性,用 1 种或 1 种以上的限制性内切酶作用于同一 DNA 片段,如果存在 SNPs 位点,酶切片段的大小和数目就会出现差异,然后进行电泳检测就可以判断是否有 SNPs 位点以及碱基替换的类型。不难看出,该方法只适合于检测酶切位点处的 SNPs,对于酶切位点以外的 SNPs 则无法检测,因而在一定程度上限制了该方法的应用。

2.3.2 突变酶学检测 (Enzymatic mutation detection, EMD) EMD 是利用 T_4 内切酶 VII 的特性检测 SNPs, T_4 内切酶 VII 又称为离解酶,能够解离重组中间体,识别并切割一些特殊 DNA 底物,如: 十字结构、分支 DNA、单链突出、单碱基错配而形成的泡状结构及因插入或缺失而形成的突变^[26]。EMD 检测的方法简述如下: PCR 扩增不同个体的 DNA 片段,混合,经变性、复性形成异源双链,用 T_4 内切酶 VII 处理,电泳分离产物片段,根据片段大小判断是否存在 SNPs 及大致位置。EMD 的优点是: 方法简单,速度快,检出率可高达 100%; 可分析几个 kb 的 DNA 片段,能检测多种类型的 SNPs,并且可以找出 SNPs 的大致位置。

2.4 基于杂交的检测法

2.4.1 TaqMan 探针技术 TaqMan 技术是以荧光共振能量传递 (Fluorescent resonance energy transfer, FRET) 为基础的检测方法,在 PCR 反应中,将供者-受者染料对 (发光基团和淬灭基团) 分别结合到 Taqman 探针的两端,探针未与目标序列结合时,通过 FRET 作用使供者不发荧光; 完全互补配对后,由于 Taq DNA 聚合酶具有 5' 核酸酶的活性,可将供者从探针上切下来从而发出荧光。如果探针与目标序列中存在错配碱基,就会减少探针与目标序列结合的紧密程度及 Taq DNA 聚合酶切割供者的活性,也就影响了供者的荧光释放量,从而使碱基突变链和正常链得以区分。TaqMan 技术将检测结合在 PCR 反应过程中,无须分离或洗脱,提高了速度并减少了 PCR 污染的可能性^[27]。其缺点主要是: 敏感性受 Taq 酶活性影响较大; 设计探针时需注意 FRET 的传递效率和 PCR 扩增效率,探针设计成本较高。

2.4.2 基因芯片 (Gene chip) 检测 基因芯片又称 DNA 芯片 (DNA chips),是在固相支持介质上进行分子杂交并原位荧光检测的一种高通量 SNPs 分析方法。根据核苷酸的碱基配对原理,设计 2 种或多种探针,在优化操作后,探针只与其完全互补的序列杂交,而与含有单个错配碱基的序列不杂交。特定序列的探针固定在特殊的载体表面上,如玻璃、硅片等,制成 DNA 芯片。待测基因经提取扩增、荧光化学标记后,与固定的探针进行杂交。然后洗去没有发生杂交的样品,即可检测杂交样品。由于目标基因和探针杂交的程度与荧光强度及种类有关,因此通过激光扫描,可根据荧光强弱或荧光的种类检测出被检序列的碱基类别^[2]。利用基因芯片技术筛查 SNPs 是随着近几年芯片技术的迅速发展而建立的一种高通量、自动化的检测手段,应用该方法可以寻找新的 SNPs 位点,并实现 SNPs 位点在基因组中的精确定位。但是该项技术还不够成熟,由于杂交条件对不同 GC 含量的 DNA 是完全不同的,还没有找到一种普遍性的杂交条件,另外还需要解决重复序列对杂交准确度的影响。

