生物多样性 2003，11(5)：370～382 ［创刊 10 周年纪念特刊］
Biodiversity Science

新一代分子标记——SNPs 及其应用

邹喻苹 葛 颂
（中国科学院植物研究所系统与进化植物学重点实验室， 北京 100093）

摘要：单核苷酸多态性（SNPs）是广泛存在于基因组中的一类 DNA 序列变异，其频率为 1% 或更高。它是由单个碱基的转换或颠换引起的点突变，稳定而可靠，并通常以二等位基因的形式出现。采用生物芯片和 DNA 微阵列技术来检测 SNP，便于对基因组进行大幅度和高通量分析。因此，作为新一代分子标记，SNP 在生物学诸多领域具有广阔应用前景。本文简要叙述 SNPs 技术的发展历史、研究动态以及相关的理论，介绍了与 SNPs 相关的基本术语、概念及其特点，列举了发现与检测 SNPs 主要技术的原理和方法，同时还根据一些具体实例介绍了 SNPs 在模式动、植物遗传图谱构建、品种鉴定、物种起源与亲缘关系、连锁不平衡与关联分析及其在群体遗传结构及其变化机制研究中的应用。最后展望了 SNPs 在群体遗传、分子育种和生物进化等研究领域中的应用前景。

关键词：单核苷酸多态性，原理和方法，现况和进展

中图分类号：Q3-3 **文献标识码**：A **文章编号**：1005－0094(2003)05－0370－13

A novel molecular marker—SNPs and its application

ZOU Yu-Ping, GE Song
Laboratory of Systematic and Evolutionary Botany, Institute of Botany, Chinese Academy of Sciences, Beijing 100093

Abstract: Single nucleotide polymorphisms (SNPs) are an abundant form of DNA variation which have a frequency of 1% or more throughout the genomes. SNPs consist of a single nucleotide base alteration including transition and transversion. They are stable and reliable mutation and are frequently referred to as bi-allelic makers. SNPs can be used conveniently for large-scale and high throughput genome analysis, in particular combining DNA chips and microarrays techniques. Therefore, SNPs provide a novel molecular marker system potentially useful for a wide range of biological disciplines. Here we briefly introduce the history and developments of SNP techniques, including its basic concept, its discovery and screening. We also discuss its applications in different research areas such as genetic mapping in mode animals and plants, DNA fingerprinting and its application in variety identification, species origin and relationship, linkage disequilibrium and associate analysis, and its application in population genetics. We anticipate that SNP markers will contribute greatly to the studies on population genetics, molecular breeding as well as evolutionary biology.

Key words: single nucleotide polymorphisms, concept and principle, application

1 前言

生命世界存在着极其丰富的多样性，这是人类早就认识到但迄今仍在试图解释的现象。如今，随着分子生物学的不断发展，来自分子水平的信息，尤其是海量的 DNA 序列数据为人们进一步认识和解释生命多样性提供了前所未有的机会。通常，基因组 DNA 序列存在三种类型的天然变异：单个核苷酸替换、一段核苷酸的插入或缺失以及卫星 DNA 重复次数的差异。1994 年，单核苷酸多态性（Single Nucleotide Polymorphism，简称 SNP）这个术语第一次出现在人类分子遗传杂志上，随后 Lander（1996）第一

基金项目：国家自然科学基金项目（30025005）和国家重点基础研究发展规划“973”项目（G2000046805）
收稿日期：2003－04－09；接受日期：2003－06－19
作者简介：邹喻苹，女，1939 年出生，研究员，主要研究方向为植物生理生化。E-mail：zouyp@ns.ibcas.ac.cn

次正式提出 SNPs 为新一代分子标记。1998 ~ 2002 年，每年召开一次“ SNPs 与复杂基因组 ”国际会议，其宗旨是探讨 SNPs 在复杂基因组研究中应用的可能性，其内容包括方法、应用与伦理。自 1999 年第二次会议开始，新增加了群体遗传学与信息学内容。近几年先后用 SNPs 构建了密度为 2 cM（Wang *et al.*，1998）以及含 142 万个 SNPs、密度达 1 SNP/1.9 kb 的人类遗传图谱（Sachidanandam *et al.*，2001）。2002 年 11 月，由美、英、中、日、加等 5 国参与的人类单倍型（haplotype）图谱正式启动，预计 3 年内完成。这些图谱对开展致病基因的定位和人类起源与进化的研究非常重要。

受到“ 人类基因组研究 ”巨大成果的刺激，近几年人类 SNPs 的研究变得异常火爆。2001 年公布的人类基因组草图的初步分析表明，人类基因组含 3 万个左右的基因，95% 的基因组区域仍是一片有待开发的荒漠，该区域的序列必定具有重要的生物功能，隐藏着复杂的科学问题和巨大的商机。尽管人与人之间基因组共性极大，但还是存在着遗传差异，从而导致了人类群体的遗传多样性，形成不同个体或群体对疾病的易感性不同，对药物和环境攻击的反应也不同。因此如何采用大规模和高通量的方法来发现、检测、研究与利用 SNPs，已引起了科学家和商业公司的极大兴趣和激烈角逐，其实质是一场“ 基因争夺战 ”。美国国立生物技术信息中心（NCBI）已建立了 SNPs 公用数据库，SNPs 登记数目成几何级数增加。截止到 2002 年 2 月，注册数已达 412 万。

随着人类 SNPs 研究的迅猛发展，动植物 SNPs 的研究和开发也受到广泛的关注，并在相关的研究机构中迅速开展起来（Primmer *et al.*，2002；Rafalski，2002a），尤其是在模式动植物和重要农作物中已取得了可喜的进展（Lindblad-Toh *et al.*，2000；Hoskins *et al.*，2001；Rafalski，2002a，b；Ching *et al.*，2002）。Henry *et al.* 在 2001 年编著的新书《Plant Genotyping：The DNA Fingerprinting of Plant》中，评述了近几年来植物基因型分析中 SNPs 的研究和应用成果，Rafalski（2002a）也综述了 SNP 在作物遗传研究中的应用。除了制作遗传图谱以外，SNPs 也能用于种质资源的 DNA 指纹分析和生物多样性检测，用于连锁不平衡的关联分析等。由于 SNP 检测与分析技术的飞速发展，特别是与 DNA 微阵列和芯片技术相结合，使其迅速成为继 RFLP 和微卫星标记（SSRs）之后最有前途的第三代分子标记，正在生物医学、农学、生物进化等众多领域发挥着巨大的作用。本文力图对 SNPs 的基本概念和特点作一简要介绍，列举发现和检测 SNPs 的主要技术，重点评述 SNPs 在植物各研究领域中的应用现状和前景，以期推动国内该领域研究的发展。

2　SNPs 的基本概念和特点

2.1　什么是 SNPs

SNPs 是 Single Nucleotide Polymorphisms 的缩写，称为单核苷酸多态性。简单地讲，它是指基因组 DNA 序列中由于单个核苷酸（A，T，C，G）的替换而引起的多态性。因此，通常所说的 SNPs 不包括碱基的插入、缺失以及重复序列拷贝数的变化（Brookes，1999）。一个 SNP 表示在基因组某个位点上有一个核苷酸的变化，主要由单个碱基的转换（以一种嘧啶置换另一种嘧啶 C↔T 或一种嘌呤置换另一种嘌呤 A↔G）以及颠换（嘌呤与嘧啶互换，C↔A，C↔G，A↔T）所引起。具有转换型变异的 SNPs 约占 2/3，其他几种 SNP 在相似的水平上，因为 CpG 二核苷酸的胞嘧啶是人类基因组中最易发生突变的位点，其中大多数是甲基化的，可自发地脱去氨基而形成胸腺嘧啶。在任何已知或未知的基因内或附近都可能找到数量不等的 SNPs，根据它们在基因组分布的位置可分为基因编码区 SNPs（cSNPs）、基因周边 SNPs（pSNPs）和基因间 SNPs（iSNPs）等三类（Wang *et al.*，1998）。总的说来，cSNP 比较少，因为在外显子内的变异率仅占周围序列的 1/5，但它在遗传病和育种的研究中却具重要意义，因此倍受关注。根据对遗传性状的影响，cSNPs 又可分为两种：一种是同义 cSNPs（Synonymous cSNPs），即 SNP 所致编码序列的改变并不影响其所翻译的氨基酸序列，突变碱基与未突变碱基的“ 含义 ”相同；另一种是非同义 cSNPs（non-synonymous cSNPs），即指碱基序列的改变将导致编码氨基酸的改变从而产生蛋白质序列的改变，可能最终影响到蛋白质的功能。大多数 SNPs 位于基因组的非编码区，它们对个体的表现型是无意义的，但对群体而言，这些 SNPs 作为遗传标记在群体遗传和生物进化的研究中却很有用（Syvanen，2001）。在描述 SNPs 时，当前的一致意见是用术语“ haplotype ”（单

