

一种基于变性高效液相色谱技术筛选单核苷酸多态性并估算基因频率的方法

杜红丽, 方梅霞, 聂庆华, 雷明明, Adam Ishag NEAMA, 张细权

(华南农业大学 动物科学学院, 广东 广州 510642)

摘要: 选取生产性能具有明显差异的 4 个鸡品种(莱航鸡、隐性白洛克鸡、丝羽乌骨鸡和杏花鸡)为材料, 利用变性高效液相色谱(DHPLC)技术检测鸡生长激素基因部分序列(518 bp)的多态性, 快速筛选到 5 个可能与生产性能相关的单核苷酸多态性(SNPs)。根据 DHPLC 杂合谱型与测序基因型一一对应的关系, 介绍了当所检测目的 DNA 片段中含有多个 SNPs 时基因频率估算的简单方法, 比较估算的和 PCR-RFLP 统计的基因频率, 发现差异不显著, 证明该估算方法有一定的可行性。

关键词: 变性高效液相色谱; 单核苷酸多态性; 基因频率

中图分类号: S831.2

文献标识码: A

文章编号: 1001-411X(2005)03-0084-05

A method for estimating allelic frequencies of single nucleotide polymorphisms (SNPs) based on denaturing high performance liquid chromatography (DHPLC)

DU Hong-li, FANG Mei-xia, NIE Qing-hua, LEI Ming-ming, Adam Ishag NEAMA, ZHANG Xi-quan

(College of Animal Science, South China Agric. Univ., Guangzhou 510642, China)

Abstract: Four breeds of chickens (white Leghorn, white recessive rocks, Taihe silkies and Xinhua) with different production performance were applied to screen potential single nucleotide polymorphisms (SNPs) related to production traits in a region of the growth hormone gene using denaturing high performance liquid chromatography (DHPLC). Five SNPs were found in a fragment of 518 bp. In the present study a simple and rapid method to estimate allelic frequencies of SNPs in the case of several SNPs existing in the same targeted DNA fragment was introduced based on the correspondence between the heterozygous DHPLC peak and sequencing genotype, of which the accuracy was evaluated by comparing with the allelic frequencies result of PCR-RFLP. It's a feasible method because no significant difference was detected between the estimated allelic frequencies and those calculated by PCR-RFLP.

Key words: denaturing high performance liquid chromatography (DHPLC); single nucleotide polymorphisms (SNPs); allelic frequency

单核苷酸多态性(single nucleotide polymorphisms, SNPs)在动物基因组中广泛分布, 具备多态信息量大、易于检测和统计分析等优点, 被称为继限制性片段长度多态性(RFLP)和微卫星多态性之后的第 3 代

收稿日期: 2004-06-24

作者简介: 杜红丽(1975-), 女, 博士研究生。通讯作者: 张细权(1963-), 男, 教授, 博士;

E-mail: xqzhang@scau.edu.cn

基金项目: 国家重大基础研究规划项目(G2000016102)

©1994-2015 China Academic Journal Electronic Publishing House. All rights reserved. <http://www.cnki.net>

遗传标记^[1]。用于检测 SNPs 的技术有多种,如单链构象多态性(SSCP)、RFLP、变性梯度凝胶电泳(DGGE)等^[2]。变性高效液相色谱(denaturing high performance liquid chromatography, DHPLC)是近年发展起来的一种筛查未知 SNPs 的有效方法,具有检测效率高、成本低、易于操作和快速准确等优点^[3],在人功能基因 SNPs 研究中已获得较好的应用^[4~8]。其原理是通过带正电荷的流动相将带负电荷的 DNA 分子吸附到固定相上,在不完全变性的条件下,通过杂合和纯合二倍体在柱中滞留时间(retention time)的差异,来检测目的 DNA 的碱基变异^[9]。但是利用 DHPLC 技术进行 SNPs 筛查时,由于不同类型纯合子的滞留时间几乎相同,很难区分,故难以统计不同群体的等位基因频率,对选择合适分子标记进行经济性状 QTL 定位和遗传多样性等后续研究带来困难。因此,建立一种合理的频率估算方法,对利用 DHPLC 技术进行 SNPs 的检测以及 SNPs 的应用研究显得尤其重要。