3 SNPs在水产动物遗传育种中的应用

SNPs 在构建高密度遗传连锁图谱、关联分析、群体遗传结构及系统发育分析、品种鉴定等方面均表现出了良好的应用前景^[28]。下面重点阐述水产

动物遗传育种中 SNPs 的应用研究及研究进展。

3.1 构建高密度遗传连锁图谱

遗传连锁图谱 (Genetic linkage map), 是生物基因组结构研究以及进行 QTL 准确定位的一个重要前提。遗传连锁图谱上包括的标记数越多, 分布越均匀则定位的基因就越精确, 这将为标记辅助选择、重要经济性状的 QTL 定位及至最终实现基因型选择创造条件。到 20 世纪 80 年代, 各种 DNA 分子标记的出现使得构建中、高密度遗传连锁图谱成为可能, 特别是微卫星 (SSR) 曾一度成为育种学家的最爱。SNPs 在基因组中分布的广泛性及其在同一位点上的双等位特性, 使之适合于自动化大规模扫描, 成为继 SSR 之后最受推崇的作图标记。人类^[13] 以及作为实验遗传的模式动物鼠 (*Mus musculus*)^[29]、果蝇 (*Drosophila melanogaster*)^[30]、模式植物拟南芥 (*Arabidopsis thaliana*)^[31] 等的 SNPs 遗传连锁图谱已经建立起来。在水产动物方面, Stickney 等^[32] 采用寡核苷酸微阵列 (Oligonucleotide microarrays) 技术, 对模式动物斑马鱼 (*Danio rerio*) 中 712 个基因和表达序列标签 (EST) 进行大规模扫描得到了 2 035 个 SNPs, 其中发生转换的 SNPs 占 54.6%, 由颠换造成的 SNPs 占 45.4%。构建了斑马鱼的第一幅 SNPs 遗传图谱, 该图谱由 25 个连锁群 (Linkage Group, LG) 组成, 其全长为 3 000 cM, 包括了 1 930 个 SNPs, 平均分辨率为 6.98 cM。利用斑马鱼的 SNPs 图谱, 将以前一些用基因图谱或微卫星图谱不能定位的突变和基因进行了定位。通过斑马鱼 SNPs 图谱, Stickney 成功地将 Talbot 等^[33] 发现的 floating head (*flh*) 突变定位于 LG13; 功能基因 *st11* 也被成功地定位到 LG2。SNPs 图谱的构建无疑将更有利于 QTL 的精确定位。目前用 SNPs 标记构建遗传连锁图谱还只限于模式种斑马鱼, 它所确立的一系列方法也可用于水产动物的 SNPs 遗传图谱的构建。

3.2 连锁不平衡、关联分析和功能学验证

SNPs 标记在群体水平上的应用研究最具有吸引力的方面是利用群体遗传学中的连锁不平衡 (Linkage disequilibrium, LD) 原理来进行关联分析 (Associate study)^[5]。连锁不平衡是一个复杂现象, 遗传距离、不同等位基因的选择压力、遗传漂变、群体的瓶颈效应以及发生新突变均对连锁不平衡有着很大的影响。但由于漂变和选择产生的不平衡

在不连锁的基因座之间将很快消失, 而紧密连锁的基因座之间的连锁不平衡消失很慢, 因而通过研究一个位标与性状相关基因座之间的连锁不平衡将有助于目标性状的精细定位。由于 SNPs 在基因组十分丰富, 它的稳定性和容易记录使得人们可以利用连锁不平衡分析来深入研究群体遗传学的一些基本理论问题, 并在此基础上进行关联研究。进行关联分析时, 采用候选基因法来分析功能基因的等位基因与表型的关联性, 是一个有效的方法^[34]。候选基因 SNPs 标记是一种重要的 QTL 定位方法, 它通过揭示直接在生理上或在生长发育过程中能得以表现的标记基因与控制数量性状的主效基因的关系, 而用于 QTL 定位。

肉质是家畜重要经济性状之一, 其中影响最为突出的是双肌性状。双肌性状是由于肌肉生长抑制素 (Myostatin, MSTN) 缺失或突变从而导致肌肉过度生长或肥大引起的。Grobet L 等^[35] 研究表明皮埃蒙特牛 *MSTN* 基因在第三外显子处发生错义突变 (G → A), 使得蛋白质成熟区酪氨酸替代胱氨酸, 导致 *MSTN* 基因失活, 从而表现出双肌性状。生长性状是水产动物中最重要的一个经济性状, 与水产动物生长性能相关的候选基因主要包括生长激素 (Growth hormone, GH)、生长激素受体 (Growth hormone receptor, GHR)、类胰岛素生长因子 (Insulin-like growth factor, IGF) 等。国内外学者通过候选基因法在寻找与水产动物生长性状相关的 SNPs 位点上也取得了初步成效。Tao 等^[36] 采用 PCR-RFLP 和双向特异等位基因扩增 (Bidirectional amplification of specific alleles, Bi-PASA) 技术对北极嘉鱼 (*Salvelinus alpinus* L.) 两个全同胞家系中与生长相关的 10 个候选基因进行了 SNPs 位点的筛选, 并进行了 SNPs 位点与生长性状的关联分析。结果表明, 10 个候选基因中有 5 个 (*GH1*、*GH2*、*IGF1*、*Pit1* 及 *GHRH/PACAP2*) 含有 SNPs 位点 (8 个), 其中位于基因 *GHRH/PACAP2* 上的 SNPs 位点与北极嘉鱼早期生长速率存在极显著性相关 ($P=0.000\ 01$); 同时表明, 对于非模式生物, 通过近缘种序列比对来设计引物扩增候选基因进行 QTL 定位是一种切实可行的方法。Gross 等^[37] 用 *TaqI* 酶切大西洋鲑 (*Salmo salar* L.) *GH1* 基因从第 1 外显子到第 4 外显子的 1 825 bp 片段, 结果在内含子 3 上发现了一个新的 *TaqI* 酶切位点, 在 1 龄大西洋鲑中共检测到了 3 个等位基因 8 种基因型, 在这 3

