

专题介绍

植物的单核苷酸多态性及其在作物遗传育种中的应用^①

1,2,3 郝岗平¹ 杨清² 吴忠义¹ 曹鸣庆¹ 黄丛林^②

¹(北京农业生物技术研究中心 北京 100089) ²(南京农业大学生命科学院 南京 210095)

³(山东泰山医学院 泰安 271000)

摘要 单核苷酸多态性(single nucleotide polymorphism, SNP)是基因组中最常见的遗传多态性,在遗传学研究的许多方面具有重要的作用。综述了单核苷酸多态性的发现、特点及其应用等方面对植物SNP的研究进展,并展望其在作物遗传育种中的应用前景。

关键词 SNP, 遗传标记, 关联分析

Single Nucleotide Polymorphism (SNP) and its Applications in Crop Genetics and Breeding

1,2,3 HAO Gang-Ping¹ YANG Qing² WU Zhong-Yi¹ CAO Ming-Qing¹ HUANG Cong-Lin^②

¹(Beijing Agro-Biotechnology Research Center, Beijing 100089)

²(College of Biological Sciences, Nanjing Agricultural University, Nanjing 210095)

³(Taishan Medical College of Shandong, Tai'an 271000)

Abstract Single nucleotide polymorphisms (SNPs) are the most frequent variations in the genome of any organism, playing an important role in many aspects of genetics research. In this paper, we reviewed the feature of SNPs, its advantage in genotyping and application in crops breeding.

Key words Single nucleotide polymorphism (SNP), Genetic marker, Association analysis

随着人类、拟南芥(*Arabidopsis thaliana*)和水稻(*Oryza sativa* L.)等多种高等生物基因组测序的完成,人们已经开始致力于生物基因组序列差异的发现和利用。单核苷酸多态性(single nucleotide polymorphism, SNP)的筛选及其检测正成为研究者们广泛关注的焦点。人类基因组研究表明:基因序列变异大多是单核苷酸的变异。在不同的人群中SNP的频率分布有差异,这些差异可以代表某一人群的遗传差异。因此,研究SNP有助于解释个体的表型差异、不同群体和个体对疾病的易感性等。另外,比较物种间SNP的差异还可以了解物种间的亲缘关系和进化的生物学信息。Goldstein(2001)通过研究人群连锁不平衡(linkage disequilibrium, LD)水平发现高连锁不平衡区域远远超过我们原先的估计。这一研究表明,分析单倍型SNP比分析单

①北京市科委合同项目(H012010240240240113)和北京市科技新星项目(H020821150130)资助。

②通讯作者。Author for correspondence. E-mail: clhuang@public.bta.net.cn

收稿日期:2003-06-17 接受日期:2003-10-23 责任编辑:崔郁英

个 SNP 能够提供更有效的等位基因与表型关联的方法。植物学研究者可以借助于人类基因组 SNP 研究的方法和结果开展植物基因组 SNP 分析,玉米(*Zea mays* L.)和小麦(*Triticum aestivum* L.)等作物的高度遗传多样性更有利于 SNP 的鉴别。可以利用作物的近等基因系直接分析单倍型 SNP,所以 SNP 在植物上的应用前景十分广泛,对植物 SNP 的开发同样具有时间的紧迫性。本文在对 SNP 作简要介绍的基础上,分析了 SNP 在作物遗传育种上的应用前景。

1 单核苷酸多态性

SNP 和碱基的插入/缺失(indels)是生物个体之间两种序列差异类型。SNP 是指单个核苷酸变异引起的 DNA 序列多态性,包括单个碱基的转换(transition),如 $T \rightarrow C$ 和 $A \rightarrow G$, 以及颠换(transversion)如 $A \rightarrow C$ 、 $T \rightarrow G$ 、 $G \rightarrow C$ 和 $A \rightarrow T$, 而且其中最少一种等位基因在群体中的频率不小于 1%。因为这种变异可以是转换也可以是颠换,理论上讲,SNP 既可能具有 2 等位多态性,也可能具有 3 或 4 等位多态性。但 3 或 4 等位多态性的情况较少见,通常所说的 SNP 都是 2 等位多态性。转换的发生率总是明显高于其他几种变异,属于转换型变异的 SNP 约占全部 SNP 的 2/3。在单个基因或整个基因组中 SNP 的分布不均匀,在非转录序列中要多于转录序列,而且在转录区也是非同义突变的频率比同义突变的频率低得多。在基因编码区的 SNP 称为编码 SNP (coding SNP, cSNP), 它又分为两类:未引起蛋白质编码氨基酸序列改变的同义编码 cSNP (synonymous cSNP, s-cSNP) 和引起蛋白质编码氨基酸序列改变的非同义编码 cSNP (nonsynonymous cSNP, ns-cSNP)。其中 ns-cSNP 会导致蛋白质功能的改变。由于 cSNP 在标记功能基因和研究基因的遗传效应等方面具有重要意义,因此它的研究备受关注。