有研究利用 DHPLC 技术筛查与鼻咽癌相关基因 SNPs 时,没有考虑野生型和变异型纯合子难以区分这一点^[9],把所有表现为单峰的个体都认为是野生型纯合子,并粗略统计了 SNPs 位点纯合与杂合个体的比例^[10]。本试验则不考虑表现为单峰的个体,只统计 DHPLC 技术检测到的肩峰或不对称峰个体,根据它们与测序图(基因型)一一对应的关系,建立方法估算各 SNPs 位点的等位基因频率,并与 PCR-RFLP 的结果相比较,验证这种方法的可行性。

1 材料与方法

1.1 试验鸡群及样本处理

20 只莱航蛋鸡来自广州市力康公司,31 只速生型隐性白洛克肉鸡、41 只丝羽乌骨鸡和 34 只杏花鸡由广东温氏食品集团南方家禽育种公司提供,供试样本均无亲缘关系。鸡翅静脉采血,0.02 g/mL EDTA 抗凝,苯酚-氯仿法抽提鸡基因组 DNA,溶解于 TE 缓冲液中,−20 °C 保存备用。

1.2 引物设计和 PCR 反应

根据鸡生长激素基因全序列(GeneBank 登陆号: AB095994)选择−429~+89 区域设计特异性引物,该区域包含调控区和部分外显子 1。引物序列为: 101F: 5'-GCCCTGGCAGCCCTGTTAAC-3'; 101R: 5'-CACCCACCATCGTATCCATC-3'。25 μL 反应体系中包含 50 ng 基因组 DNA,1×PCR 缓冲液,0.2 mmol/L

dNTP, 1.5 mmol/L Mg²⁺, 0.25 μmol/L 引物, 0.8 U Taq 聚合酶。PCR 反应程序: 94 °C 3 min; 94 °C 30 s, 62 °C 45 s, 72 °C 1 min, 35 个循环; 后延伸 72 °C 5 min。反应完成后用 0.01 g/mL 的琼脂糖凝胶电泳检测 PCR 产物。

1.3 DHPLC 检测

1.3.1 参数设置 突变检测的分离梯度: 输入 DNA 序列及选择检测方式, 软件系统自动模拟选择最佳分离梯度; 温度: 软件系统根据 DNA 序列预测 t_m 值, 检测温度一般在 $(t_m \pm 2)$ °C 范围内合理选择, 根据引物 101FR 扩增片段的 t_m 值及溶解曲线, 选择 56.4、58.4 和 60.4 °C 作为 DHPLC 筛查 SNPs 的温度。

1.3.2 突变检测 设定使恒温自动取样器将 7 μL PCR 产物注入 WAVE[®] DNA 片段分析系统, PCR 产物被流动相 A 液 [0.1 mol/L TEAA, φ(乙腈)=0.025%] 和 B 液 [0.1 mol/L TEAA, φ(乙腈)=25%] 以 0.9 mL/min 的速度带至分离柱被洗脱分离。

1.3.3 谱型鉴别 如果谱型峰的个数或峰的对称性改变, 则说明 DNA 片段中存在变异。

1.4 测序和 BLAST 分析

对每种肩峰或不对称峰的谱型, 随机抽取 2 个样品 PCR 产物, 送交上海博亚生物公司进行产物纯化测序, 每个样品均测定前后 2 个反应; 利用 DNASTAR 软件对测序结果进行比较分析以确定 SNPs。

