

## 利用连锁不平衡法发掘作物种质资源中 优异基因的研究进展

张鹏, 李金泉

(华南农业大学农学院, 广州 510642)

**摘要:**作物种质资源中保留着大量的优异基因。由于缺乏合适的统计方法,作物种质资源中蕴藏的丰富遗传变异还未能充分发掘和利用。连锁不平衡分析是对基因(或已定位的高分辨率QTL)的功能进行鉴定的有效方法。DNA测序技术和高通量SNP分析等技术的发展,使利用连锁不平衡法进行优异基因的发掘成为了可能。综述了有关利用连锁不平衡发掘作物种质资源中优异基因的研究进展。

**关键词:**连锁不平衡;种质资源群体;优异基因

**中图分类号:**S326 **文献标识码:**A

### Reviews of the Advancement on Mining Exotic Genes in Crop Germplasm Resources by Linkage Disequilibrium Method

Zhang Peng, Li Jinquan

(College of Agriculture, South China Agricultural University, Guangzhou 510642)

**Abstract:** A large number of exotic genes exist in crop germplasm resources. The rich genetic variations in crop germplasm resources haven't been fully explored and utilized because of being lack of appropriate statistical methods. Linkage disequilibrium analysis is an effective way to identify the function of the gene (or targeted high-resolution QTL). Development of DNA sequencing technology and high-throughput SNP analysis technology make it possible that discovers the exotic genes by using linkage disequilibrium method. The authors summarize the research of mining excellent exotic genes with linkage disequilibrium method in crop germplasm resources.

**Key words:** linkage disequilibrium, germplasm resources, exotic genes

早在20世纪初期, Jinnings就提出了连锁不平衡(Linkage disequilibrium, LD)这一概念。连锁不平衡指的是不同遗传标记间存在着的非随机组合现象<sup>[1]</sup>。座位间的遗传是使连锁不平衡在群体中得以长期保持的主要机制。因此通过测定标记位点与潜在的QTL之间的连锁不平衡程度即可断定QTL所在的位置。通过检测遍布基因组中的大量遗传标记位点或者候选基因附近的遗传标记,寻找与QTL紧密连锁而表现出与QTL相关的位点,这就是连锁不平衡定位基因的基本思想<sup>[2]</sup>。

目前中国的三大作物中,水稻种质资源已达67 841份,小麦种质资源达38 200份,玉米种质资源达16 901份,其它重要作物的种质资源数量也均在数千份以上<sup>[3]</sup>。以往对种质资源群体中的基因发掘一般采用常规的QTL定位,必须首先创建由一个或几个精心筛选的亲本杂交而来的分离群体(如F<sub>2</sub>、DH、BC群体),所定位的QTL很大程度上依赖于所选择的有限亲本。由于缺乏合适的统计方法,种质资源群体中蕴藏的丰富遗传变异还未能充分开发利用。如果按照常规的QTL定位方法来利用这些遗传变异,就需要设计包

**基金项目:**国家自然科学基金资助项目“利用连锁不平衡法发掘水稻种质资源群体的优异基因研究”(30700494)。

**第一作者简介:**张鹏,男,1983年出生,在读硕士研究生,研究方向:植物种质创新与生物技术,通信地址:510642广东省广州市天河区五山路483号华南农业大学农学院植物育种系。Tel: 020-85280205, E-mail: zyzhang\_311@stu.scau.edu.cn。

**通讯作者:**李金泉,男,1974年出生,博士,副教授, E-mail: lijinquan@scau.edu.cn。

**收稿日期:**2009-01-08, **修回日期:**2009-01-21。

括所用亲本的双列杂交,在时间、资金、空间和分析方法上都非常困难<sup>[4]</sup>。如何有效地鉴定种质资源群体中蕴藏的优异基因值得深入研究。

连锁不平衡分析是对基因(或已定位的高分辨率QTL)的功能进行鉴定的有效方法,也是评价自然植物群体性状表型变异的候选基因贡献率的有效方法。DNA测序技术和高通量SNP分析技术的发展及2001年LD作图(LD mapping)在植物中的首次成功运用,使利用其进行优异基因的发掘成为了可能。笔者对有关利用连锁不平衡发掘作物种质资源中优异基因的研究进展进行综述。

