

Single Nucleotide Polymorphism and Its Utilization in Plant Research

Ying Shi, Mu Li, Ping He, Fusheng Li*

Agronomy and Biotechnology College, Yunnan Agricultural University, Kunming
Email: ashyinga@163.com, Lfs810@sina.com

Received: Apr. 22nd, 2014; revised: May 7th, 2014; accepted: May 19th, 2014

Copyright © 2014 by authors and Hans Publishers Inc.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0/>



Open Access

Abstract

SNP (single nucleotide polymorphisms) had become the third generation molecular marker technology in recent years. It had the characteristics of wide distribution, the polymorphism information content big, easy detection and statistics analysis etc. In crops research, SNP had many applications, such as genetic marker, molecular marker assisted selection (MAS), construction of genetic linkage map with high density, mapping expressed sequence tag (EST), and so on.

Keywords

SNP, Molecular Markers, Crops

SNP技术及其在作物研究上的应用

石莹, 李穆, 何平, 李富生*

云南农业大学农学与生物技术学院, 昆明
Email: ashyinga@163.com, Lfs810@sina.com

收稿日期: 2014年4月22日; 修回日期: 2014年5月7日; 录用日期: 2014年5月19日

摘要

SNP(单核苷酸多态性)是最新发展起来的第三代分子标记技术, 具有分布广、多态信息量大、易于检测

*通讯作者。

和统计分析等特点。目前, SNP技术在水稻、玉米、大豆等作物研究中应用较多。本文综述该技术在这些作物的遗传标记、分子标记辅助选择、构建高密度遗传连锁图谱、绘制EST图谱等方面的研究进展。

关键词

SNP, 分子标记, 作物

1. 引言

生物个体之间存在的差异, 实质上是生物个体间基因型的差异, 而 DNA 分子标记可以直接反映基因水平上的差异。随着分子生物学的发展, DNA 分子标记已经逐渐取代蛋白质(酶)水平上的生化标记, 在遗传标记研究中成为新的热点[1]。而目前以 SNP(单核苷酸多态性)分子标记最为研究者所关注。SNP 所反映的 DNA 遗传变异更多体现在单个碱基的变异水平, 由于其有许多优点, 在一定程度上弥补了第一代分子标记(如: 限制性片段长度多态性, RFLP)[2] [3]和第二代分子标记(如: 微卫星 DNA 多态性, SSR)方法的不足, 所以被称为第三代基因遗传标记方法[4]。SNP 可以定位在每一个基因的附近区域, 在对每个基因的等位基因与表现型相关性方面进行检测具有巨大应用潜力[5] [6]。本文主要介绍 SNP 技术及其目前在一些主要作物研究上的应用情况。

2. SNP 技术简介

单核苷酸多态性(Single Nucleotide Polymorphisms, SNP)指在基因组上单个核苷酸变异所形成的遗传变异, 理论上包括单碱基的缺失、转换、插入及颠换等, 但实际发生目前有两种: 转换和颠换, 转换是嘌呤突变成嘌呤(或嘧啶突变成嘧啶), 颠换指嘧啶和嘌呤之间的转换, 而且颠换发生频率是转换的 1/2[7]。其突变的碱基可以是 A、T、G、C 中任何一个, 但就目前观察到的情况而言, SNP 基本都是发生在相连碱基之间, 而且多数为 C 和 T 之间[4]。

SNP 的特点可归纳为以下 6 个方面: 1) 遗传稳定性高; 2) 位点丰富且分布广泛, 如人类基因组中大约每 1000 bp 出现 1 个 SNP[8]、其他哺乳类动物中每 500~1000 bp 出现 1 个 SNP[9], 而在水稻中则每 232 bp 出现 1 个[10]; 3) 位于基因编码区的 SNP 可导致蛋白质功能的改变, 是生物发生变异或病变的可能的直接原因, 因此富有代表性; 4) 二态性和等位基因性: SNP 一般有 2 种等位型的碱基组成, 具有二态性, 且是双等位的, 故杂合期望值较低; 5) 可快速检测, 易实现自动化分析; 6) 密度高、分布广。