1.5 PCR-RFLP 分析

以 4 个鸡品种 126 个个体为材料, 对 SNP 位点 C-12IT 进行 PCR-RFLP 分析。8 μL PCR 产物中加入 3 U Pae I 内切酶和 0.9 μL 10× Buffer, 37 °C 水浴过夜。酶切产物用 0.015 g/mL 琼脂糖凝胶电泳检测, 得出 RFLP 基因型。

1.6 频率统计和 χ^2 检验

1.6.1 频率统计 假定谱型和测序得到的基因型(本试验有 5 个 SNPs 位点的基因型)是一一对应的, 而 SNPs 通常都为双等位基因, 因此, 根据基因型频率和基因频率的定义以及它们之间的关系, 提出以下公式估算各 SNPs 位点的等位基因频率:

$$F_i = (\sum A_{ii} + 1/2 \sum A_{ij}) / n,$$

式中, F_i 表示某 SNP 位点等位基因 i 的频率, A_{ii} 和 A_{ij} 分别表示品种内具有不同谱型的相应 SNP 位点为纯合(ii)和杂合(ij)的个体数(即 DHPLC 检测得到的 a、b、c、d 谱型的个体数, SNP 位点是杂合或纯合需要结合测序图来确定), n 表示品种内具有不同谱型的总个体数(即 DHPLC 检测具有 a、b、c、d 谱型的总个体

数).

1.6.2 χ^2 检验 对 DHPLC 估算的 -121 位点 C 等位基因频率与 RFLP 得到的频率进行适合度检验, χ^2 值计算公式参照文献[11]的方法.

2 结果与分析

2.1 DHPLC 检测及测序结果

在 DHPLC 检测中, 4 个品种 126 个个体共筛查

到 4 种肩峰或不对称峰的谱型(图 1-a、b、c 和 d)和 1 种单峰谱型, 从每种肩峰或不对称峰谱型中选择 2 个个体进行 PCR 产物测序, 利用 DNASTAR 软件分析测序结果, 发现 5 个 SNPs 位点: G-391A, G-360A, T-359G, G-334A 和 T-121C(图 1). 由测序图可见, DHPLC 分析中出现谱型 a 的片段在相应的 5 个 SNPs 位点分别为 GG、GA、TT、GG、TT; 以此类推 b、c、d 谱型片段其相应的 5 个 SNPs 位点的基因型.