倍型)代替术语“allele”(等位基因)。单倍型定义为在给定的一条染色体的紧密连锁的位点上多个等位基因的集合,通常3~4个相邻等位基因彼此靠近而构成的单倍型可作为一个整体而遗传(称为单倍型块,haploblock)(Edwards *et al.*, 2001;Rafalski, 2002),见表1。

从表1可见,在所研究玉米的8个基因型中有3个单倍型(H1,H2,H3),它们分别由3~4个相邻的SNP构成。H2和H1比较,在外显子IV的57处(C→T)、内含子IV的232(C→T)和236处(C→T)有3个SNPs,这两个单倍型至少是3个等位基因的集合。H3和H2比较,在碱基130处(T→C)、232(C→T)、165(A→T)、236处(C→T)有4个变异,它们是4个等位基因的集合。

2.2 SNPs是基因组中分布最广泛而稳定的点突变

在人类群体中,SNPs的频率约占1%甚至更高。据估计,在人类DNA的30亿对碱基中,平均每300~1000个碱基对就有1个SNP发生,因此一个个体应至少携带300万个SNPs,全部人群的SNPs总数可达千万(Brooks,1999)。在植物基因组中特别是在玉米和水稻中,几十个或100多个碱基出现1个SNPs,其频率超过人类和果蝇,可见SNPs是生物基因组中最大量、最普遍的点突变(Rafalski, 2002a;Nasu, 2002)。SNPs又是一种古老而高度稳定的突变(尤其是处在编码区的cSNP),因此对于大规模的研究而言,它比微卫星标记和其他的重复序列更可靠,后者在经历了少数几代后,就会明显发生变化(Marshall, 1997)。

2.3 SNPs是二等位基因分子标记

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很罕见,故SNPs通常被简单地称为二等位基因分子标记(Brookes, 1999)。因此,在对基因组筛选时往往只需对SNPs进行+/-的分析,而无需进行DNA片段的长度分析,这就有利于发展自动化技术来筛选或检测SNPs。就生物技术本身的发展而言,当今在许多实验室都拥有多种分子标记技术,然而检测大多数分子标记的方法如RFLP、RAPD、AFLP、SSR等都要依赖DNA片段在Agarose胶或PG胶上的电泳分离。尽管毛细管电泳和荧光DNA分析技术能同时进行多位点检测,它们能达到相当高的自动化程度,但毕竟都未摆脱“凝胶电泳系统”这个瓶颈。这种基于凝胶电泳基础上的分析仍然是费时和费力的,远远不能满足大规模和高通量对基因型分析的要求。SNPs本身的特点使其具备其他分子标记无可比拟的优越性,奠定了应用DNA微阵列、DNA芯片等高新技术来发现与检测生物基因组之间差异的基础。然而,SNPs的二态性也使它在应用时受到一定的限制。许多评论强调,使用二等位基因的SNPs系统难以取代多等位基因系统如RFLP和SSR。但SNPs在基因组的分布比SSR要广泛得多,因此这个不足可以通过使用更多的位点,加大分布密度来弥补,一张完整的SNPs图谱可以提供远高于基因精确定位所要求的密度。Kruglyak *et al.* (1999)指出,一个二核苷酸重复多态性标记的信息大约是SNPs的

表1 8个用于SNPs分析的玉米基因型间储藏蛋白基因 *globulin*1-s 位点中的单倍型(引自 Bhattramakki & Rafalski,2001)
Table 1 The major haplotypes found in the region spanning intron IV of the *globulin*1-s locus among eight maize genotypes used for SNP analysis

单倍型 Haplotype	基因型名称/编号 Genotype name/no.	多态碱基的位置 Polymorphic base position						
		外显子 Exon IV			内含子 Intron IV			外显子 Exon V
		57	130	165	232	236	274	374
H1	H-4	T	T	A	T	T	ND	C
	H-3	T	T	A	T	T	ND	C
	H-1	T	T	A	T	T	ND	C
	E-2	T	T	A	T	T	ND	C
	B-73	T	T	A	T	T	ND	C
H2	MO17	C	T	A	C	C	D	C
	B-1	C	T	A	C	C	D	C
H3	H60	C	C	T	T	T	D	C

注:所检序列长度为403 bp,只显示多态碱基位置,ND示未检测,D示缺失。

2.25 ~ 2.5 倍，也就是说一个有 900 ~ 1000 个均匀分布的 SNPs 图谱在进行基因组扫描时，它所能提供的信息量就足以和目前最常用的有 400 个标记位点的多态性图谱的信息量相当，而且 SNPs 的检测速度很快，最终能取代 SSR，用于编码复杂性状的多基因研究。

3　SNPs 的发现与鉴定

3.1　SNPs 的发现

由于 SNPs 作为一种新型的分子标记在理论研究和实际应用上均具有极大的潜力，国际上学术机构和商业公司均投入了很大的力量寻找 SNPs。目前，可采用几种不同的路线来发现 SNPs，包括直接对 PCR 产物进行测序、用鸟枪法从基因组文库以及 EST 文库筛选序列等（Rafalski，2002）。

DNA 片段的直接测序是发现 SNPs 的最直接和最常用的方法。最好是选择一套差异大但又是近交的纯合个体进行 PCR 反应，它们既能代表待研究群体的多样性，又便于检测个体的多态性水平和清楚地确定单倍型。通常从待测定的基因或 ESTs 提供的序列设计引物，扩增出 400 ~ 700 bp 的片段，对扩增产物进行双链测序。然后，还要对所得序列进行排序并仔细鉴别真正的多态性和由于测序误差而产生的差异。在此，必须强调指出，直接测序鉴定 SNPs 的主要问题是它的误差。一般序列分析的误差率刚好是 1 bp/100 bp，相当于许多植物种内 SNPs 发生的频率（Edwards *et al.*，2001）。如果测序误差没有及时检测出来，将造成在设计等位基因特异的寡核苷酸（Allele-Specific Oligonucleotide 简称 ASO）引物和进行 SNPs 评估方面的浪费。此外，详细的搜索要检测许多基因型，这样就又混合了序列误差。直接测序更为明显的问题是在许多植物种中有杂合体或多倍体存在，在上述两种情况下，实际上经 PCR 得到的是从同源和部分同源的位点而衍生的多个产物，不能准确鉴定碱基的变化，因此事先要对 PCR 产物进行克隆，这样就大大增加了工作量（Bhattramakk & Rafalski，2001）。用上述直接测序方法在美国优良玉米品种资源中发现了高水平的多态性，在非编码区平均每 48 bp 出现 1 个 SNP，在编码区每 131 bp 有 1 个 SNP，在不同群体中多态性比例不同，这些多态性分布在少数几个保守的单倍型中，意味着在美国栽培玉米群体的历史中存在瓶颈效应（Rafalski，2002b）。最近对大豆序列多样性的研究表明，在 22 个多态性高的大豆基因型中，SNPs 的出现频率在编码区每 kb 发现 1.64 个 SNP，在非编码区每 kb 发现 4.85 个 SNP，被测序基因的 3′端有 1/3 含 SNPs（Rafalski，2002a）。

迄今为止，在相关的数据库中存在大量的序列信息（如 ESTs 数据库等），因此采集和分析现有序列数据也是发现 SNPs 的重要方法。拟南芥（*Arabidopsis thaliana*）的全基因组测序已经完成，水稻基因组序列草图已发表，第 1 和第 4 号染色体精确图也已公布，这些都为发现 SNPs 提供了极大的方便。Cereon 公司比较了拟南芥两个生态型基因组数据库的 DNA 序列，发现了 56 000 个多态性，其中 SNPs 将近 40 000。日本植物基因组中心通过分析 3 个日本水稻品种（Nipponbare，Kasalath 和 Kitaake）、2 个印度品种（Kasalath 和 Guang-lu-ai 4）以及野生稻 *O. rufipogon*（WI 1943）基因组 417 个推测的基因间隔区，发现了 2800 个序列变异（包括碱基替换、插入与缺失、SSR），其频率为平均每 89 bp 出现 1 个 SNP，或在两个随机挑选的品系间每 232 bp 出现 1 个 SNP，如此丰富的碱基变异仅次于玉米。我国科学家新近完成了水稻粳稻（*Oryza sativa* ssp. *japonica*）第 4 号染色体的精确测序（Feng *et al.*，2002），在选自该染色体（全长 34.6 Mb）上 2.3 Mb 的片段中鉴定出了 9056 个 SNPs，平均每 268 bp 出现 1 个（Nasu *et al.*，2002）。