### 1 连锁不平衡方法的建立及相关理论研究

从连锁不平衡方法开始建立至今已经取得了一系列的进展。在玉米自交系、拟南芥和野生大麦中的研究表明LD水平衰减很快,显示LD分析可以得到足够的分辨率来确定候选基因区域<sup>[5-6]</sup>。Gupta等<sup>[7]</sup>认为LD可使用于植物基因组学研究,主要集中于两方面:第一方面用于性状与标记的关联分析(不需要使用作图群体),随后可用于分子标记辅助选择(MAS);第二方面应用于研究自然群体和种质资源库的遗传多样性、群体结构和作物的遗传改良。Norbrog等<sup>[8]</sup>检测并建立了拟南芥全基因组的连锁不平衡结构图,其研究表明利用连锁不平衡法鉴定拟南芥的基因功能是一个有效的策略。Zhang等<sup>[4]</sup>利用404份商用玉米自交系的性状值、分子标记信息和系谱信息,定位了8个控制玉米雄花开花性状的QTL,并通过BLUP方法评价了所有自交系的QTL效应。Parisseaux等<sup>[9]</sup>发现这种方法检测的QTL能在不同的群体中高度重复。近年来,随着生物信息学的飞速发展和水稻基因组测序工作的完成,使得SSR(simple sequence repeats)标记的引物设计非常容易和快捷,水稻基因组的SSR标记已发展到2700多个<sup>[10]</sup>。基于PCR技术的SSR标记表现出共显性遗传、数量多、多态性高和信息量大等优点。丰富的SSR和SNP标记将为深入研究作物种质资源群体的连锁不平衡结构、基因的精细定位和关联分析等提供重要的技术支持。

到目前为止,国内外基于连锁不平衡的关联分析法在水稻种质资源中的研究刚刚起步。Agrama等<sup>[11]</sup>用120个SSR标记分析了172份水稻种质资源中的LD水平,有显著证据表明LD延续至15 cM范围,相邻位点的LD强度足可以通过全基因组扫描有效应用于重要农艺性状的关联分析。Thomson等<sup>[12]</sup>为水稻的连锁不平衡定位准备了大量的SSR和SNP数据,并认为连锁不平衡定位和关联分析的最新进展为高效率地探索

种质资源库中有用的遗传变异提供了新的发掘途径<sup>[12]</sup>。Mather等<sup>[13]</sup>发现栽培稻的连锁不平衡程度大于普通野生稻,最大的连锁不平衡数量存在于粳稻亚种中,认为SNP分析和LD研究将为鉴定水稻重要表型性状的基因奠定基础。

美国康奈尔大学已经设计出了进行分子标记与表型关联分析的软件——TASSEL<sup>[14]</sup>,运用此软件可以直接进行连锁不平衡分析和基因定位。

## 2 连锁不平衡分析在发掘优异基因方面的应用

### 2.1 通过LD作图发掘优异基因

LD作图亦称为关联作图(association mapping),是一种基于LD的将遗传变异与目标性状表型联系起来的技术,具有更广泛的遗传变异和分辨率较高的两大优势<sup>[5]</sup>。LD作图利用的是自然群体中的自然变异,即通过分析自然群体中标记与紧密连锁QTL间的LD关系来鉴定和定位QTL,而且可以鉴定由QTL所代表的真正与被研究目的性状相关联的基因。

LD作图包括基于全基因组扫描和基于候选基因两种策略。在玉米<sup>[8]</sup>和大麦<sup>[15]</sup>中也提出了一种被称之为“two-tiered”的颇具前景的LD作图策略,即首先利用在较长范围内存在LD的骨干种质材料进行低分辨率的基于全基因组扫描的LD作图以确定候选基因区域,然后再利用地方品种和野生种质进行高分辨率的LD精细作图以发掘候选基因。