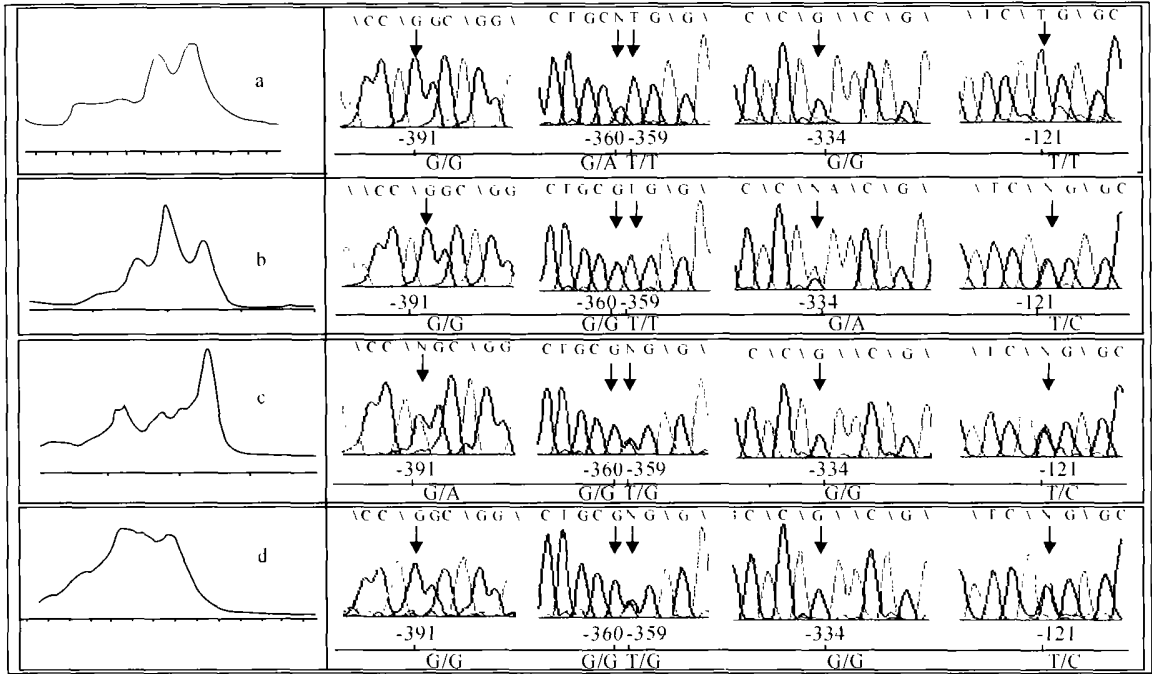


图 1 DHPLC 谱型及测序得到的 SNPs 情况图

Fig. 1 DHPLC chromatography types and corresponding SNPs by sequencing

2.2 谱型统计及等位基因频率估算

根据 DHPLC 筛查的 4 种肩峰或不对称峰谱型, 对 4 个鸡品种 126 个个体进行归类, 结果如表 1 所示. 根据公式 $F_i = (\sum A_{ii} + 1/2 \sum A_{ij}) / n$ 对 T-121C 位点等位基因在杏花鸡中的频率进行估算(缺乏 CC 型纯合子):

$$F_T = (\sum A_{TT} + 1/2 \sum A_{TC}) / n = [7 + 1/2(1 + 4 + 5)] / 17 = 0.71,$$

$$F_C = (\sum A_{CC} + 1/2 \sum A_{TC}) / n = [0 + 1/2(1 + 4 + 5)] / 17 = 0.29.$$

同理计算其他 4 个 SNPs 位点等位基因(-391A、-360A、-359G 和 -334A)的频率, 估算结果如表 2 所示. 从估算的频率来看, 可以初步了解这 5 个 SNPs 等位基因在 4 个群体中的分布情况, 其中 2

个本地品种(丝羽乌骨鸡和杏花鸡)与 2 个引进品种(隐性白和莱航鸡)在 -391、-360、-359 和 -121 位点的等位基因频率有明显的不同, 这可以为以后选择合适的遗传标记提供参考.

表 1 4 个鸡品种中肩峰或不对称峰谱型个体数

Tab. 1 Numbers of each chromatography type in 4 chicken breeds

品种 breeds	谱型 DHPLC profiles				n
	a	b	c	d	
丝羽鸡 Taihe silkies	7	4	5	3	19
隐性白 white recessive rock	7	0	0	1	8
莱航鸡 white Leghom	10	0	0	2	12
杏花鸡 Xinghua	7	1	4	5	17
总计 total	31	5	9	11	56

表 2 4 个鸡品种 5 个 SNPs 位点等位基因频率估算结果及 -121 C 频率的比较

Tab. 2 Estimated allelic frequencies of 5 SNPs in 4 chicken breeds and comparison of allelic frequencies of -121 C

品种 breeds	SNP ¹⁾					-121 C 基因频率 ²⁾		
	-391 A	-360 A	-359 G	-334 A	-121 C	-121 C allelic frequencies		
						RFLP	DHPLC	χ^2
丝羽鸡 Taihe silkies (41)	0.13	0.18	0.21	0.11	0.32	0.24	0.32	0.65
隐性白 white recessive rock(31)	0.00	0.44	0.06	0.00	0.06	0.00	0.06	1.92
莱航鸡 white Leghorn(20)	0.00	0.42	0.08	0.00	0.08	0.00	0.08	1.67
杏花鸡 Xinghua(34)	0.12	0.21	0.26	0.03	0.29	0.35	0.29	0.28