EST 数据库是一种新的和有潜力的获得丰富 SNPs 的来源，其优点是包括许多表达基因的序列资料。人们有理由期待至少某些 EST 会涉及重要的农艺性状。与一些无名（anonymous）标记如微卫星标记等不一样，来自 EST 的 SNPs 本身可能就是所研究性状的决定者。再者，从特异和不同的基因型衍生而来的独立的 EST 序列分析，将使得在 DNA 芯片上寻找 SNPs 成为可能，而这是进一步发现植物 SNPs 更好的方法。如今，许多生物技术公司创建了多种农作物如玉米和大豆的 ESTs 的专有数据库，这些数据库是存在于物种间、群体间、个体间多态性的丰富来源。当前面临的挑战是如何充分利用这种资源来挖掘 SNPs，通过数据采掘来鉴定 SNPs 可能是一种比较经济的途径，这种方法已用于人类 SNPs 的鉴定（Piccoult *et al.*，1999）。一些专业化的称为“Polybayes”的软件（Marth *et al.*，1999）可在几

秒钟之内分析大批量样品,为高通量地发掘 SNPs 提供了重要手段(Syvanen, 2001)。

3.2 SNPs 的检测

从技术上来讲,凡是检测点突变的方法都可用于鉴定 SNPs,但迄今为止还没有任何一种方法是万能的。在实际应用时要根据研究目的、实验室设备与技术条件以及经费情况进行选择。在众多的方法中,根据等位基因特异的寡核苷酸(Allelic Specific Oligonucleotide 简称 ASO)PCR 或杂交、限制性位点酶切、寡核苷酸连接以及引物延伸等方法较为常见。有关 SNPs 检测的方法学已有不少很好的评述可供参考(Gut, 2001;Syvanen, 2001)。

等位基因特异的 PCR 或杂交已成为广泛使用的方法,特定地设计 PCR 引物可用于区分 SNPs 等位基因。See *et al.* (2000)在大麦中发展了这种设计引物的方法。为了扩增一个特定位点和准确地区分 SNPs 的等位基因,他们在 PCR 反应中用了 3 个引物 2 个正向引物,在其 5′端分别用不同颜色的荧光标记,在其 3′端分别与 SNPs 二等位基因中的一个相匹配;另有 1 个反向引物。表 2 表示根据大麦基因组位点 *ABG*65 的序列设计的一套等位基因特异引物。

扩增的具荧光的 PCR 产物经电泳分离后可在一种特殊的荧光仪(例如 FMBIO Ⅱ)上检测,也可以直接在 377DNA 自动测序仪上分离和检测。等位基因可根据荧光颜色和 PCR 产物大小记录下来。在此例中绿色荧光的电泳带为等位基因 A,红色荧光带为等位基因 G,若发现同时存在红绿两种荧光的带则为两个品种的杂交后代,或同一野生群体中的杂合个体。由此可见,采用等位基因专一的 PCR 方法确定杂合体是很方便的。

Ausbuel 实验室(Drenkard *et al*., 2001)在构建拟南芥的遗传图谱时发展了另一套简单的等位基因特异的扩增策略,其改进是在引物最后 4 个碱基中引入了另外一个额外的错配碱基,它与原先在 3′端天然存在的错配偶联,导致非特异性等位基因扩增产物大量减少,而对特异性等位基因的扩增影响较小,因此明显增加了引物的专一性。他们称这种改进的方法为 SNAP,并根据一套经验数据编写了一个名为 SNAPER 的计算机程序(http://www. patho. mgh. Harvard. edu/ausbuelweb),以评估不同的外加错配碱基对改变 PCR 扩增的影响。SNAP 策略的优点是:通过在引物的 3′端另外加入错配碱基,大大提高了引物的专一性;只要是每两个等位基因专一的引物都用于其中一个等位基因的扩增,就有可能检查出杂合体或异源双链,在此可作为共显性标记;并可方便地在标准的琼脂糖凝胶电泳上检测,不要求特殊的设备和复杂的方法,因此可普及到任何分子生物学实验室。

等位基因特异的杂交已用于大批量和高通量的 SNPs 鉴定,例如 DNA 芯片的设计和使用。当涉及有关 SNPs 的芯片时,指的是处理基因型分析的芯片,而不是通常所讨论的 EST 基础上的表达芯片。所有基因型分析的芯片的核心是 ASO,它的序列对一个等位基因是专一的。选择简单或复杂的芯片取决于能获得多少覆盖众多位点的等位基因序列(Edwards *et al*.,2001)。在使用较复杂的芯片时须用到 DNA 微陈列技术,例如 ASOs 微阵列的"铺砖途径"(tiling path)是为芸台属(*Brassica*)植物基因型分析芯片而设计的(Lemieux, 2001)。在芯片操作中多用激光共聚胶扫描仪判读芯片上的荧光信号以确定基因型,然后根据一个专有的规则系统进行

表 2 大麦基因组位点 ***ABG*65** 的序列及其内部的 **SNP** 引物(选自 See *et al*., 2000)
Table 2 The sequence and internal primers at the *ABG*65 loci in barley genome

位点 ABG5 **的序列** Sequence at the *ABG*65 loci
5′-TGGCGACTCTGATGCTACGATTGGATCAGCAGAGCCCACAAGTTGGCCCCCGCCTGGGAAGGACCGT
TCATCYCTCTAAGGTGCAGAACAACGGAGCATATCGACTTTACAACCTCGACAGGGAAACGGACGAGCCGC
GAGCATGGAATGGAGATCTACTGAAGCGCTTCTACACATAACCGCCGATAGATCCTCA-3′

位点 ABG **的内部** SNP **引物** Internal primers at the *ABG*65 loci
HEX-CTGGGAAGGACCGTTCATCG
FAM-CTGGGAAGGACCGTTCATCA,

数据分析。可以预见，在不久的将来，拟南芥和主要农作物如玉米、水稻、小麦和大豆的基因型分析芯片会形成商品。由于基因型分析芯片的显著优点，它能在主要农业性状的 ESTs 内检测多样性，毫无疑问将成为大批量、高产率 DNA 指纹的首选方法（Edwards *et al.*，2001；Bhattrmakki *et al.*，2001）。

近几年动态等位基因特异性杂交（dynamic allele-specific hybridization 简称 DASH）受到青睐，此技术是根据在连续升温条件下的荧光强度动态变化而求得不同样品的 Tm 值来分辨不同的基因型（Howell *et al.*，1999；曾朝阳等，2002）。我国湘雅医学院等单位（曾朝阳等，2002）用 DASH 技术对 96 份人外周血样品成功地进行了 2 个 SNPs 位点的基因分型。他们计算了在大规模进行 SNPs 检测时，每份样品的费用仅合人民币 4 元，因此认为 DASH 是一种相对准确、经济、高通量鉴定 SNPs 的技术。

基质辅助激光解吸电离飞行时间质谱法（matrix-assisted laser desorption time of flight mass spectrometry 简称 MALDI-TOF），依赖于扩增的 DNA 片段与特定的 ASO 的忠实杂交，然后洗提下来这种结合的 ASO 用质谱仪分析，是一种强有力的分析 SNPs 的工具，在几秒钟之内可分析大批量样品，但须与高度多重化 PCR 接轨（Griffin *et al.*，1999）。变性高压液相（denaturing high-performance liquid chromatography 简称 DHPLC）有赖于 DNA 同源双链与异源双链之间物理性质的差异，根据异源双链和同源双链在变性反向高压液相离子对柱层析过程中，滞留时间不一致而分离（Oefner *et al.*，1998）。单链 DNA 片段构象变异多态性（Single-Strand Conformation Polymorphism 简称 SSCP）是指由于单个碱基替换而引发的小片段单链 DNA 三维构象的变异而产生的多态性。近几年 SSCP 技术有新的改进，利用变性梯度胶电泳（DGGE）和温度梯度胶电泳（TGGE）可检测 500 ~ 1000 bp 的片段，检出率可达 100%（Etcheid *et al.*，1998）。