SNP 是继限制性片段长度多态性(RFLP)和微卫星多态性(SSR)之后发展起来的第三代分子标记技术。与前两代分子标记技术相比,它具有较多优点。

1.1 数量多,分布广泛

SNP 是目前为止分布最为广泛、存在数量最多且标记密度最高的一种遗传多态性标记。The Arabidopsis Genome Initiative (2000) 将拟南芥 Columbia 生态型(ecotype)82 Mb 和 Landsberg 生态型 92.1 Mb 的基因组序列进行了比较分析,发现了 25 274 个 SNP,平均每 3.3 kb 有一个 SNP;发现了 14 570 个 Indel,平均每 6.1 kb 有一个 Indel。我们与法国 INRA 合作,以 94 个拟南芥生态型(ecotype)为材料,系统研究了位于第 IV 条染色体上 CBF1、CBF2 和 CBF3 位点区域的 3 907 个碱基对,发现 SNP 的分布频率更高:在基因编码区域平均每 33 bp 有一个 SNP,每 2 kb 有一个 Indel;在 5' 端非编码区域平均每 29 bp 有一个 SNP,平均每 60 bp 有一个 Indel;在 3' 端非编码区域平均每 47 bp 有一个 SNP,平均每 142 bp 有一个 Indel(据未发表资料)。

1.2 遗传稳定性高,遗传分析重现性好且准确性高

虽然目前 SSR 在基因定位研究中的应用较多,但由于在基因组中 SSR 存在不稳定性,分布的密度相对较低,基因分型较难实现完全自动化等问题,使其在基因定位中的应用受到局限(Weber and Wong, 1993)。SNP 标记的遗传稳定性要比 SSR 等标记高得多,而且在群体中也是按孟德尔规律遗传用于遗传分析或基因诊断,重现性和准确性大大提高。

1.3 易于快速且高通量地进行基因型分型

由于 SNP 的二态性,非此即彼,在基因组中往往只需 + / - 的分析,而无须象检测 SSR 标记那样分析片段的长度,这就有利于自动化的筛选或检测技术的开发。尽管 SNP 只有两种等

位基因型,在个体中的多态信息量比SSR等多等位基因型的信息量少,但SNP二态性、高频率和稳定性的特性弥补了信息量上的不足。3~4个相邻的SNP双等位标记构成的单倍型就有8~16种,相当于1个SSR形成的多态性,且突变率很低(每个核苷酸的突变率约为 10^{-9})(吴刚和李果,2001)。由于SNP自身的特性,注定了它更适于复杂性状的遗传分析和引起群体差异的基因识别等方面的研究。

2 植物SNP的发现及基因型分型

对植物SNP的研究主要分两个层次:对目的性状相关SNP的发现和群体中筛查已知的目的性状相关SNP;对样本中已知位点的SNP进行基因型分型(genotyping),结合群体的表型分析,获得目标性状的基因标记。

2.1 植物目的性状相关SNP的发现

目前植物SNP的研究多处于SNP的发现阶段。植物SNP的发现必须考虑如下两个方面的工作:目的性状相关基因位点或EST的选择以及核心群体的确定。对于小麦、玉米和水稻等农作物来说,最好从有代表性的近等基因系选择个体。通常可以参考RFLP标记的资料来确定核心群体的构成。

2.1.1 不同个体的PCR扩增片段直接测序是发现SNP的最常用方法 利用目的性状相关的基因位点或EST序列设计引物,通过PCR扩增,对PCR产物进行测序,然后应用SNP发现的专业软件Genalys或DNASTAR,结合Clustal等软件,分析测序结果,排除测序错误,发现SNP。