个群体中不同等位基因和基因型的分布存在显著性差异 ($P<0.05$)。Prudence 等^[38]对太平洋牡蛎 (*Crassostrea gigas*) *amylase* 基因 (*AMYA* 和 *AMYB*) 进行了 PCR-RFLP 分析, 结果表明, 牡蛎肉重与 *amylase* 基因型显著相关, 这对于遗传育种中提高牡蛎生长速度具有重要的潜在价值。Xu 等^[39]克隆了与尖吻鲈 (*Lates calcarifer*) 肌肉生长相关的小清蛋白基因 (*PVALB1* 和 *PVALB2*), 在 *PVALB1* 基因的 3' 非编码区发现了 1 个微卫星位点, 该位点与尖吻鲈孵出 90 天后的体质量、体长极显著相关 ($P<0.01$); 同时通过直接测序法在 *PVALB2* 基因的 3 个内含子上检测到了 3 个 SNPs 位点, 但没有发现与尖吻鲈的生长性状存在相关性。Kang 等^[40]利用 Southern 杂交技术对牙鲆 (*Paralichthys olivaceus*) GH 第 1 外显子到第 5 外显子的 2 114 bp 进行了分析, *Sau3AI* 限制性内酶在牙鲆大中小 3 个群体中检测到了 6 个等位基因 15 种基因型, 这些等位基因和基因型在 3 个不同群体中的分布具有显著性差异 ($P<0.05$), 可以将这些位点作为与牙鲆生长相关的标记。Case 等^[41]用 *Dra I* 酶切大西洋鳕 (*Gadus morhua*) 和其他海域鳕 *pan I* 基因中的 313 bp 片段, 发现大西洋鳕 *pan I* 基因的变异对其生长性状有着重要的影响, 其中属于 *pan I ab* 基因型的鳕在出生 10 周后的体质量和体长明显高于 *pan I bb* 型的鳕 ($P<0.01$)。

国内学者倪静等^[42]根据牙鲆生长激素 (GH) 基因的 5 个外显子序列设计引物, 通过 PCR-SSCP 分析技术对其进行检测, 结果表明, 该群体生长激素基因的第 4 个外显子存在多态性 (C → T), 进行关联分析发现不同基因型的个体在体质量和头长上存在显著性差异 ($P<0.05$)。本实验室于凌云等通过候选基因法, 对大口黑鲈 (*Micropterus salmoides*) *MyoD* 基因进行了 SNPs 位点的筛选, 在内含子上发现有 7 个 SNPs 位点, 这些位点在分析群体中的突变频率为 4.2%~35.3%, 均大于 1% 因而可以认为是突变点, 为接下来分析这 7 个 SNPs 位点与大口黑鲈生长性能的相关性奠定了基础 (待发表)。相对畜牧和家禽而言, SNPs 标记应用于水产动物关联分析方面起步较晚, 研究还很少, 但随着 SNPs 技术的迅速发展和各国学者对水产养殖业的重视, 必将在水产动物分子标记辅助育种中得到广泛的应用。