此外，基于限制性酶切的 RFLP，PCR-RFLP，CAPS（即 PCR-SSCP）已成为实验室检测 SNPs 的常规手段。用 ddNTP 诱导引物延伸的技术也受到重视，已在大麦育种中应用（Paris *et al.*，2001）。一些公司基于核苷酸连接发展了用于检测 SNPs 的裂解酶和荧光探针例如 TaqMan 荧光探针（PE Biosystems）和称为分子信标（Molecular beacon，MBs）的荧光探针（Tyagi *et al.*，1996）等也用于 SNPs 检测。

4　SNPs 的应用

SNPs 在基因组广泛而稳定的存在，提供了一批很好的分子标记，在高密度遗传图谱构建、性状作图和基因的精确定位、群体遗传结构分析以及系统发育分析等方面均具有广阔的应用前景（Brooks，1999；Rafalski，2002a，b）。下面我们将简要介绍几个主要的应用领域。

4.1　遗传图谱的构建

除了在人类基因组的应用外，SNPs 的信息丰富、稳定遗传、适于做近交系之间的杂交分析等优点同样为实验遗传的模式动物——鼠（*Mus musculus*）的研究提供了有力的工具。Lindblad-Toh *et al.*（2000）报道了在鼠基因组中大规模的 SNPs 分析。他们在 8 个鼠的品系中揭示了核苷酸多态性的比例，采用高密度 ASO 阵列在 1755 个 STSs（序列标记位点，sequence tagged sites）中鉴定了 2848 个 SNPs，其中的 3/4 已用于鼠基因组作图，构建了鼠的第一代 SNP 图谱。近一个世纪以来，果蝇（*Drosophila melanogaster*）的遗传分析已成为研究基因功能的强有力工具，但长期缺乏好的构建分子遗传图谱的标记。继人、鼠、植物的 SNP-遗传图谱之后，Hoskins *et al.*（2001）基于果蝇 STS 的物理图谱，发展了贯穿整个基因组的一套高密度的 474 个 SNPs 标记，用于构建高分辨率的二等位基因遗传图谱。这个新的图谱立即在果蝇研究中起了很好的促进作用。

在模式植物拟南芥（*Arabidopsis thaliana*）以及许多农作物中，一系列重要的生物学过程诸如开花、授粉、发育式样、细胞程序化、抗病机理等研究已经模式化。要迅速有效地鉴定和研究某个物种的变异，必须有一个高密度、含有易于基因型分析的遗传图谱。Cereon Ginomics 公司（2001）公开的信息指出，在拟南芥（*A. thaliana*）的两个生态型序列中，平均每 3.3 kb 有 1 个 SNP，对这个拥有 130 Mb 的基因组而言，这种碱基转换的变异大约为 40 000 个 SNPs。Cho *et al.*（1999）用一种抗真菌病基因 *Eds*16 序列中的 237 个 SNPs 标记在拟南芥（*A. thaliana*）中构建了分辨率为 3.5 cM 的二等位基因遗传图谱，这个过程简要说明如表 3。这是第一次在二倍体植物中构建的二等位基因标记的遗传图谱，它所确立的一系列方法也可用于其他植物的

表 3 在拟南芥(*A. thaliana*)中用于构建图谱的 SNPs 的鉴定(引自 Cho *et al.*, 1999)
Table 3 Characterization of SNPs identified in *A. thaliana* for mapping

SNPs 标记的性质 Nature of SNP markers	占 412 个转换型 SNPs 的百分比 Percentage of AT412 SNPs (%)	总数 Total
DHPLC/测序鉴定 Identified by DHPLC/dideoxy sequencing		487
格式化 412 个 ASO 的微阵列 Formatetted on AT412 oligonucleotide array		412
成功的 PCR 扩增 Successful PCR amplification	95	390
单个扩增后的强杂交信号 High hybridization signal following singleplex amplification	85	351
多重扩增后的强杂交信号 High hybridization signal following multiplex amplification	81	332
识别两个生态型的 Discrimination between Columbia and Landsberg *etecta* homozygotes	64	262
识别纯合子与杂合子的 Discrimination between homozygotes and heterozygotes	57	235
识别纯合子与定位到一条染色体上的 Discrimination between homozygotes and mapping to unique chromosomal position	58	237

SNPs 遗传图谱构建。在 F_2 代收获以后,这个二等位基因图谱构建的整个过程花了不到两天的时间,而要用 RFLP 标记得到同样的结果须花几个月时间(Cho *et al.*,1999)。

在许多物种中发现、收集与鉴定的 SNPs 已经构成了庞大的序列变异数据库,极大地促进了遗传图谱的构建工作,使得原来冗长乏味的事情变得轻松而有趣。在玉米中,保守地估计可获得 2 千万个以上的 SNPs 供分析用,故用 SNPs 构建 EST 遗传图谱变得比较容易(Rafalski,2002b)。为此目的,首先在作图群体的父本、母本中用直接测序方法检测目标基因 3′端未转录区域的 SNPs,然后通过一个记录 SNPs 的流程对子代的基因型作图。Davis *et al.* (2001)通过这个途径成功地用高分辨率的 B73 × Mo17 作图群体(http ://www. agron. missouri. edu/Coop/Conf/2001/03Abstracts. pdf)构建了玉米的遗传图谱。如此超高密度的 SNPs 贯穿整个玉米基因组,也大大促进了基于连锁不平衡的关联图谱构建。

用 SNP 标记还有助于将遗传图谱和物理图谱进行进一步的整合(Rafalski, 2002a)。物理图谱通常是由 BAC 克隆的重叠群组成,这样可以通过检测 BAC 末端的重复序列来发现 SNPs,然后再将这些来自 BAC 末端的 SNP 标记进行遗传作图,从而达到将遗传图谱和物理图谱进行整合的目的。据报道,在玉米中大约 20% 的 BAC 末端序列是单拷贝或低拷贝,因此可以用于图谱的整合(Rafalski, 2002a)。

4.2 DNA 指纹的确立

作为一种新型的 DNA 指纹,SNPs 标记在物种鉴定、物种起源与亲缘关系、遗传育种等领域得到了广泛的应用。Primmer *et al.* (2002)在两种雀型目的食虫鸟(*Ficedula hypoleuca* 和 *F. albicollis*)中揭示了鸟类基因组的高核苷酸多态性。迄今为止在 GenBank 数据库中能查到的鸟类序列信息非常有限,只占脊椎动物序列的 0.4%,而且其中的 2/3 来自一种家养鸡。*Ficedula* 属食虫鸟种的复合体已成为研究物种形成和杂交的重要模式系统,因此研究 SNPs 标记的应用潜力非常重要。他们从其他鸟类现有的序列设计引物,然后用 PCR 扩增同源的食虫鸟序列以及从上述两种鸟的基因组文库得到的随机克隆进行测序两种策略,成功地在两个种种内的约 9.1 kb 序列分别鉴定了 51 个和 61 个 SNPs,其频率分别为 1/175 bp 和 1/150 bp 此外在 17 个同源位点鉴定了两个种种间的碱基差异的频率 > 50%,而在比较的 692 bp 的 SSR 序列中没有种间 SNPs。食虫鸟总的核苷酸多态性为 0.0023 ~ 0.0027,比人类的 SNPs 高 3 ~ 6 倍。鸟类基因组较高的核苷酸多态性可能是由于食虫鸟有效种群的历史比人类古老。

Savolainen *et al.* (2000)用测序和 SSCP 方法比较了异交和自交的拟南芥属(*Arabidopsis*)植物在乙醇脱氢酶基因(*Adh*)位点的核苷酸多态性式样。在此基因区域发现了 8 个不同的插入与缺失,8 ~ 15 个“T”重复以及 15 个核苷酸替换。进一步分析表明,*Arabidopsis lyrata* 中的核苷酸多样性为 0.0038,低于 *A. thaliana* 中的0.0069,但 *A. lyrata* 群体中的序列多样性要比 *A. thaliana* 群体中的高得多。两个种该基因区域中分离位点(segregating sites)的分布也不同,在 *A. lyrata* 的北美群体中存在过量的多态性位点,而在 *A. thaliana* 中却罕见这种现象;两个种多态性的地理分布差异也很大。这些差异是不同的交配系统所致,也反映了 *A. thaliana* 在迅速扩张过