Nasu等(2002)以水稻3个栽培种和1个野生种(*Oryza rufipogon*)为实验材料,对分布于全基因组的417个位点进行SNP研究,发现了2800个SNP,发生频率平均每89bp有一个SNP。

Tenaillon等(2001)以25个玉米近等系为材料,对分布在1号染色体上的21个遗传位点序列多样性进行了研究,发现平均每104bp有一个SNP。Bhatramakki等(2002)以8个玉米近等系为材料,对502个EST位点的Indel多态性进行了研究,发现43%的EST具有Indels多态性,共发现了655个Indel,其中单碱基的Indel占54.8%,其次是2个和3个碱基的Indel,也发现了6个碱基(3.4%)和8个碱基(2.3%)的Indel。Ching等(2002)构建了36个玉米近等系组成的核心群体,研究18个基因位点的遗传多态性,发现在基因非编码区域平均每48bp有一个SNP,每85bp有一个Indel;在基因编码区域平均每130bp有一个SNP,而Indel很少,在2.35kb编码区域只发现了一个3bp的Indel。

Cregan等(http://www.intl-pag.org/pag/9/abstracts/P3e_11.html)研究了22个不同大豆(*Glycine man* (L.) Merr.)基因型的SNP,发现在基因编码区平均每610bp有一个SNP,在非编码区平均每229bp有一个SNP。

Kota(2001)等从19000个EST中选择了目的性状相关的EST180个,用7个基因型的大麦(*Hordeum vulgare* L.)进行了SNP研究,发现了72个SNP。

2.1.2 SNP发现的其他方法 基于公共数据库的直接方法也常用来搜索新的SNP。公共数据库中已有大量的表达序列标签(ESTs)、序列标签位点(STSs)、cDNA文库和基因组测序公开的序列等信息。在这些序列之间必然存在大量的重叠区域,通过比较这些重叠区域,并运用一些软件(如XGAP)删除由测序造成的碱基错误,就可得到候选SNP甚至真正的SNP,这种策略可大大降低成本,已被用于构建SNP标记(Gu et al., 1998; Picoult-Newberg et al., 1999; Marth et

al., 1999)。利用拟南芥数据库发现了 37 344 个 SNP, 通过 <http://www.arabidopsis.org/cereon/index.html> 可以查到这些数据。

另外, 基因芯片技术也被用于 SNP 的发现。Wang 等(1998)报道采用 DNA 芯片技术从 16 725 个 STS (包含 2 Mb 的人类 DNA)中得到 2 748 个 SNP, 平均每 721 bp 有一个 SNP。

2.2 植物 SNP 的基因型分型

植物 SNP 的基因型分型研究包括三步: 首先是分型群体目的性状表型的精确鉴定; 其次是已知的植物目的性状相关 SNP 的基因型分型; 最后通过基因型和表型的关联分析, 获得植物目的性状 SNP。

表型鉴定的精确性对于 SNP 的应用至关重要, 尤其是多基因控制的数量性状。

总的来说, 用来发现新的 SNP 的方法都可以用来对已知的目的性状相关 SNP 进行基因型分型。对 SNP 进行基因型分型的主要技术包括微测序、多重反向点杂交、DNA 芯片或微阵列以及基质辅助激光解吸电离飞行时间(MALDI-TOF)质谱法等。目前用得较多的是 DNA 微阵列技术和 MALDI-TOF 技术。

3 SNP在作物遗传育种上的应用

目前已经开发出许多基于非直接检测序列多态性的分子标记, 其中 SSR 标记优于其他标记。SSR 标记不适合于关联分析, 原因在于进化上不同的物种可以出现片段大小相等的 SSR 等位标记, 相反也有可能在相同的单倍型里发现片段大小不同的 SSR 等位标记。SNP 分析不需要 DNA 大小的分离, 可以使用自动化的微阵列技术批量检测, 同时也比 SSR 更易于在基因组的单拷贝区定位。另一方面, SNP 属于双等位标记, 它的期望杂合度比较低, 如 SNP 分析玉米种质的杂合度为 0.263, 而 SSR 标记为 0.77(Taramino and Tingey, 1996)。当在筛查区域出现几个相距较近的 SNP 决定单倍型时, SNP 可以提供非常有用的信息。与连锁不平衡结合, 决定单倍型的 SNP 是十分丰富的, 一个亚区域的 SNP 就足以决定单倍型, 这种决定单倍型的 SNP 就叫做单倍型标签(haplotype-tag)(Johnson *et al.*, 2001)。