SNP 有不同的分类: 1) SNP 在基因组分布的位置: 基因间 SNP(iSNP)、基因编码区 SNP(cSNP)、基因周边 SNP(pSNP)[11]; 2) SNP 对生物遗传性状的影响: 蛋白编码 SNP、非蛋白编码 SNP, 而在蛋白编码 SNP 中, 如果不引起所编码氨基酸序列的改变, 称为同义蛋白编码 SNP, 反之, 则称非同义蛋白编码 SNP[7] [12]-[14]。

SNP 分析技术按其研究对象分为: 1) 对未知 SNP 进行研究, 即寻找未知的 SNP 或确定某一未知 SNP 与某遗传病的关系; 2) 对已知 SNP 进行分析, 即对不同群体 SNP 遗传多样性或在临床上对已知治病基因的遗传病进行基因诊断。检测未知 SNP 的方法有: 变性梯度凝胶电泳(DGGE)、限制性片段长度多态性(RFLP)、温度梯度凝胶电泳(TGGE)、随机扩增多态性 DNA(RAPD)、变性的液相色谱检测(DHPLC)、单链构象多态性(SSCP)等; 检测已知 SNP 的方法有: 核苷酸片段分析(ASO)、错配扩增突变分析(MAMA)、(基因芯片, gene chips)等。

3. SNP 标记在作物研究中的应用

近年来,已在不少作物中发现了 SNP 位点并进行具体应用。例如,玉米基因组中含有丰富的 SNP。Selinger 等对 11 个玉米栽培种和 7 个原始种进行 SNP 鉴定,在调节基因中发现 116 个 SNP[15]。Rafalski 等报道,玉米基因组序列中平均每 70bp 有 1 个 SNP[16]。在水稻中,最高 SNP 数量可达 20127 个,每 154 个碱基就存在一个 SNP,在水稻亚种内的品种间也含有丰富的 SNP[17]。Nasu 等在 3 个粳稻种,2 个籼稻种和 1 个野生稻种的 DNA 区段中检测,在约 250,000 个碱基的序列中一共找到 2800 个 SNP,平均每 89 个碱基就会发现一个 SNP[18]。Zhu 等人对 22 个有差异性表型的大豆基因型序列多态性的研究表明编码区和非编码区均有 SNP 出现,且出现频率是:编码区少于非编码区[5]。Yang 等[19]在对番茄 SNP 标记开发研究中,利用 3 个品种的 EST 数据,在 44 个基因中发掘了 101 个候选 SNP,经实验检测证明其中的 83%是真实的。Bundock 等的研究结果表明,在大麦的某些基因中也存在 SNP,频率为每 131 个碱基就出现一个 SNP[20]。以上研究表明,作物基因组序列均含有丰富的 SNP,对其进行 SNP 研究,有利于揭示作物生长发育规律。目前,SNP 技术在作物研究中的应用可归纳为以下几个方面。

3.1. 遗传标记

遗传标记(genetic markers)是遗传分析上用作标记的基因,它在发展过程中形成了四种不同类型的标记:形态学标记、细胞学标记、生化标记和分子标记。其中,分子标记是目前较受欢迎的,已得到广泛的应用,主要有 RFLP、RAPD、SSR、AFLP 等,但这些分子标记都有各自的优缺点。如:试验发现 SSR 不适用于关联分析(association analysis)[21],所以 SSR 分子标记的使用受到限制,但 SNP 正好可以弥补 SSR 标记所受到的限制,而且 SNP 是双等位的,杂合期望值较低,SNP 的杂合期望值是 0.263,而 SSR 的却达到 0.77[22] [23]。研究发现,两个基因座上的等位基因不是随机组合和分离的[24]。如,玉米的一个基因座上,只在很少数的碱基对区间上就能鉴别出 SNP 单元型,而这个基因组片段所鉴别出来的 SNP 基因座是连锁不平衡(Linkage Disequilibrium, LD)的[7] [23] [25]。近年来,随着某些作物种质多样性的持续减少,LD 增长,方便了目的基因座上 SNP 的单元型与表型之间相关性的分析[26]。此外,因为大豆是自交的,使其具有高度的连锁不平衡。如果有足够多的连锁不平衡存在于一个作物的基因组中,全基因组扫描会有利于鉴别与目的性状相关的基因组区域[23]。此策略的采用,为将 SSR 标记转换为 SNP 标记开辟了道路。如:郝岗平等(2004)在拟南芥中抗旱基因 *CBF4* 单核苷酸多态性的研究分析,显示出各拟南芥抗旱生态型品种之间的表型差异与 SNP 的相关性;韩继成等(2003)对水果基因组 DNA 中乙烯受体基因保守序列进行 SNP 分析,发现了 19 个 SNP 的位点可能与呼吸跃变型果实有关;Kanazin 等(2002)在大麦中 112 个 SNP 的进化研究发现大麦野生种与大麦栽培种间可能存在多重驯化行为。