对于 SNPs 位点与某一性状的关联性研究, 不同实验室得出的结果往往也存在很大差异甚至完

全相反, 而且与某一性状显著关联的 SNPs 到底是发挥功能性作用, 还是仅仅与功能性 SNPs 相连锁的遗传标记甚至二者只是某种联系上的假象, 这都需要进行功能学研究来加以证实。预计引起细胞功能学改变的 SNPs 在所有 SNPs 只占极小一部分, 大多数是分布在基因启动子区域可能发挥调节转录效应的 SNPs 和蛋白质编码区域引起编码氨基酸改变的 SNPs。分子生物学技术的迅速发展, 极大地推动了 SNPs 功能学验证的深入研究。现在比较成熟的对于启动子区域 SNPs 功能学研究技术主要包括: (1) 报告基因转染技术。这一技术主要用于研究启动子 SNPs 对于 mRNA 转录效率的影响, 通过观察转录结果来判断 SNPs 是否具有功能。(2) 凝胶迁移滞后实验 (Electrophoretic mobility shift assays, EMSA)。该技术通过在体外合成含 SNPs 位点的寡核苷酸与转录因子特异性结合, 观察二者结合的强度和效率, 但是该技术只是人工合成较短长度的寡核苷酸, 没有考虑 SNPs 位点周围遗传背景环境的影响^[43]。(3) 染色质免疫沉淀分析 (Chromatin immunoprecipitation assay, CHIP)。该技术通过超声将染色体碎片化, 再将碎片化的核酸片段与转录因子结合, 然后通过 PCR 技术扩增观察判断二者结合的效率 and 强度^[44]。

现阶段对于 SNPs 功能学的研究主要集中在人类复杂疾病上。Sun 等^[45]通过关联分析发现 *CASP8* 基因启动子区域 -652 6N 缺失与肺癌风险的降低存在显著相关性。同时对其进行了 SNPs 功能学验证, 发现由于 -652 6N 的缺失使得 *CASP8* 基因对应的 mRNA 表达量有所下降, 最终导致 T 淋巴细胞中 *CASP8* 的活性和在癌细胞抗原诱导下的细胞死亡率降低。SNPs 的功能学验证这一特点是其他分子标记所不具备的, 因而用 SNPs 标记来进行关联性研究再进行功能学验证很大程度上提高了标记的可靠性, 具有其他分子标记无可比拟的优越性。虽然 SNPs 功能学验证在水产动物中还没有相关研究, 但随着水产动物分子生物技术的不断发展, 该技术独特的优点在水产动物遗传育种中将具有广泛的应用前景。

3.3 种群进化和亲缘关系的研究

SNPs 具有分布广泛位点丰富、低突变率 (10^{-8} ~ 10^{-9}) 能稳定遗传及二等位基因标记易于实现自动化等特点, 非常适合群体遗传学中种群进化和亲缘关系的研究^[46]。线粒体 DNA (mtDNA)

具有分子结构简单、严格的母性遗传、几乎不发生重组、进化速度快、不同区域进化速度存在差异等特点,使其成为群体遗传学和分子系统学研究的重要标记。在水产动物方面,将 SNPs 标记应用于 mtDNA 来研究特种的进化和亲缘关系已经取得了一系列进展。Chow 等^[47]用 4 种限制性内酶 (*Alu*, *Dde* I, *Hha* I 和 *Rsa* I) 对 13 个不同海域的 456 尾剑鱼 (*Xiphias gladius*) 的线粒体 DNA 进行了 PCR-RFLP 分析,共检测到了 52 种基因型,遗传距离为 0.702 ~ 0.962。*Rsa* I 在地中海剑鱼群体中没有发现多态性,说明很少有外来剑鱼进入该水体中;同时还成功地将大西洋与太平洋的剑鱼群体进行了区分。Aranishi F 等^[48]对亲缘关系非常接近的 3 种鳕科鱼类的细胞色素 *b* 基因进行了 PCR-RFLP 分析,结果在阿拉斯加州鳕中扩增出了 3 条带 (106 bp, 161 bp 和 291 bp),太平洋鳕为两条带 (106 bp 和 452 bp),而在大西洋鳕中没有切开,从而很好地将 3 种鳕区分开来。Itoi 等^[49]运用 PCR-RFLP 和 *TaqMan* MGB 探针技术,得到的不同大小的酶切片段及荧光密度的不同能够很好地将日本鳗鲡 (*Anguilla japonica*) 和欧洲鳗鲡 (*Anguilla anguilla*) 区分开来。由于等位酶标记受到取样数量大小的限制,而微卫星标记在不同实验室或国家之间的数据传输困难,Smith 等^[50]利用 10 个 SNPs 位点成功地将分别位于美国和加拿大流域的大鳞大麻哈鱼区分开来,其精确度达到 95%。汪登强等^[51]利用 PCR-RFLP 对 13 种鲟形目鱼类 mtDNA 进行了分析,结果显示,鲟科的 6 种鱼与匙吻鲟和白鲟分成两大支;鲟科中两种鳇没有聚到一起,支持了鳇不能作独立分类单元的观点,为鲟形目鱼类遗传和进化研究提供了科学依据。