程中瓶颈效应的影响。

Selinger *et al.*（1999）在研究 18 个玉米调控基因（*b* 基因）时，对该基因一个 594 bp 区域测序发现了 116 个 SNPs 和 30 个 1～5 bp 插入和缺失。在栽培与祖先玉米基因型中有 3 个主要的单倍型。*b* 基因调控玉米花青素形成的水平和式样，它的多样性反映了表现型和进化的关系。Wang *et al.*（1999）尝试采用玉米中称为“进化位点”的 *tb*1 基因的 SNP 数据来回答栽培玉米的起源问题。他们从 17 个玉米（*Zea mays* ssp. *mays*）基因型的转录区和非转录的上游区大约 2.9 kb 序列搜寻 SNPs，其中 12 个基因型属于 *Z. mays parviglumis*（玉米的祖先），5 个基因型属于 *Z. mays mexicana*。野生玉米在编码 *tb*1 蛋白的基因部分的序列仅有 3% 变异，而在启动子区域的变异却非常高；但所有栽培玉米在该启动子区域的变异却非常低。因此，他们认为在玉米的长期栽培过程中，影响启动子区域的选择要比影响编码区强得多，这些选择导致玉米表现型的变化。Rafalski *et al.*（1999）对 8 个玉米近交系几百个位点的序列分析表明，每 83 bp 有 1 个 SNPs，每 250 bp 有一个小的插入或缺失，并发现玉米储藏蛋白基因 *glb*1 中仅有少数几个单倍型（见表 1）。因为玉米是一种异花授粉物种，在群体内和群体间如此高的遗传变异主要决定于交配系统的类型。他们观察到，在当今玉米及其祖先中的某些单倍型的保守性与 Selinger *et al.*（1999）所得的结果一致，只不过是从更多的样品中得到证实。Tenaillon *et al.*（2001）对玉米（*Z. mays* ssp. *mays*）1 号染色体上分布的 21 个位点的 DNA 序列多态性式样曾进行过仔细研究。他们对来自美国以外的 16 个地方品种和 9 个美国近交系总共 25 个个体进行了 21 个基因位点的测序，结果表明，在任何两个样品之间，每 104 bp 就有 1 个 SNP。比较地方品种和近交系两者之间的遗传多样性发现，近交系平均保留了地方品种多样性的 77%。Mogg *et al.*（2002）最早用基因型分析芯片来检测玉米 SNPs。他们从 11 个玉米近交系的 2 个基因片段和 52 个 SSR 位点序列的两翼鉴定了 324 个序列多态性，其中含 218 个 SNPs，106 个插入或缺失。该研究的主要目的是找到一种提供大量序列多态性的方便来源，以用于玉米基因型分析。他们根据 32 个 SSR 连锁位点和 2 个基因连锁位点设计和合成了 123 个 20 碱基的 ASOs，这 123 个 ASOs 以一定的格式共价结合到一块玻璃片上，然后这块芯片与 ^{33}P 标记的 PCR 产物杂交，并根据 ASO 的杂交数据构建了 11 个玉米近交系的亲缘关系树状图。

在水稻中，直链淀粉的含量常常是决定烹调和米食加工质量的一个重要指标，含量低者米质好，软、粘而有光泽，含量高者米质发干而松散。负责合成直链淀粉的淀粉合成酶是由 *Waxy* 基因（*Wx*）编码的，曾报道有一个多态的 SSR 序列 $(CT)_n$ 紧密连锁到 *Wx* 基因上。为了确定这个序列的多态性是否与直链淀粉的含量相关，Ayres *et al.*（1997）测试了 92 个水稻品种的家系，鉴定出 *Wx* 的 7 个 $(CT)_n$ 等位基因。为了进一步确定 *Wx* SSR 等位基因与直链淀粉含量之间密切相关性的分子基础，他们对随机选取的 42 个品种中上述 8 个等位基因的 $(CT)_n$ 的 200 bp 片段（在 *Wx* 编码区推测的先导内含子 5′裂解位置）进行了序列分析，发现所有直链淀粉含量低于 18% 的品种其序列都是 AATTATA；而所有直链淀粉含量在中等水平以上的 26 个品种在该位置的序列都是 AAGTATA。这种 TG 转换与直链淀粉含量的相关性在育种实践中很有价值（Ayres *et al.*，1997）。Nasu *et al.*（2002）比较了不同水稻品种的 SNPs 多态性后发现，品种 Kasalath 和野生稻 W1943 之间的多态频率最高（0.75%）；其次是 G4 和 W1943 之间（0.71%）；粳稻品种和 W1943 之间分别为 0.33% 和 0.34%。这些数据表明与野生稻 W1943 遗传关系更近的是日本品种，而不是籼品种。他们还发现水稻 12 条不同染色体分布的式样也不同。两个密切相关的品种 Nipponbare 和 Koshihikari 之间的 94 个 SNPs 能转换成分子标记，并确立了贯穿整个基因组分布的 213 个共显性标记，这些作为分子标记的 SNPs 在水稻基因组研究、群体遗传研究和分子育种中有巨大的潜力。

SNPs 指纹在区分近缘种的研究中非常有效。红云杉（*Picea rubens*）和黑云杉（*P. mariana*）在形态上难以区分，它们和白云杉（*P. glauca*）在北美地区又是分布区重叠的种。为了鉴定 3 个近缘种之间的多态性，Germano *et al.*（1999）从每个种代表不同地理群体的 4 个个体取样，对叶绿体 *trn*K 的内含子、*trn*K、*rpl*33-*psa*J-*trn*P 区域以及核 DNA 的 ITS 区域进行测序，结果发现了 13 个叶绿体的和 12 个核 DNA 的种间 SNPs，其分布如图 1 所示。他们用群体采样方式（样品采自种苗原产地，具有广泛的代表性，

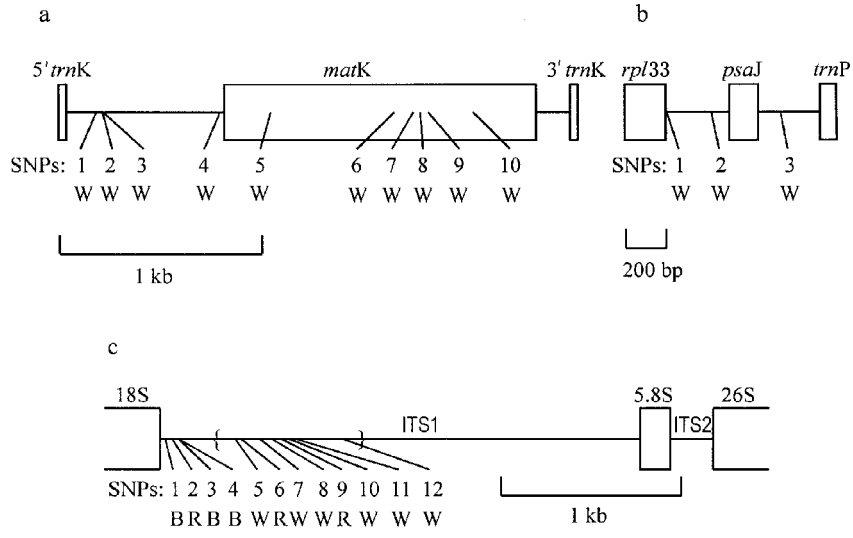

图 1 3 个云杉种间 SNPs 在叶绿体 *trn*K 内含子、*rpl*33-*psa*J-*trn*P 区域以及核 rDNA 的 ITS 区域的分布
长方形盒子示基因编码区：a，*trn*K 区域；b，*rpl*33-*psa*J-*trn*P 区域；c，核 DNA 的 ITS 区域。数字示种间 SNPs 的相对位置；SNP 下面的字母示每个 SNP 能区分的种：W 表明其上的 SNP 将白云杉与红云杉和黑云杉分开，B 表明其上的 SNP 将黑云杉与红云杉和白云杉分开，R 表明其上的 SNP 将红云杉和白云杉分开（根据 Germano J. 1999 改绘）。

Fig. 1 The relative positions of interspecific SNPs in the gene regions (boxed areas) of *Picea* are shown.

a. *trn*k intron region, b. *rpl*33-*psa*J-*trn*P region, c. nuclear ITS region. Numbers indicate the ralative position of interspecies SNPs. Letters below SNPs indicate the species that each SNP distinguishes: W distinguishes white spruce from red and black spruce, B distinguishes black spruce from white and red spruce, R distinguishes red spruces from white and black spruce.