3.2. 分子标记辅助选择

依赖于植株的表现型选择的传统育种遇到了很多困难,一个优良品种的培育往往需要花费十几年的时间。为了提高选择效率,育种专家在实践中不断探索运用遗传标记来进行育种选择。因为能利用的 SNP 标记的具有分布广和数量多的优势,所以,和其他的遗传标记相比较,SNP 标记在标记辅助选择(marker-assisted selection, MAS)育种方面更具有潜力。如:Hayashi[27]使用等位基因特异 PCR 方法进行水稻抗瘟病基因 *Piz* 位点的等位基因间 SNP 基因分型,结果表明利用这种 PCR 方法对 SNP 进行基因分型在水稻 MAS 育种中是一个十分方便、有价值的工具;Huang X 等[28]对 514 份水稻核心群体的 14 种不同农艺性状的全基因组 SNP 进行相关分析,揭示了不同的水稻种质农艺性状的表型差异与 SNP 的相关性;随后[29]又通过进一步对 950 份水稻种质资源开花期及产量相关性状进行全基因组 SNP 关联分析,揭示了其相关性。

3.3. 构建高密度的遗传连锁图谱

如果要绘制高分辨率的遗传图谱,就要构建大小不受限制的作图群体。基因组中的 SNP,因为其分布的广泛性及其双等位性等,使 SNP 非常适合用于自动化大规模扫描,成为了最受欢迎的作图标记,而且具有高密度性的 SNP 遗传图谱的建成,使我们可以更精确的进行标记辅助选择(MAS),降低或消除在目的基因之外的遗传背景带来的坏的影响[30]。大豆、番茄、玉米、小麦等其它重要作物的 SNP 遗传图谱的构建工作已取得重大进展。

物理图谱的组成是因为一系列 BAC 克隆(按顺序排列的)的重叠,在将传统遗传图谱和物理图谱整合时,需要筛选和检测 BAC 克隆的末端,用来筛选出没有重复序列的末端[23]。鉴别出的序列用于找出父本和母本基因型间具有多态性的 SNP,绘制 SNP 图谱。如,玉米中,1/5 的 BAC 末端序列可用于传统遗传图谱与物理图谱的整合[31]。

3.4. 绘制 EST 图谱

表达序列标签(Expressed Sequence Tag, EST)是遗传物质 DNA 与蛋白质之间的过度产物,反映了遗传信息传递过程的[23] [24]。EST 来源于一定环境下一个组织总 mRNA 所构建的 cDNA 文库,而 EST 最后会形成蛋白质,决定了生命的差异现象,因此,EST 能更有效地反映遗传信息对生命所起到的作用[24]。

SNP 反映的是相对应染色体基因座上的遗传多态性的状态,所以可以用来绘制遗传图谱[23]。利用大豆、玉米、水稻等作物的 SNP 多态性,就易绘制出 EST 图谱。通用的技术方法是先对作图群体父本和母本的目的基因 3' 端的翻译区进行单核苷酸多态性分析,然后对上诉群体中所有个体进行单核苷酸多态性基因型的检测[1] [23] [24] [32]。例如, Bhatramakki 等人利用 B73 × Mo17 这个分辨率高的群体和单核苷酸多态性焦磷酸测序技术来将玉米 EST 谱的绘制完成。相同的,缺失或插入多态性也能用来绘制 EST 谱[33]。目前, Useche 等正在从玉米 EST 数据库中大规模开发 SNP[34]。