4 问题与展望

SNPs 是生命遗传物质基因变异的主要存在形式,既可用于高密度遗传图谱的构建,也可用于基于候选基因或整个基因组的关联研究。近年来,SNPs 在人类流行病学关联研究中取得的成果,极大地促进了 SNPs 在动物基因组研究中的应用。一系列发现和检测 SNPs 的方法、构建图谱的策略以及连锁不平衡和关联分析等技术正被广泛应用于动物基因组研究领域。水产动物的主要经济性状(如生长性状)大都属于数量性状(QTL),其遗传机制比较复杂,受到众多基因控制。要想精确

定位这些 QTL 从而辅助育种,就必须建立高密度的遗传图谱。SNPs 标记的出现可以很好地解决这一问题,与微卫星相比,SNPs 在基因组中的分布更为广泛,拥有更庞大的遗传信息,并且易于自动化。因而,更适合于建立水产动物遗传图谱,进行 QTL 精细定位。但由于 SNPs 应用于水产动物研究起步较晚,缺乏足够的 SNPs 位点来构建遗传图谱或进行关联分析。在今后的研究中,应致力于用高效的检测方法大规模筛选 SNPs 位点,进行关联分析以及 SNPs 功能学验证研究;另一方面可从已知的与重要经济性状相关的基因着手,研究这些基因外显子、内含子 SNPs 多态性,特别是调控序列上的 SNPs 多态性,这样更容易找到与特定功能相关的分子标记,提高水产动物选择育种的效率。

参考文献:

- [1] Lander E S. The new genomics: global views of biology[J]. Science, 1996, 274: 536-539.
- [2] Wang D G, Fan J B, Siao C J, et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphism in the human genome[J]. Science, 1998, 280: 1 077-1 082.
- [3] Saxena R, Voight B F, Lyssenko V, et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels[J]. Science, 2007, 316: 1 331-1 336.
- [4] Alain V, Denis M, Magali S C, et al. A review on SNP and other types of molecular markers and their use in animal genetics[J]. Genet Sel Evol, 2002, 34: 275 305.
- [5] Brookes A J. The essence of SNPs[J]. Gene, 1999, 234: 177 186.
- [6] Syvanen A C. Accessing genetic variation: genotyping single nucleotide polymorphisms[J]. Nat Rev Genet, 2001, 2: 930 942.
- [7] Halush M K, Fan J B, Bentley K, et al. Patterns of single-nucleotide polymorphisms in candidate genes for blood-pressure homeostasis[J]. Nat Genet, 1999, 22: 239-247.
- [8] Henikoff S. Accounting for human polymorphisms predicted to affect protein function[J]. Genome Res, 2002, 12(3): 436-446.
- [9] Sachidanandam R, Weissman D, Schmidt S C, et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms[J]. Nature, 2001, 409: 928-933.
- [10] Kryuglyak L. The use of a genetic map of biallelic markers in linkage studies[J]. Nat Genet, 1997, 17: 21-24.
- [11] Weber J L, Wong C. Mutation of human short tandem repeats[J]. Hum Mol Genet, 1993, 2: 1 123 1 128.