1) 表示 SNP 位点及其中的 1 个等位基因;“-”表示起始密码子之前的位置; 2) 表示 DHPLC 估算的 -121 位点 C 等位基因频率与 PCR-RFLP 检测得到的 C 等位基因频率的比较; 当 $\chi^2 > 3.84$ 时, $P < 0.05$

2.3 T-121C 位点的 PCR-RFLP 分析及与 DHPLC 估算结果的比较

利用 PCR-RFLP 技术对莱航蛋鸡等 4 个群体共 126 个个体进行 T-121C 位点检测分型(图 2), 按照孟德尔基因分离定律计算等位基因频率(表 2)。根据公式计算每个品种 2 种方法得到频数的 χ^2 值, 差异均不显著($P > 0.05$), 说明 DHPLC 估算的 C 等位基因频率与 PCR-RFLP 得到的频率基本符合。

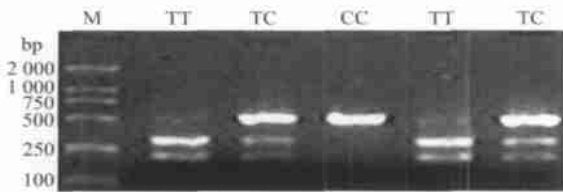


图 2 T-121C 位点的 PCR-RFLP 分析电泳图谱

Fig. 2 Electrophoretic profiles of T-121C by PCR-RFLP

3 讨论

利用 DHPLC 技术, 本试验在鸡生长激素基因 5' 调控区 518 bp 范围内成功检测到 5 个 SNPs; 按照所提出的公式, 估算了该 5 处 SNPs 位点等位基因在 4 个鸡品种的频率, 通过比较, 发现与 PCR-RFLP 所得出的基因频率基本一致, 证实了该估算方法的可行性。按照该方法估算基因频率, 可以初步了解 SNPs 在不同群体中的分布情况, 预测该 SNPs 可能的效应及其应用价值, 为将来选择合适的 SNPs 标记进行遗传多样性和生长等性状 QTL 定位等后续研究奠定基础。

本文估算基因频率是建立在 DHPLC 谱型与测序基因型一一对应的基础上的, 事实上 DHPLC 与 SSCP 一样, 都基于 DNA 构象检测变异, 相对而言, DHPLC 更加灵敏和准确, 与 SSCP 一样将构象型和基因型一一对应可行的^[12]。但是利用该方法进行基

因频率估算也存在一定的偏差, 主要表现为忽略了 DHPLC 检测时为单峰的个体, 因此, 应用这种方法估算的频率是不能用于 QTL 定位或作遗传距离图, 因为基因频率的误差会导致 QTL 定位时置信区间扩大或遗传距离图有误差^[13, 14]。需要指出的是当群体中表现为肩峰或不对称峰的个体数越多, 而表现单峰的个体数越少时, 估算的基因频率会越准确。另外, 虽然理论上 SNPs 可以为二、三、四等位基因, 但实际上人类一般为二等位基因^[15], 所以我们提出的等位基因频率估算方法有普遍的适用性, 有研究表明若结合引物延伸技术(primer extension)和 DNA 池来估算基因频率, 会使估计结果更准确^[16, 17]。

参考文献:

- [1] VIGNAL A, MILAN D, SANCRISTOBAL M, et al. A review on SNP and other types of molecular makers and their use in animal genetics[J]. Genet Sel Evol. 2002, 34: 275-305.
- [2] 聂庆华, 张细权, 雷明明. 单核苷酸多态性及其在鸡 QTL 定位上的应用[J]. 遗传, 2003, 25(6): 729-734.
- [3] KRISTENSEN V N, KELEFIOTIS D, KRISTENSEN T, et al. High-throughput methods for detection of genetic variation[J]. Biotechniques, 2001, 30(2): 318-322.
- [4] SPIEGELMAN J L, MINDRINOS M N, OEFNER P J. High-accuracy DNA sequence variation screening by DHPLC[J]. Biotechniques, 2000, 29: 1084-1090.
- [5] MATYAS G, de PAEPE A, HALLIDAY D, et al. Evaluation and application of denaturing HPLC for mutation detection in marfan syndrome: identification of 20 novel mutations and two novel polymorphisms in the FBN1 gene[J]. Hum Mutat, 2002, 19: 443-456.
- [6] WU G, WU W, HEGDE M, et al. Detection of sequence variations in the adenomatous polyposis coli (APC) gene using denaturing high-performance liquid chromatography[J]. Genet Test, 2001, 5(4): 281-290.
- [7] RAVNIK-GLAVAC M, ATKINSON A, GLAVAC D, et al.