4. 小结

4.1. SNP 技术存在的问题

虽然 SNP 作为新一代遗传标记对各方面的发展都有应用价值,但就现在来说,SNP 技术应用的制约因素很多。第一,全基因组 SNP 标记的开发仍有赖于基因组草图搜索法,这种开发策略只能应用于少数动植物种类。第二,SNP 标记的开发费用仍相对较高。第三,大规模的进行 SNP 的分析需 SNP 检测技术的进一步成熟。第四,缺乏大规模和具有普遍应用价值的 SNP 数据库,是科研工作者获取 SNP 位点信息受限。第五,SNP 在植物研究方面仍处于起步阶段。

4.2. 展望

具有多态性丰富、密度大、具有代表性和检测速度快等特点的 SNP,在生物领域有广阔的应用前景和发展潜力[35]。首先,由于生物技术的发展,SNP 在某些方面弥补了前两代分子标记方法的不足之处。其次,由于 SNP 检测和分析技术的逐渐成熟和不断完善,上述问题也会逐步解决,使实验结果更加可靠。相信在不就的将来,SNP 作为新一代的分子遗传标记会对生物各个领域产生较为深远的影响。

基金项目

云南省现代农业甘蔗产业技术体系建设专项(2009~2014);云南省重点新产品开发计划(2012BB014);云南省创新团队之“高原山地特色作物种质创新及利用”项目(云科人发[2012]18号)。

参考文献 (References)