LD作图已在许多植物简单或复杂数量性状的遗传剖析分析中进行了成功运用<sup>[7]</sup>,并且涉及许多性状。如在玉米中,LD作图已对基因多样性与开花时间<sup>[16]</sup>、胚乳颜色<sup>[17]</sup>、淀粉合成<sup>[18]</sup>、玉米的球蛋白和绿原酸含量<sup>[19]</sup>、饲用品质<sup>[20]</sup>等性状变异之间的关系进行了研究。Thornsberry等<sup>[16]</sup>利用92个玉米自交系的开花期数据和Dwarf8序列多态数据进行的关联分析取得了与早期突变体分析和QTL作图相一致的结果。Virk等<sup>[21]</sup>进行的关联分析为水稻的6个农艺性状(分蘖数、株高、开花期、穗长、叶长和粒宽)找到了与之关联的RAPD标记,发现一个可以解释其分蘖数变异49.84%的QTL。LD作图是一种比较有效的作图方法,可以实现数量性状的精确作图,但还有一些问题有待进一步解决和完善,譬如如何排除群体结构的影响。LD作图也存在作图的“盲区”,对遗传多样性偏低群体的作图效果还不如QTL作图<sup>[22]</sup>。

Yu等<sup>[23]</sup>对运用连锁不平衡原理进行的关联作图和已经被广泛应用的其它作图方法在分辨率、检测的等位变异基因数目和研究所需时间上进行了较为详细的比较(图1)。

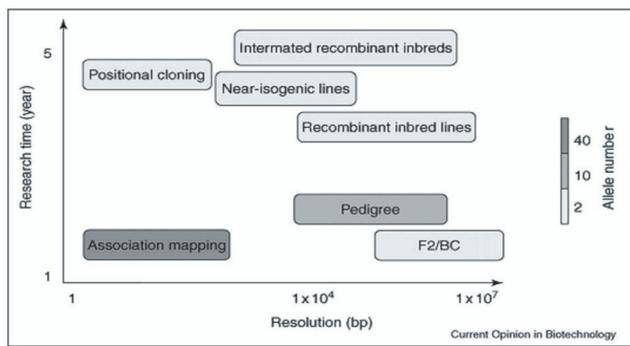


图1 关联分析与其它作图方法在分辨率、检测的等位变异基因数目和研究所需时间上的比较

## 2.2 利用连锁不平衡分析进行QTL精细定位发掘基因

利用连锁不平衡方法和现有作物群体进行QTL定位有以下的优点：(1)品种的基因型在不同世代中是一个常量；(2)每一个品种的表型值能够在不同的环境中重复地试验，可以减少环境和测量误差；(3)应用世代累积的重组能够使QTL定位到一个精细的范围；(4)不需要试验杂交种及其分离后代的作图群体；(5)QTL定位之后，每个品种的QTL效应可通过线性无偏预测(the best linear unbiased prediction, BLUP)方法估计，使育种家能方便地选择到优异的品系和品系组合<sup>[4]</sup>。

基于连锁不平衡分析实现人类致病突变基因高分辨率定位的理论与方法已经广泛应用于某些致病突变基因的精确精细定位和克隆<sup>[24-26]</sup>。Meuwissen等<sup>[27]</sup>用连锁分析与连锁不平衡分析相结合的方法将影响牛双胎率的QTL定位到5号染色体上CSSM22~ILSTS66两个标记之间，这两个标记与该QTL紧密连锁，在试验中没有观察到重组，结果与USMARC基因组数据库所提供的两个标记相距该QTL为0.6 cM是一致的，可以肯定地得出结论：QTL被定位在小于1 cM的区域，这种情况下通过图位克隆来寻找主效基因是可行的。

## 2.3 单倍型标签SNP与连锁不平衡结合发掘基因

SNP(single nucleotide polymorphism, 单核苷酸多态性)是指在某种生物不同个体DNA序列中，存在单个核苷酸变异的多态现象。SNP的最大优点是它们在基因组中丰富的多态性。利用SNPs分子标记，通过LD作图法可以鉴定导致生物特定性状的基因。