- DHPLC screening of cystic fibrosis gene mutations[J]. Hum Mutat, 2002, 19: 374-383.
- [8] COOKSEY R C, MORLOCK G P, HOLLOWAY B P, et al. Temperature mediated heteroduplex analysis performed by using denaturing high-performance liquid chromatography to identify sequence polymorphisms in mycobacterium tuberculosis complex organisms[J]. J Clin Microbiol, 2002, 40(5): 1 610-1 616.
- [9] UNDERHILL P A, JIN L, LIN A A, et al. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography[J]. Genome Res, 1997, 7(10): 947-949.
- [10] 黄 华, 周宇凤, 江培洲, 等. 变性高效液相色谱筛选鼻咽癌相关基因单核苷酸多态性[J]. 第一军医大学学报, 2002, 22(7): 602-604.
- [11] 董时富. 生物统计学[M]. 北京: 科学出版社, 2002. 172-173.
- [12] BUNN C F, LINTOTT C J, SCOTT R S, et al. Comparison of SSCP and DHPLC for the detection of LDLR mutations in a New Zealand cohort[J]. Hum Mutat, 2002, 19(3): 311.
- [13] 盛志廉, 陈瑶生. 数量遗传学[M]. 北京: 科学出版社, 1999. 340-356.
- [14] NEI M, KUMAR S. 分子进化与系统发育[M]. 北京: 高等教育出版社, 2002. 232-254.
- [15] BROOKES A J. The essence of SNPs[J]. Gene, 1999, 234: 177-186.
- [16] GIORDANO M, MELLAI M, HOOGENDOORN B, et al. Determination of SNP allele frequencies in pooled DNAs by primer extension genotyping and denaturing high-performance liquid chromatography[J]. J Biochem Biophys Methods, 2001, 47: 101-110.
- [17] HOOGENDOORN B, NORTON N, KIROV G, et al. Cheap, accurate and rapid allele frequency estimation of single nucleotide polymorphisms by primer extension and DHPLC in DNA pools[J]. Hum Genet, 2000, 107: 488-493.

【责任编辑 柴 焰】

·简讯·

航天水稻选育新品种——华航一号

“华航一号”是华南农业大学1996年利用返地式卫星将稻种搭载15 d,进行空间诱变选育而成的优质稻新品种,具有优质、高产、抗逆、广适等特性,于2001年通过广东省品种审定,审定编号为:粤审稻200108;并于2003年通过国家品种审定,审定编号为2003032,是我国第一个通过国家级品种审定的航天育种水稻新品种。目前,该品种已经申请国家农作物新品种权保护(申请编号为20040225.0)。该品种属早、晚兼种型,全生育期早造123~126 d,晚造105~108 d。全省各地进行的较大面积的晚造生产表明,其产量、熟期、后期熟色等方面都表现甚佳。几年来,“华航一号”在广东省湛江、汕头、潮州和汕尾等地迅速应用于生产,并推广到广西、海南、江西等省区,截止至2004年,全国推广面积共12.848万 hm^2 ,其中广东省累计推广面积达到8.395万 hm^2 ,产量均超过7 500 kg/hm^2 ,高产超9 000 kg/hm^2 ,最高产量达10 845 kg/hm^2 。

华南农业大学科技处 全 锋供稿