包括来自 11 个原产地的 46 株红云杉，来自 30 个原产地的 84 株黑云杉以及来自 22 个原产地的 90 株白云杉）对几个种专一的叶绿体 SNPs 和 ITS SNPs 通过限制性酶切、SSCP 和 ASO-PCR 方法进行了鉴定。有 2 个 SNPs（1 个叶绿体 DNA 的，1 个核 DNA 的）将黑云杉与红云杉和白云杉分开，在所检测的个体中 96% ~100% 一致。例如，其中的 *trn*K SNP 10 定位于 *mat*K 密码的第 410 碱基位置，在黑云杉中它编码异亮氨酸，而在红云杉和白云杉中却编码亮氨酸；有 5 个 SNPs 将白云杉与红云杉和黑云杉分开（4 个叶绿体 SNPs，1 个核的 SNP），在所检测的个体中 100% 一致。他们还进一步用双盲法对 45 个匿名的红、白、黑云杉样品（从微量针叶提取的 DNA 样品）进行了 SNPs 鉴定。为了鉴定白云杉，用 *trn*K SNP1、2 和 3 进行 SSCP 分析以及 ITS 的 SNP 7 进行 *Bsp*12861 分析；然后用 *trn*K SNP10 的 ASO-PCR 分析以及 ITS SNP1 的 *Bst*U1 分析将黑云杉与白、红云杉区分，其结果 100%（45/45）正确。这些方法简单而灵敏，通常比等位酶、RFLP 和 RAPD 等分子标记更为有效。

4.3 群体遗传和连锁不平衡

群体遗传学研究群体的基因库或者说群体遗传组成及其变化机制。从它的发展历史可以发现，群体遗传学研究在很大程度上依赖于特定的遗传标记，每种新的遗传标记的发现和应用都对其发展产生重要影响（葛颂，洪德元，1994；Buckler & Thornsberry, 2002）。因此，在分子水平上的群体遗传学研究取决于能否得到足够的 DNA 多态性。尽管与一些常用的分子标记（如 SSR）相比，二等位基因的 SNP 显得多态性并不高（期望杂合度较低），但如果考虑许多单碱基位点变异所构成的单倍型时，SNP 所含的信息量就非常大了。因为如前所述，SNP 在基因组中非常丰富，这一丰富的变异性为开展群体水平的研究提供了有力的工具。除了前述进行指纹分析、物种鉴定和种间关系研究外，SNP 标记在群体水平上的应用最具有吸引力的方面是利用群体遗传学中的连锁不平衡原理来进行高密度图谱的构建和进行关联分析（Syvanen, 2001）。从某种意义上说，当前人们对 SNPs 产生如此之大的兴趣在于它作为一种标记，通过连锁不平衡作图法可以用于鉴定导致生物特定性状（如人类疾病）的基因。其基本原理在于首先确定一批按一定间隔存在、覆盖整个基因组的 SNP 标记，然后在特定群体中寻找这些 SNP 标记与待研究特征之间的关系，即确定与特征相关

的 SNP 基因型,从而确定导致生物出现特定性状的基因组区域。这一方法最初提出是用来定位人类疾病基因,作为复杂疾病基因精确作图的工具,随后扩展到在基因组整体水平上的关联研究,它同时还用于人类基因组的重组研究以及探讨人类的起源和进化历史等(Pritchard & Przeworski, 2001)。

在此,我们简要介绍一下与 SNPs 应用直接相关的两个基本概念:连锁不平衡和关联分析。连锁不平衡(Linkage Disequilibrium,简称 LD),指群体内不同位点上等位基因间的非随机组合式样(Goldstein,2001)。由群体遗传学的基本定理 Hardy-Weinberg 平衡可知:如果有两个位点,每个位点上均有两个等位基因 A 和 B,它们在群体中的频率分别为 f(A)和 f(B),那么经过一个世代的随机交配,单倍型 AB 的频率应为 A 频率和 B 频率的乘积,即 f(AB) = f(A)f(B)。如果上述条件不能满足,那么等位基因 A 和 B 就处于连锁不平衡(LD)(Stumpf, 2002)。简言之,连锁不平衡就是不同位点的等位基因之间存在连锁关系而不是自由组合关系。由于位点之间的这种连锁关系,就可以利用连锁不平衡技术通过标记位点的变化来预期其他位点的变异(Brooks, 1999)。关联研究(associate study),又称关联分析,主要是研究群体中的分子变异与表型变异之间的相关关系。关联研究的优点在于不需要构建作图群体,因而对于不能进行控制实验的人类群体非常有效,而且得到结果快,当存在高度 LD 的情况下也有很高的分辨率(Buckler & Thornsberry, 2002)。由此可见,从分子标记的角度看,连锁不平衡与关联研究有着密切关系,关联研究的成效如何在很大程度上取决于群体中连锁不平衡的强弱和式样(Brooks, 1999;Pritchard & Przeworski, 2001)。

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(LD)技术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。在理想的高水平 LD 状况下,可以根据某位点上的等位基因来预测另一个位点上将出现哪一种等位基因,从而通过确定基因组中很小部分区域的多态位点基因型来分析整个基因组大部分区域的变异。但实际情况并非如此理想。因为 LD 是一个十分复杂的现象,诸如突变、选择、随机漂变等遗传因素以及群体的交配系统和进化历史等群体动态因素均对 LD 有很大的影响(Brooks, 1999;Stumpf, 2002)。以人类研究为例,在北欧后裔的人群中 LD 通常可以延伸达 60 kb,而在非洲人群中 LD 要小得多。这种差别大概反映了北欧人在历史上经历过严重的瓶颈效应(Rafalski, 2002a)。

在植物遗传研究中,重组自交系被成功地用于 10 ~ 30 cM 区域的 QTL 作图,但建立在 LD 基础上的关联研究可以确定 QTLs 所代表的实际基因,提供了比遗传图谱要高得多的分辨率(Remington *et al.*, 2001)。Tenaillon *et al.* (2001)收集了美国以外 16 个地理品种和 9 个美国近交系共 25 个个体,并对第一号染色体上的 21 个基因位点进行了测序和各种统计分析。结果发现,在玉米中 LD 消失得特别快,只在平均几百个碱基内存在(见图 2),不同位点之间不存在 LD。他们还发现,在所检测的 20 个位点中,11 个在 250 bp 的距离内 LD 降到 p = 0.2,5 个在 500 bp 内 LD 的 p 值降到 0.2,还有 4 个

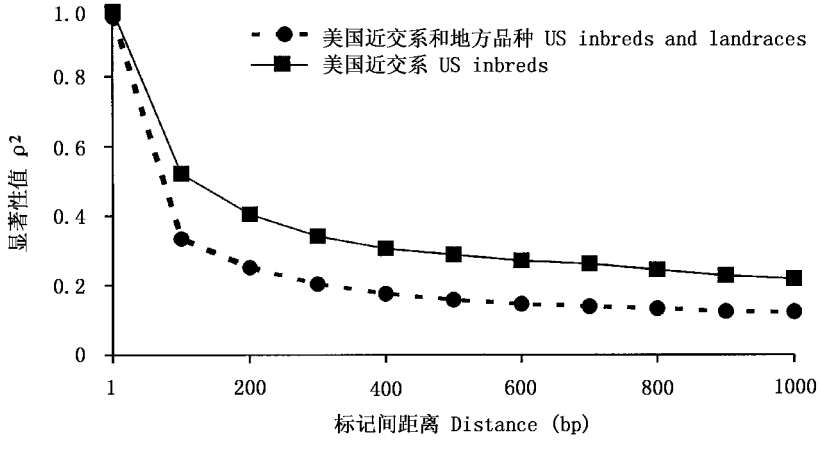

图 2　玉米 SNP 标记的连锁不平衡与标记距离之间的关系
横坐标为标记间距离,纵坐标为显著性值(根据 Tenaillon *et al.*, 2001 改绘)。

Fig. 2　The correlation between the SNP linkage disequilibrium and base pair distance in maize. Ordinate indicate significant value, and abscissa indicate distance.