3.1 利用 SNP 标记构建高密度遗传连锁图谱

SNP 在基因组中分布广泛, 发生频率很高, 如水稻和玉米 SNP 的发生频率分别为平均每 89 bp 有一个 SNP(Nasu *et al.*, 2002), 平均每 61 bp 有一个 SNP(Ching *et al.*, 2002)。SNP 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, 将对作物遗传作图及其精细程度产生深远的影响。随着新的 SNP 标记的发现和定位, 作物遗传作图的标记密度将日益增高, 这将为作物育种提供前所未有的便利工具。随着标记密度的升高, 基因组扫描能够将数量性状位点(QTL)定位于更小的染色体区域内, 从而为新的主效基因的发现和定位克隆打下良好的基础; 而且高密度 SNP 遗传图谱的建成使我们更精确的进行标记辅助选择(MAS), 降低或消除目的基因之外的遗传背景对这些技术带来的不良影响。除此之外, 高密度分子标记的定位也会给品种资源和品种纯度的鉴定带来崭新的信息。

3.2 遗传图谱和物理图谱的整合

在整合物理图谱和传统的遗传图谱时, 需要筛选 BAC 末端序列来找到没有重复序列的区域, 然后在此区域发现作图亲本之间的 SNP 多态性, 这些 SNP 标记再用于传统遗传作图。玉

米大约 20% 的 BAC 末端序列为单或低拷贝序列区域, 适合于这两种图谱的整合(Meyers *et al.*, 2001)。

3.3 群体遗传学和连锁不平衡

群体遗传学是研究群体的遗传组成和其相互关系, 所使用的主要研究工具是 DNA 多态性。随着 SNP 时代的到来, 群体遗传学研究者将能利用适当的 SNP 标记来进行更详细和快速的作物基因组研究, 有效的分析基因型和表型之间的关系。

要评估 SNP 在群体遗传学中的作用, 首先要熟悉连锁不平衡(linkage disequilibrium, LD)。对两个连锁基因座的等位基因 A 与 B 之间的关联以 Δ 值来衡量。 Δ 值 = AB 基因型频率 - A 基因型频率 $\times$ B 基因型频率, 并且 Δ 值以 $1-r$ 的速率逐代递减, 直到 0。 r 为两点之间的重组值。因而, 经 n 代的随机交配后, $\Delta_n = \Delta_0(1-r)^n$ 。若 $\Delta = 0$, 则 A 和 B 随机关联, 也即两点连锁平衡; 若 $\Delta = 1$, 则 A 和 B 完全关联; 若 $0 < \Delta < 1$, 表示非随机关联, 即连锁不平衡(吴国俊, 1997)。

LD 是一个复杂现象, 遗传距离、不同等位基因的选择压力、遗传飘变、群体的瓶颈效应以及发生新的突变都影响 LD。但由于飘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基因的精细定位。连锁不平衡分析常用于定位克隆的后期阶段, 也就是基因在染色体上的定位经连锁分析大致确定后, 这时位标与性状相关基因相距很近。由于缺乏适当的和高密度的能够使用有效的方法进行染色体区段比较的多态性标记, 所以在作物上利用 LD 的分析受到一定的限制。现在这种状况可望依赖于 SNP 的使用来改善。这种非常稳定和丰富的标记, 将促进这个领域在作物育种上的快速发展。

3.4 基于 SNP 的关联分析

目前作物多基因控制的抗病及抗逆等性状基因的研究并没有取得象单基因控制性状那样快的进展, 但这方面的研究因为候选基因关联分析的提出而出现新的曙光。关联分析不需要大的研究群体, 而是比较某个或某一套标记在抗性和非抗性个体的分布程度, 某种标记如果在抗性个体中分布十分明显, 那么就可以认为该标记与抗性表型相关联, 这种关联可能是目标位点与邻近的多态标记间存在连锁不平衡的结果。