Olsen等<sup>[28]</sup>运用基于单倍型(而非基于单个SNP位点)的方法阐明了拟南芥中开花基因CRY2的自然等位基因变异，研究发现开花相关基因CRY2在31个生态型拟南芥中有A和B两种明显不同的单倍型，3个多态性位点HAP AQ、HAP AS和HAP B作为单倍型标签SNP基本上可以将这些材料区分开来。进一步研究表明，短日照条件下较常见的HAP AQ单倍型而言，HAP AS和HAP

B单倍型与提早开花这一性状呈显著关联。这是首例在拟南芥中运用基于单倍型SNP的LD作图对QTL进行精细作图的成功运用，从而可以发掘基因。

## 3 连锁不平衡的研究热点与展望

### 3.1 重组与连锁不平衡

目前，关于基因组结构和重组与基因组内LD水平关系较为一致观点是：基因组水平内不同区域的重组率不同，整个基因组内存在重组热点，通常具有较高LD水平的单倍型区域和LD水平较低的重组热点呈散布状态；基因富集区(染色体末端)比基因较少区域(着丝粒处)具有较高的重组率和较低的LD<sup>[29]</sup>。基因内部比基因间具有较高的重组率和较低的LD，如玉米bz基因内部的重重组率比整个基因组的平均重组水平要高出100倍<sup>[30]</sup>。

重组可以导致本来连锁的座位彼此独立存在。重组是在减数分裂过程中通过同源交换和基因转换发生的。交换(crossing-over)使原来连锁的基因不再联合传递，打破较长范围内的LD。Drouaud等<sup>[31]</sup>最近发现拟南芥第4染色体上交换率较高的几个区即为重组的热点区域。而基因转变(gene conversion)则使减数分裂过程中基因的一个等位基因被转换成同座位的另一个等位基因，它可以打破基因组局部范围内的LD，但不会影响群体内较长范围的LD。因此，仅是紧密连锁的标记其LD水平降低，而其侧翼标记仍处于完全关联状态。但最近的研究表明，基因转变是影响某些座位重组和LD的重要因素。

### 3.2 上位性互作和基因型与环境互作与连锁不平衡

作物的许多重要农艺性状，如产量及其相关性状、株型、生育期和抗性等均属于由多基因控制的复杂数量性状。这些性状由多基因控制的特性、不同座位之间的相互作用以及环境因素的影响更是增加了对其研究的难度。近年来，新的统计方法的建立<sup>[32]</sup>使存在上位性互作的基因座的发掘成为了可能，而上位性互作导致的相关座位间LD的存在为更好地研究上位性效应及基因型与环境的互作提供了一条新的思路。Lou等<sup>[33]</sup>提出了一种新的在自然群体中鉴定加性、显性和上位性效应基因座的统计方法，即基于单倍型算法的多座位LD分析。该方法强大精确的统计能力已在人类身高研究中得到成功运用，但到目前为止还没有在植物中应用的报道。

### 3.3 功能性标记的开发与连锁不平衡

功能性标记的开发为解决基于连锁的分子标记辅助选择和基于特定分离群体的定位结果进行选择所遇到的问题提出了新的思路。在多个材料中对目标性状进行调查，对目标基因进行序列分析，结合性状和基因序列信息进行基于连锁不平衡的关联分析，有可能直