基因在600 bp到3500 bp距离内LD的p值降到0.2。这些结果和人类的研究明显不同。在人类中,在基因组的某些区域LD可以延伸2.2~6.4 cM,而且人类不同基因的LD值有很大差异,一些基因的LD通常超过2.5 kb(甚至可以延伸至10 kb),但另些基因的LD则下降很快(Tenaillon *et al.*, 2001)。与此同时,Remington *et al.* (2001)利用6个基因以及47个SSR标记对玉米基因组进行了连锁不平衡研究,都到了基本相似的结果,但在他们的研究中不同的基因LD程度有较大差异。这些研究对玉米的关联作图有很重要的指导作用,因为连锁不平衡结构将影响到关联研究的设计和实施。由于玉米基因组LD下降太快,因而无法象人类那样利用SNPs来进行基因组水平的关联测定,但在QTL作图之后再对候选基因进行关联分析,对寻找具体基因甚至基因内的具体区域的数量性状效果具有很大的应用前景(Remington *et al.*,2001)。但要达到此目的,SNP检测应该保证标记密度在100~200 bp 1个SNP的水平(Tenaillon *et al.*,2001)。

模式植物拟南芥(*A. thaliana*)也是植物中开展连锁不平衡研究最早的植物,由于是自交植物,其结果与玉米明显不同。Nordborg *et al.* (2002)从拟南芥中选择了20个样本,对决定花期的基因位点*FRI*及其邻近250 kb的区域进行了测序,结果发现LD明显地随着距离加大而下降,但基本上在1 cM(250 kb)的距离消失,这一数值远远高于玉米的结果,也明显高于其他的生物类群(例如果蝇)的结果,即LD下降的速度只相当其他生物的1/50。进一步对一些地方群体的分析发现,由于奠基者效应的缘故,LD的程度会更高。因此,Nordborg等认为建立在单倍型基础上的LD作图在拟南芥以及其他自交植物中会非常有用。

群体遗传理论预期,群体的瓶颈效应和自交会明显增加LD。由于美国的栽培大豆在从亚洲引进的过程中曾经历了严重的瓶颈效应,4/5的多样性大约来自7~10个引种的个体,而且大豆是自交物种。因此,Rafalski(2002a)认为,美国的大豆群体可能会显示高水平的LD。由此可见,基因组中LD的程度和式样绝不仅仅是理论上的研究,而且它将决定选择什么样的关联作图方法。迄今对LD总体式样的大致分析表明,不管是在基因上还是在非编码区,总的情况很不一致,取决于基因组的不同区域以及不同的群体。

近年来的人类连锁不平衡研究表明,人类基因组是由许多"区块(blocks)"组成,这些区块似乎是共同遗传的并由一些长度在1~5 kb的重组热点分开。这些重组热点使得不同区块之间发生频繁的交换,但区块内的重组却很少(Stumpf, 2002)。这一结构特点很可能也存在于其他生物类群中。因此,连锁不平衡研究已经基本上从评估LD如何随遗传距离变化而变化,转化到揭示基因组中"区块"的边界和它们之间的关联强度。一旦区块确定下来,相对较少数量的标记就可以代表基因组中大部分的单倍型,而那些重组热点区域就要相应增加标记的数目(Goldstein, 2001)。可以预见,人们对基因组结构和功能的认识的不断深化,必将推动群体遗传学理论和实践的发展。

5 结束语

从某种意义上说,我们已进入SNPs时代,对人类SNPs的发现、描述及其在确定表现型中的成功标志着一个新里程碑的出现(Brooks,1999)。通过大批量、高通量SNPs的发现与鉴定,人类SNP-Haplotype遗传图谱的构建,在连锁不平衡基础上的关联分析等,有望为人体致病基因的寻找和疾病的防治提供快速和有效的途径。

近几年SNPs在生物医学和人类起源与进化研究中的成果极大地促进了SNPs在动植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在动植物研究领域中受到广泛的关注,毫无疑问将在分子和群体遗传、动植物育种和生物进化等研究领域中发挥越来越大的作用。正如Rafalski *et al.* (2002a)认为,通过SNPs标记,人们将能广泛收集到植物基因组中的信息,并进行超高分辨率分析,从而有效地利用几乎是无穷无尽的遗传信息宝库,并将开发出与表现型相关的的基因芯片进行更有效的关联分析。当前,在植物学研究中面临最大的挑战仍然来自作图群体的构建和准确判断表现型的困难,特别是对于那些容易受环境影响的性状,因而限制了在芯片上作图。因此,上述目标的实现,仍需要群体和分子遗传学、生物化学以及生物信息学等领域学者专家的携手合作。

参考文献

Ayres N. M., McClung A. M., Larkin P. D., Bligh H. F. J., Jones C. A. and Park W. D. 1997. Microsatellites and a single-nucleotide polymorphism in an extended pedigree of US rice germ plasm. *Theoretical and Applied Genetics*, **94**: 773 – 781.

Bhattramakki D. and Rafalski A. 2001. Discovery and application of single nucleotide polymorphisms markers in plant. In: Henry R. J. (ed.), *Plant Genotyping: The DNA Fingerprinting of Plant*. CAB International, 179 – 192.

Brookes A. J. 1999. The essence of SNPs. *Gene*, **234**: 177 – 186.

Buckler E. S. IV and Thornsberry J. M. 2002. Plant molecular diversity and applications to genomics. *Current Opinion in Plant Biology*, **5**: 107 – 111.

Ching A., Caldwell K. S., Jung M., Dolan M., Smith O. S., Tingey S., Morgante M. and Rafalki A. J. 2002. SNP frequency, haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genetics*, **3**: 19.

Cho R. J., Mindrinos M., Richards D. R., Sapolsky R. J., Anderson M., Drenkard E., Dewdney J., Reuber T. L., Stammers M., Federspiel N., Theologis A., Yang W. H., Hubbell E., Au M., Chung E. Y., Lashkari D., Lemieux B., Dean C., Lipshutz R. J., Ausubel F. M., Davis R. W. and Oefner P. J. 1999. Genome-wide mapping with biallelic markers in *Arabidopsis thaliana*. *Nature Genetics*, **23**: 203 – 207.

Drenkard E., Richter S. R., Rozen S., Stutius L. M., Angell N. A., Mindrinos M., Cho R. J., Oefner P. J., Davis R. W. and Ausubel F. M. 2000. A simple procedure for the analysis of single nucleotide polymorphisms facilitates map-based cloning in *Arabidopsis*. *Plant Physiology*, **124**: 1483 – 1492.

Edwards K. J. and Mogg R. 2001. Plant genotyping by analysis of single nucleotide polymorphisms. In: Henry R. J. (eds.), *Plant Genotyping: The DNA Fingerprinting of Plant*. CAB International, 1 – 13.

Etscheid M. and Riesner D. 1998. TGGE and DGGE. In: Karp A., Issac P. G. and Ingram D. S. (eds.), *Molecular Tools for Screening Biodiversity*. Chapman & Hall, London, 133 – 156.

Feng Q., Zhang Y., Hao P., Wang S., Fu G., Hung Y., Li Y., Zhu J., Liu Y., Hu X., Jia P., Zhang Y., Zhao Q., Ying K., Yu S., Tang Y., Weng Q., Zhang L., Lu Y., Mu J., Lu Y., Zhang L. S., Yu Z., Fan D., Liu X., Lu T., Li C., Wu Y., Sun T., Lei H., Li T., Hu H., Guan J., Wu M., Zhang R., Zhou B., Chen Z., Chen L., Jin Z., Wang R., Yin H., Cai Z., Ren S., Lv G., Gu W., Zhu G., Tu Y., Jia J., Zhang Y., Chen J., Kang H., Chen X., SHAhao C., Sun Y., Hu Q., Zhang X., Zhang W., Wang L., Ding C, Sheng H., Gu J., Chen S., Ni L., Zhu F., Chen W., Lan L., Lai Y., Cheng Z., Gu M., Jiang J., Li J., Hong G., Xue Y. and Han B. 2002. Sequence and analysis of rice chromosome 4. *Nature*, **420**: 316 – 320.

Ge S(**葛颂**) and Hong D-Y(**洪德元**). 1994. Genetic diversity and its detection(**遗传多样性及其检测方法**). In: Qian Y-Q (**钱迎倩**) and Ma K-P(**马克平**) (eds.), *Principles and Methodologies of Biodiversity Studies* (**生物多样性研究的原理与方法**). Chinese Science and Technology Press, Beijing, 123 – 140. (in Chinese)

Germano J. and Klein A. S. 1999. Species-specific nuclear and chloroplast single nucleotide polymorphisms to distinguish *Picea glauca*, *P. mariana* and *P. rubens*. *Theoretical and Applied Genetics*, **99**: 37 – 49.

Goldstein D. B. 2001. Islands of linkage disequilibrium. *Nature Genetics*, **29**: 109 – 111.

Griffin T. J., Hall J. G., Prudent J. R. and Smith L. M. 1999. Direct genetic analysis by matrix-assisted laser desorption/ionization mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America*, **96**: 6301 – 6306.

Gut I. G. 2001. Automation on genotyping single nucleotide polymorphisms. *Human Mutation*, **17**: 475 – 492.

Hoskins R. A., Phan A., Naeemuddin M., Mapa F. A., Ruddy D. A., Ryan J. J., Young L. M., Wells T., Kopczynski C. and Ellis C. 2001. Single nucleotide polymorphism markers for genetic mapping in *Drosophila melanogaster*. *Genome Research*, **11**: 1100 – 1113.