- [12] Brookes A, Day I. SNP attack on complex traits[J]. Nat Genet, 1998, 20(3): 217-218.
- [13] Syvanen A C. Accessing genetic variation: genotyping single nucleotide polymorphisms[J]. Nat Rev Genet, 2001, 2: 930-942.
- [14] Altshule D, Pollara V J, Cowles C R, et al. An SNP map of the human genome generated by reduced representation shotgun sequencing[J]. Nature, 2000, 407: 513-516.
- [15] Orita M, Iwahana H, Kanazawa H, et al. Detection of polymorphisms of human DNA by gel electrophoresis as single-strand conformation polymorphisms[J]. Proc Natl Acad Sci USA, 1989, 86: 2766-2770.
- [16] Orita M, Suzuki Y, Sekiya T, et al. Rapid and sensitive detection of point mutations and DNA polymorphisms using the polymerase chain reaction[J]. Genomics, 1989, 5: 874-879.
- [17] Murakami Y, Katahira M, Makino R, et al. Inactivation of the retinoblastoma gene in a human lung carcinoma cell line detected by single-strand conformation polymorphism analysis of the polymerase chain reaction product of cDNA[J]. Oncogene, 1991, 6: 37-42.
- [18] Michaud J, Brody L C, Steel G, et al. Strand-separating conformational polymorphism analysis: efficacy of detection of point mutations in the human ornithine delta- α aminotransferase gene[J]. Genomics, 1992, 13: 389-394.
- [19] Sarkar G, Yoon H S, Sommer S S. Screening for mutations by RNA single-strand conformation polymorphism (rSSCP): comparison with DNA-SSCP[J]. Nucleic Acids Res, 1992, 20(4): 871-878.
- [20] Ravnik-Glavac M, Glavac D, Dean M. Sensitivity of single-strand conformation polymorphism and heteroduplex method for mutation detection in the cystic fibrosis gene[J]. Hum Mol Genet, 1994, 3: 801-807.
- [21] Risesner D, Steger G, Wiese U, et al. Temperature-gradient gel electrophoresis for the detection of polymorphic DNA and for quantitative polymerase chain reaction[J]. Electrophoresis, 1992, 13(9): 63-66.
- [22] Cotton R G H, Malcolm A D B. Mutation detection[J]. Nature, 1991, 353: 582-583.
- [23] Oefner P J, Underhill P A. DNA mutation detection using denaturing high performance liquid chromatography (DHPLC) [J]. Am J Hum Genet, 1995, 57: 226.
- [24] Botstein D. Construction of a genetic linkage map in the man using restriction fragment length polymorphism[J]. Hum Genet, 1980, 32: 314-331.
- [25] Donis-Keller H, Knowlton R G, Braman J C, et al. Genotyping by restriction fragment length polymorphisms [P]. European Patent; 0221633, 05/13/1987.
- [26] Del Tito B J, Poff H E, Novotny M A, et al. Automated fluorescent analysis procedure for enzymatic mutation detection[J]. Clin Chem, 1998, 44: 731-739.
- [27] Ranade K, Chang M S, Ting C T, et al. High-Throughput genotyping with single nucleotide polymorphisms[J]. Genome Res, 2001, 11: 1262-1268.
- [28] Liu Z J, Cordes J F. DNA marker technologies and their applications in aquaculture genetics[J]. Aquaculture, 2004, 238: 1-37.
- [29] Lindblad-Toh K, Winchester E, Daly M J, et al. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse [J]. Nat Genet, 2000, 24: 381-386.
- [30] Hoskins R A, Phan A, Naeemuddin M, et al. Single nucleotide polymorphism markers for genetic mapping in *Drosophila melanogaster* [J]. Genome Res, 2001, 11: 1100-1113.
- [31] Cho R J, Mindrinos M, Richards D R, et al. Genome-wide mapping with biallelic markers in *Arabidopsis thaliana* [J]. Nature, 1999, 23: 203-207.
- [32] Stickney H L, Schmutz J, Woods L G, et al. Rapid mapping of zebrafish mutations with SNPs and oligonucleotide microarrays[J]. Genome Res, 2002, 12: 1929-1934.
- [33] Talbot W S, Trevarrow B, Halpern M E, et al. A homeobox gene essential for zebrafish notochord development[J]. Nature, 1995, 378: 150-157.
- [34] Lynch M, Walsh B. Genetics and analysis of quantitative traits [M]. Sinauer Assoc Inc. Sunderland, MA, USA, 1998, 980p.
- [35] Grobet L, Poncelet D, Royo L J, et al. Molecular definition of an allelic series of mutations disrupting the myostatin function and causing double-muscling in cattle[J]. Mammal Genome, 1998, 9: 210-213.
- [36] Tao W J, Boulding E G. Association between single nucleotide polymorphisms in candidate genes and growth rate in Arctic charr (*Salvelinus alpinus* L.) [J]. Heredity, 2003, 91: 60-69.
- [37] Gross R, Nilsson J. Restriction fragment length polymorphism at the growth hormone 1 gene in Atlantic salmon (*Salmo salar* L.) and its association with weight among the offspring of a hatchery stock[J]. Aquaculture, 1999, 173: 73-80.
- [38] Prudence M, Moal J, Boudry P, et al. An *amylase* gene poly-