进行关联分析时, 需要检测候选基因的等位基因与表型的关联性 or 全基因组扫描来确定与表型相关的区域。基于这种连锁不平衡的相关分析需要一套高密度的多态标记进行系统的基因组扫描。双等位的 SNP 由于其分布广、突变率低且易于自动化检测而成为首选。基因组扫描所需最少的 SNP 标记数目取决于连锁不平衡的染色体区域长度。在植物群体中, LD 的区域长度取决于群体的选择, 群体的瓶颈效应和近亲繁殖会增加 LD 区域长度。如果群体在演化过程中经历了瓶颈效应后会表现为长的 LD 区域和高重组率。美国的栽培大豆在由亚洲到美洲的迁移过程中就经历了好几次瓶颈效应(Delanney *et al.*, 1983), 另外大豆又是近亲繁殖种类, 因此美国大豆很可能具有很长的 LD 区域。尽管没有美国大豆 LD 区域的确切数字, 但是在 22 个大豆基因型里发现的 SNP 也在 3 个作图亲本 Minsoy、Noir 和 Archer 中发现, 因此有可能发现一些大区域的单倍型(Rafalski, 2002)。据最新报道, 北美玉米群体的最大 LD 区域可能大于 100 kb (Rafalski, 2002)。通过 RFLP 分析发现一些甘蔗(*Saccharum sinensis* Roxb.) 杂交后代 LD 区域达到几个里摩(Jannoo *et al.*, 1999)。染色体着丝粒附近的低重组率区比编码区具有更长的 LD 区

域,推测多数重组发生于基因附近。要解释基因密度、重组和 LD 之间的关系还有待于进行更深入系统的研究。

3.5 用 SNP 标记作为连锁图谱进行连锁分析

继第一代 RFLP 和第二代 SSR 标记连锁图谱后,SNP 标记策略的提出又为遗传图谱的研究提供了新的契机。乍一看起来,这一策略似乎又回到了低多态性的 RFLP 标记上来。然而,目前各种新的技术和手段的应用,可以使人们通过快速高通量地检测大量的 SNP 来弥补其低多态性的不足。

SNP 图谱的构建与应用应当说是遗传学研究史上的第三次飞跃。SNP 技术的应用使连锁分析技术逐步趋于自动化,这是 SNP 技术上对分子遗传学研究的最大贡献。这种变化与各种新方法的建立是分不开的,尤其是检测手段的变化。目前应用的最多的是 DNA 芯片与微阵列技术和 MALDI-TOF 技术。它们用于 SNP 的直接分型,使得分型工作可以快速且高通量地对大样本进行处理,从而大大提高了分型速度,降低了成本。然后再依照参数分析和 LOD 值的方法来衡量 SNP 标记与目的性状基因间的连锁关系(刘万清和贺林,1999)。

3.6 5SNP 应用于进化和种群多样性的研究

生物界的进化与进化过程中物种多样性的形成、基因组的突变以及突变的选择密切相关,构建整个基因组的 SNP 图谱对于直接研究物种起源和进化具有重大的意义。

Kota 等(2001)对大麦 7 个基因型的 180 个 EST 位点进行了 SNP 研究,发现了 72 个 SNP,并且将这些 SNP 标记应用于大麦亲缘关系的研究。Kanazin 等(2002)进一步对大麦 5 个基因型的 54 个 EST 位点进行了 SNP 研究,发现了 112 个 SNP,并且将这些 SNP 标记应用于大麦进化的研究,发现大麦和其野生祖先 *H. spontaneum* 之间存在多重驯化行为或基因的多重转移。

4 结束语

SNP 是一种用来构建高密度遗传连锁图谱和基于全基因组或候选基因的关联分析的重要分子标记,它作为一种新的遗传学研究工具不但可以大大加快植物基因组的研究,而且对于整个生物学研究和生物产业也将带来巨大的变化。植物学研究者借助于人类基因组在这方面的研究方法和成果,建成作物高密度的 SNP 遗传连锁图谱已为时不远。但是,表型鉴定的精确性和群体的选择是 SNP 研究的限制因素,尤其是数量性状的研究。因此,适合于 SNP 关联分析高效技术的建立和核心群体的构建对于植物 SNP 研究十分重要。总之,植物的 SNP 研究虽然刚刚开始,但已经显示出其诱人的应用前景,它的开发将和人类 SNP 的开发一样,同样为植物基因组研究绘制出更加精细且更加绚丽多彩的蓝图。