Howell W. M., Jobs J., Gyllensten U. and Brookes A. J. 1999. Dynamic allele specific hybridization: a new method for scoring single nucleotide polymorphisms. *Nature Biotechnology*, **17**: 87 – 88.

Kruglyak L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genetics*, **22**: 139 – 144.

Lander E. S. 1996. The new genomics: global views of biology. *Science*, **274**: 536 – 539.

Lemieux B. 2001. Plant genotyping based on analysis of single nucleotide polymorphisms using microarrays. In: Henry R. J. (ed.), *Plant Genotyping: The DNA Fingerprinting of Plant*. CAB International, 47 – 57.

Lindblad-Toh K., Winchester E., Daly M. J., Wang D. G., Hirschhorn J. N., Laviolette J. P., Ardlie K., Reich D. E., Robinson E., Sklar P., Shah N., Thomas D., Fan J. B., Gingeras T., Warrington J., Patil N., Hudson T. J. and Lander E. S. 2000. Large-scale discovery and genotyping of single-nucleotide polymorphisms in the mouse. *Nature Genetics*, **24**: 381 – 386.

Marth G. T., Korf I., Mark D. Y., Raymond T. Y., Zhijie G., Hamideh Z., Nathan O. S., LaDeana H., Pui-Yan K., Warren R. G. and Gish W. R. 1999. A general approach to single-nucleotide polymorphism discovery. *Nature Genetics*, **23**: 452 – 456.

Marshall E. 1997. Snipping away at genome patenting. *Science*, **277**: 1752 – 1753.

Mogg R., Hanley S. and Edwards K. J. 1999. Generation of maize allele specific oligonucleotides from the flanking regions of microsatellite markers. In: Schergo International linc. (ed.), *Plant and Animal Genome* VII *Conference Abstract Guide*, 941.

Mogg R., Batley J., Hanley S., Edwards D., O'Sullivan H. and Edwards K. J. 2002. Characterization of the flanking regions of *Zea mays* microsatellites reveals a large number of useful sequence polymorphisms. *Theoretical and Applied*

Genetics, **105**: 532 – 543.

Nasu S., Suzuki J., Ohta R., Hasegawa K., Yui R., Kitazawa N. and Minobe Y. 2002. Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa, Oryza rufipogon*) and establishment of SNP markers. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes*, **9**: 163 – 171.

Nordborg M., Borevitz J. O., Bergelson J., Berry C. C., Chory J., Hagenblad J., Kreitman M., Maloof J. N., Noyes T., Oefner P. J., Stahl E. A. and Weigel D. 2002. The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nature Genetics*, **30**: 190 – 193.

Oefner P. J. and Underhill P. A. 1998. DNA mutation detection using denaturing high performance liquid chromatography (DHPLC). In: *Current Protocol in Human Genetics*. Supp. 19, PP. 7. 10. 1 – 7. 10. 12, Wiley and Sons, New York.

Paris M., Lance R. and Jones M. G. K. 2001. Single nucleotide primer extensions to type SNPs in barley. In: *Proceedings of the 10th Australian Barley Technical Symposium.*, Canberra, Australia, 16 – 20 Sep. 2001.

Picoult-Newberg L., Ideker T. E., Pohl M. G., Taylor S. L., Donaldson M. A., Nickerson D. A. and Boyce-Jacino M. 1999. Mining SNPs from ETS databases. *Genome Research*, **9**: 167 – 174.

Primmer C. R., Borge T., Lindell J. and Saetre G. P. 2002. Single-nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. *Molecular Ecology*, **11**: 603 – 612.

Pritchard J. K. and Przeworski M. 2001. Linkage disequilibrium in humans, models and data. *American Journal of Human Genetics*, **69**: 1 – 14.

Rafalski J. A., Ching A., Bhattramakki D., Henderson K., Jung M., Morgante M., Dolan M., Register J., Smith O. and Tingey S. 1999. Single nucleotide polymorphisms (SNPs) in the 3′untranslated flanks of maize genes reveal conserved ancestral haplotypes. In: *Cold Spring Harbor Meeting on Genome Sequencing and Biology*, Cold Spring Harbor, New York.

Rafalski J. A. 2002a. Application of single nucleotide polymorphisms in crop genetics. *Current Opinion in Plant Biology*, **5**: 94 – 100.

Rafalski J. A. 2002b. Novel genetic mapping tools in plants: SNP and LD-based approaches. *Plant Science*, **162**: 329 – 333.

Remington D. L., Thornsberry J. M., Matsuoka Y., Wilson L. M., Whitt S. R., Doebley J., Kresovich S., Goodman M. M. and Buckler E. S. IV. 2001. Structure linkage disequilibrium and phenotypic associations in the maize genome. *Proceedings of the National Academy of Sciences, USA*, **98**: 11479 – 11484.

Sachidanandam R., Weissman D., Schmidt S. C., Kakol J. M., Marth L. D. S., Mullikin S. S. C., Mortimore B. J., Willey D. L., Hunt S. E., Cole C. G., Coggill P. C., Rice C. M., Ning Z., Rogers J., Kwok D. R. B., Mardis E. R., Yeh R. T., Schultz B., Cook L., Davenport R., Dante M., Fulton L., Hillier L., Waterston R. H., Gilman J. D. M., Schaffner S., Van Etten W. J., Reich D., Higgins J., Daly M. J., Blumenstiel B., Baldwin J., Stange-Thomann N., Zody M. C., Linton L., Lander E. S. and David Altshuler D. 2001. A map of human genome sequence variation containing 1. 42 million single nucleotide polymorphism. *Nature*, **409**: 928 – 933.

Savolainen O., Langley C. H., Lazzaro B. P. and Freville H. 2000. Contrasting patterns of nucleotide polymorphisms at alcohol dehydrogenase locus in the outcrossing *Arabidopsis lyrata* and the *Arabidopsis thaliana*. *Molecular Biology and Evolution*, **17**(4): 645 – 655.

See D., Kanazin V., Talbert H. and Blake T. 2000. Electrophoresis detection of single nucleotide polymorphisms. *Biotechniques*, **28**: 710 – 716.

Selinger D. A. and Chandler V. L. 1999. Major recent and independent changes in levels and patterns of expression have occurred at the *b* gene, a regulatory locus in maize. *Proceedings of the National Academy of Sciences, USA*, **96**: 15007 – 15012.

Stumpf M. P. H. 2002. Haplotype diversity and the block structure of linkage disequilibrium. *Trends in Genetics*, **18**: 226 – 228.

Syvanen A. C. 2001. Accessing genetic variation: genotyping single nucleotide polymorphisms: *Nature Review Genetics*, **2**: 930 – 942.

Tenaillon M. I., Sawkins M. C., Long A. D., Gaut R. L., Doebley J. F. and Gaut B. S. 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proceedings of the National Academy of Sciences, USA*, **98**: 9161 – 9166.

Tyagi S. and Kramer F. R. 1996. Molecular beacons: probes that fluorescence upon hybridization. *Nature Biotechnology*, **14**: 533 – 536.

Wang D. G., Fan J. B., Siao C. J., Berno A., Young P., Sapolsky R., Ghandour G., Perkins N., Winchester E., Spencer J., Kruglyak L., Stein L., Hsie L., Topaloglou T., Hubbell E., Robinson E., Mittmann M., Morris M. S., Shen N. P., Kilburn D., Rioux J., Nusbaum C., Rozen S., Hudson T. J., Lipshutz R., Chee M. and Lander E. S. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphism in the human genome. *Science*, **280**: 1077 – 1082.

Wang R. L., Syeca A., Hey J., Lukens L. and Doebley J. 1999. The limits of selection during maize domestication. *Nature*, **398**: 236 – 239.

Zeng Z-Y(**曾朝阳**), Xiong W(**熊炜**), Shen S-R(**沈守荣**), Zhu S-G(**朱诗国**), Li X-L(**李小玲**), Li W-F(**李伟芳**), Li J(**李江**), Zhou M(**周鸣**), Fan S-Q(**范松青**), Ma J(**马健**), Zhou J(**周洁**), Xiao B-Y(**肖炳燚**), He L(**贺林**), Li G-Y(**李桂源**). 2002. High-throughput single nucleotide polymophisms genotyping by dynamic allele-specific hybridization. *Progress in Biochemistry and Biophysics* (**生物化学与生物物理进展**), **29**(5): 806 – 810. (in Chinese)

(责任编辑：孙大川)