- morphism is associated with growth differences in the Pacific cupped oyster *Crassostrea gigas* [J]. *Anim Genet*, 2006, 37: 348–351.
- [39] Xu Y X, Zhu Z Y, Lo L C, et al. Characterization of two parvalbumin genes and their association with growth traits in Asian seabass (*Lates calcarifer*) [J]. *Anim Genet*, 2006, 37: 266–268.
- [40] Kang J H, Lee S J, Park S R, et al. DNA polymorphism in the growth hormone gene and its association with weight in olive flounder *paralichthys olivaceus* [J]. *Fish Sci*, 2002, 68: 494–498.
- [41] Case R A J, Hutchinson W F. Association between growth and Pan I genotype within Atlantic cod full-sibling families [J]. *American Fisheries Society*, 2006, 135: 241–250.
- [42] 倪静, 尤锋, 张培军, 等. 牙鲆 GH 基因外显子多态性与生长性状关系的初步研究 [J]. *高技术通讯*, 2006, 16(3): 307–312.
- [43] Pitarque M, Richter O, Rodriguez-Antona C, et al. A nicotine C-oxidase gene (CYP2A6) polymorphism important for promoter activity [J]. *Hum Mutat*, 2004, 23: 258–266.
- [44] Buck M J, Lieb J D. CHIP-chip: considerations for the design, analysis, and application of genome-wide chromatin immunoprecipitation experiments [J]. *Genomics*, 2004, 83: 349–360.
- [45] Sun T, Gao Y, Tan W, et al. A six-nucleotide insertion-deletion polymorphism in the *CASP8* promoter is associated with susceptibility to multiple cancers [J]. *Nat Genet*, 2007, 39(5): 605–613.
- [46] Brumfield R T, Beerli P, Nickerson D A, et al. The utility of single nucleotide polymorphisms in inferences of population history [J]. *Trends Ecol Evol*, 2003, 18: 249–256.
- [47] Chow S, Okamoto H, Uozumi, et al. Genetic stock structure of the swordfish (*Xiphias gladius*) inferred by PCR-RFLP analysis of the mitochondrial DNA control region [J]. *Marine Biology*, 1997, 127: 359–367.
- [48] Aranishi F, Okimoto T, Izumi S. Identification of gadoid species (Pisces, Gadidae) by PCR-RFLP analysis [J]. *J Appl Genet*, 2005, 46(1): 69–73.
- [49] Itoi S, Nakaya M, Kaneko G, et al. Rapid identification of eels *Anguilla japonica* and *Anguilla anguilla* by polymerase chain reaction with single nucleotide polymorphism-based specific probes [J]. *Fish Sci*, 2005, 71: 1356–1364.
- [50] Smith C T, Templin W D, Seeb J E, et al. Single nucleotide polymorphisms provide rapid and accurate estimates of the proportions of U.S. and Canadian Chinook salmon caught in Yukon River fisheries [J]. *North American J Fish Manag*, 2005, 25: 944–953.
- [51] 汪登强, 危起伟, 王朝明, 等. 13 种鲟形目鱼类线粒体 DNA 的 PCR-RFLP 分析 [J]. *中国水产科学*, 2005, 12(4): 383–389.

Single nucleotide polymorphisms and its application in genetic breeding of aquatic animals

LIU Fu-ping^{1,2}, BAI Jun-jie¹

(1. Pearl River Fisheries Research Institute, Chinese Academy of Fishery Sciences, Key Laboratory of Tropical and Subtropical Fish Breeding & Cultivation, Chinese Academy of Fishery Sciences, Guangzhou 510380, China; 2. College of Aqua-life Science and Technology, Shanghai Ocean University, Shanghai 200090, China)

Abstract: Single nucleotide polymorphisms (SNPs) consist of a single nucleotide base alteration which included transition and transversion. As a new molecular marker following restriction fragment length polymorphism (RFLP) and simple sequence repeat (SSR), SNPs are used widely in genetic linkage maps, quantitative trait loci location, association studies, population genetics structure and genetics relationship. The conception, characteristics, detection methods, and advancement in genetic breeding of aquatic animals were introduced in this review. It is anticipated that SNPs marker will contribute greatly to studies on aquatic animals population genetics, molecular breeding as well as evolutionary biology. [*Journal of Fishery Sciences of China*, 2008, 15(4): 704–712]

Key words: single nucleotide polymorphisms; molecular markers; aquatic animals; genetic breeding

Corresponding author: BAI Jun-jie. E-mail: jjbai@163.net