- [1] 许阳, 李东潮, 李晓丽, 等 (2004) 单核苷酸多态性在大豆育种中的应用. *安徽农业科学*, **5**, 1000-1020.
- [2] 程备久, 凌杏元 (1994) 利用 RFLP 进行数量基因定位及效应分析的原理和方法. *生物数学学报*, **5**, 200-206.
- [3] 曹永国, 王国英, 王守才, 等 (1999) 玉米 RFLP 遗传图谱的构建及矮生基因定位. *科学通报*, **40**, 2178-2182.
- [4] 孙红霞, 杜玮南, 方福德 (2000) 单核苷酸多态性研究进展. *中国医学科学院学报*, **4**, 392-394.
- [5] 刘惠民, 杜春芳, 李润植, 等 (2003) 单核苷酸多态性在作物遗传及改良中的应用. *遗传*, **6**, 735-739.
- [6] Liu, W.Q. and He, L. (1998) SNP: Tracing new blueprint for human genome. *Hereditas*, **20**, 38-40.
- [7] 王艳芳 (2007) 针茅属植物水孔蛋白基因多态性分析及分子进化的研究. 硕士论文, 内蒙古农业大学, 呼和浩特.
- [8] Selinger, D.A., Handler, C. and Bolivia, V.L.B. (2001) An allele of the *aizel1* gene with variable expression, contains a high copy retro transposon related sequence immediately upstream. *Plant Physiology*, **1**, 1363-1379.
- [9] 曾燕如, 黄敏仁, 王明麻 (2003) 一种新的分子标记——单核苷酸多态(SNP). *南京林业大学学报*, **3**, 84-88.
- [10] Rafalski, A. (2002) Applications of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biotechnology*, **5**, 94-100.
- [11] Wang, D.G., Fan, J.B. and Siao, C.J. (1998) Large scale identification, mapping, and genotype of single nucleotide polymorphism in the human genome. *Science*, **280**, 1077-1082.
- [12] Gottgens, B., Barton, L.M., Gilbert, J.G., et al. (2000) Analysis of vertebrate SCL loci identifies conserved enhancers. *Nature Biotechnology*, **18**, 181-186.
- [13] Loots, G.G., Locksley, R.M., Blankespoor, C.M., et al. (2000) Identification of a coordinate regulator of interleukins 4, 13 and 5 by cross-species sequence comparisons. *Science*, **288**, 136-140.
- [14] Tao, H., Cox, D.R. and Frazer, K.A. (2006) Allele-specific KRTI expression is a complex trait. *PLOS Genetics*, **2**, 93.
- [15] 刘传光, 张桂全 (2006) 水稻单核苷酸多态性及其应用. *遗传*, **28**, 737-744.
- [16] Nasu, S., Suzuki, J., Ohta, R., Hasegawa, K., et al. (2002) Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. *DNA Research*, **9**, 163-171.
- [17] Yang, W.C., Bai, X.D., Kabelka, E., et al. (2004) Discovery of single nucleotide polymorphisms in *Lycopersicon esculentum* by computer aided analysis of expressed sequence tags. *Molecular Breeding*, **14**, 21-34.
- [18] Bundock, C., Christopher, T., Egger, P., et al. (2003) Single nucleotide polymorphisms in cytochrome *P450* genes from barley. *Theoretical and Applied Genetics*, **106**, 676-682.
- [19] Phillips, R.L. and Vasil, I.K. (2001) DNA-based markers in plants. Knuckler Academic Publishers, Dordrecht.
- [20] Powell, W., Morgante, M., Andre, C., et al. (1996) The comparison of RFLP, RAPD, AFLP and SSR (micro-satellite) markers for germ plasm analysis. *Molecular Breeding*, **2**, 225-238.
- [21] Viard, F., Franck, P., Dubois, M.P., et al. (1998) Variation of micro-satellite size homoplasy across electrophoretic morphs, loci, and populations in three invertebrate species. *Journal of Molecular Evolution*, **47**, 42-51.
- [22] Taramino, G. and Tingey, S. (1996) Simple sequence repeats for germ plasm analysis and mapping in maize. *Genome*, **39**, 277-287.
- [23] 任启军 (2004) 基于棉花微卫星序列的 SNP 变异和系统进化. 硕士论文, 南京农业大学, 南京.
- [24] 梁惠珍, 李卫东, 王辉, 等 (2004) SNPs 在大豆等作物遗传及改良中的应用. *农业生物技术科学*, **5**, 37-53.
- [25] 郑璐璐 (2008) SNAP 分子标记检测玉米 *Ael*, *Sul*, *D9*, *ZmRap* 基因多态性. 硕士论文, 华中师范大学, 武汉.
- [26] 聂庆华, 张细权, 雷明明 (2003) 单核苷酸多态性及其在鸡 QTL 定位上的应用. *遗传*, **6**, 729-734.
- [27] Hayashi, K., Hashimoto, N., Daigen, M., et al. (2004) Development of PCR-based SNP markers for rice blast resistance genes at the *piz* locus. *Theoretical and Applied Genetics*, **108**, 1212-1220.
- [28] Huang, X., Wei, X., Sang, T., et al. (2010) Genome-wide association studies of 14 agronomic traits in rice landraces. *Nature Genetics*, **42**, 961-969.
- [29] Huang, X., Zhao, Y., Wei, X., et al. (2012) Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm. *Nature Genetics*, **44**, 32-41.
- [30] 郝岗平, 杨清, 吴忠义, 等 (2004) 植物的单核苷酸多态性及其在作物遗传育种中的应用. *植物学通报*, **5**, 618-624.
- [31] Meyers, B.C., Tingey, S.V. and Morgante, M. (2001) Abundance distribution and transcriptional activity of repetitive

elements in the maize genome. *Genome Research*, **11**, 1660-1676.

- [32] 王萱 (2008) 玉米直链淀粉扩充基因 ae SNP 位点查找及 ae 基因对有关性状的影响分析. 硕士学位论文, 甘肃农业大学, 兰州.
- [33] Bhatramakki, D., Dolan, M., Hanafey, M., et al. (2002) Insertion deletion polymorphisms in 3' regions of maize genes occur frequently and can be used as highly informative genetic markers. *Plant Molecular Biology*, **48**, 537-547.
- [34] Useche, F., Morgant, M., Hanafey, M., et al. (2001) Computer Detection of Single Nucleotide Polymorphisms (SNP) in Maize ESTs. *Plant and Animal Genome*, **IX**, 333.
- [35] 唐立群, 肖层林, 王伟平 (2012) SNP 分子标记的研究及其应用进展. *中国农学通报*, **12**, 154-158.