参 考 文 献

- 刘万清,贺林,1999. 利用 SNP 进行遗传病致病基因搜索的策略. 生命科学, 11(5): 197~200
- 吴刚,李果,2001. 单核苷酸多态性及其在糖尿病易感基因筛选中的应用. 国外医学内分泌学分册, 21(6):288~290
- 吴国俊,1997. 多基因疾病定位的策略和研究进展. 国外医学遗传学分册, 20(4): 169~172
- Bhattarakki D, Dolan M, Hanafey M, Wineland R, Vaske D, Register J C, Tingey S V, Rafalski A, 2002. Insertion-deletion polymorphisms in 3'- regions of maize genes occur frequently and can be used as highly informative genetic markers. *Plant Mol Biol*, 48:539~547
- Ching A, Caldwell K S, Jung M, Dolan M, Smith O S, Tingey S, Morgante M, Rafalski A J, 2002. SNP frequency,

- haplotype structure and linkage disequilibrium in elite maize inbred lines. *BMC Genet*, **3**: 19
- Delanney X, Rodgers D M, Palmer R G, 1983. Relative genetic contributions among ancestral lines to North American soybean cultivars. *Crop Sci*, **23**: 944~949
- Goldstein D B, 2001. Islands of linkage disequilibrium. *Nat Genet*, **29**: 109~111
- Gu Z, Hillier L, Kwok P Y, 1998. Single nucleotide polymorphism hunting in cyberspace. *Hum Mutat*, **12**: 221~225
- Johnson G C, Esposito L, Barratt B J, Smith A N, Heward J, Di Genova G, Ueda H, Cordell H J, Eaves I A, Dudbridge F, Twells R C, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough S C, Clayton D G, Todd J A, 2001. Haplotype tagging for the identification of common disease genes. *Nat Genet*, **29**: 233~237
- Jannoo N, Grivet L, Dookun A, D 'Hont A, Glaszmann J C, 1999. Linkage disequilibrium among modern sugarcane cultivars. *Theor Appl Genet*, **99**: 1053~1060
- Kota R, Varshney R K, Thiel T, Dehmer K J, Graner A, 2001. Generation and comparison of EST-derived SSRs and SNPs in barley (*Hordeum vulgare* L.). *Hereditas*, **135**: 145~151
- Kanazin V, Talbert H, See D, DeCamp P, Nevo E, Blake T, 2002. Discovery and assay of single-nucleotide polymorphisms in barley (*Hordeum vulgare*) . *Plant Mol Biol*, **48**: 529~537
- Marth G T, Korf I, Yandell M D, Yeh R T, Gu Z, Zakeri H, Stitzel N O, Hillier L, Kwok P Y, Gish W R, 1999. A general approach to single-nucleotide polymorphism discovery. *Nat Genet*, **23**: 452~456
- Meyers B C, Tingey S V, Morgante M, 2001. Abundance, distribution, and transcriptional activity of repetitive elements in the maize genome. *Genome Res*, **11**: 1660~1676
- Nasu S, Suzuki J, Ohta R, Hasegawa K, Yui R, Kitazawa N, Monna L, Minobe Y, 2002. Search for and analysis of single nucleotide polymorphisms (SNPs) in rice (*Oryza sativa*, *Oryza rufipogon*) and establishment of SNP markers. *DNA Res*, **9**: 163~171
- Picoult-Newberg L, Ideker T E, Pohl M G, Taylor S L, Donaldson M A, Nickerson D A, Boyce-Jacino M, 1999. Mining SNPs from EST databases. *Genome Res*, **9**: 167~174
- Rafalski A, 2002. Application of single nucleotide polymorphisms in crop genetics. *Curr Opin Plant Biol*, **5**: 94~100
- Taramino G, Tingey S, 1996. Simple sequence repeats for germplasm analysis and mapping in maize. *Genome*, **39**: 277~287
- Tenaillon M I, Sawkins M C, Long A D, Doebley J F, Gaut B S, 2001. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc Natl Acad Sci USA*, **98**: 9161~9166
- The Arabidopsis Genome Initiative, 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature*, **408**: 796~815
- Weber J L, Wong C, 1993. Mutation of human short tandem repeats. *Hum Mol Genet*, **2**: 1123~1128
- 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, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson T J, Lander E S, 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science*, **280**: 1077~1082