接得到控制目标性状的基因。

### 参考文献

- [1] Rafalski A. Applications of single nucleotide polymorphisms in crop genetics. *Genome studies and molecular genetics*,2002,5: 94-100.
- [2] 胡中立,董卫国,宋运淳.人类复杂遗传疾病 QTL 分析方法的理论探讨. *遗传学报*,2002,29(2):101-104.
- [3] 刘旭.作物种质资源//刘旭.中国生物种质资源科学报告,北京:科学出版社,2003:61-80.
- [4] Zhang YM, Mao YC, Xie CQ, et al. Mapping Quantitative Trait Loci Using Naturally Occurring Genetic Variance Among Commercial Inbred Lines of Maize (*Zea mays* L.). *Genetics*, 2005,169:2267-2275.
- [5] Remington DL, Thornsberry JM, Matsuoka Y, et al. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Natl Acad Sci USA*,2001,98(20):11479-11484.
- [6] Morrell PL, Toleno DM, Lundy KE, et al. Low level of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp. *Spontaneum*) despite high rates of self-fertilization. *PNAS*,2005,102(7): 2442-2447.
- [7] Gupta PK, Rustgi S, Kulwal PL. Linkage disequilibrium and association studies in higher plants: Present status and future prospects. *Plant Mol Biol*,2005,57:461-485.
- [8] Nordborg M, Borevitz JO, Bergelson J, et al. The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nat Genet*, 2002,30:190-193.
- [9] Parisseaux B, Bernardo R. In silico mapping of quantitative trait loci in maize. *Theoretical Applied Genetics*,2004,109:508-514.
- [10] McCouch SR, Teytelman L, Xu YB, et al. Development and mapping of 2240 new SSR markers for rice (*Oryza sativa* L.). *DNA research*,2002, 9:199-207.
- [11] Agrama HA, Eizenga GC. Evaluation of linkage disequilibrium in rice and its wild relatives (abstract) // The Organizing Committee of Plant & Animal Genomes XIV Conference. Proceedings of the XIV Annual International Plant & Animal Genome Conference, San Diego, California,2006:14.
- [12] Thomson MJ, Septiningsib EM, Fatimah, et al. 2005. Exploring the genetic diversity of Indonesian rice germplasm//The International Rice Research Institute. The 5th international rice genetics symposium and 3rd international rice functional genomics symposium (Abstracts), Manila, Philippines,2005:16.
- [13] Marther KA, Caicedo AL, Olsen KM, et al. Population genomics and LD in rice.//The International Rice Research Institute. The 5th international rice genetics symposium and 3rd international rice functional genomics symposium (Abstracts),Manila, Philippines.2005:17.
- [14] Bradbury P,Buchler E,Kroon D, et al. Trait analysis by association, evolution and linkage (TASSEL) [EB/OL]. [2006]. [http:// www. Maizegenetics.net/ tassel](http://www.Maizegenetics.net/tassel).
- [15] Caldwell KS, Russell J, Langridge P, et al. Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics*,2006,172: 557-567.
- [16] Thornsberry JM, Goodman MM, Doebley J, et al Dwarf8 polymorphisms associate with variation in flowering time. *Nat Genet*, 2001,28:286-289.
- [17] Palaisa KA, Morgante M, Williams M, et al. Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. *Plant Cell*, 2003,15:1795-1806.
- [18] Wilson LM, Whitt SR, Ibanez AM, R et al. Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell*,2004,16:2719-2733.
- [19] Szalma SJ, Buckler IV ES, Snook ME, et al Association analysis of candidate genes for *maysin* and chlorogenic acid accumulation in maize silks. *Theor Appl Genet*,2005,110: 1324-1333.
- [20] Andersen JR, Zein I, Wenzel G, et al. High levels of linkage disequilibrium and associations with forage quality at a Phenylalanine Ammonia-Lyase locus in European maize (*Zea mays* L.) inbreds. *Theor Appl Genet*,2007,114:307-319.
- [21] Virk PS, Ford-lloyd BV, Jackson MT, et al. Predicting quantitative variation within rice germplasm using molecular markers. *Heredity*, 1996,76(3):296-304.
- [22] Flint-Garcia SA, ThUILlet A, Yu J, et al. Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J*,2005,44(6):1054-1064.
- [23] Yu JM, Buckler ES. Genetic association mapping and genome organization of maize. *Curr Opin Biotech*,2006,17(2):155-160.
- [24] 初少莉,朱鼎良,熊墨淼,等.糖及脂代谢调节基因与高血压病的连锁分析. *中华医学杂志*,2001,81(1):20-22.
- [25] 梁云,周初.人类基因组中的连锁不平衡方式. *国外医学(生理、病理科学与临床分册)*,2005,25(3):247-250.
- [26] Risch N, Merikangas K.. The future of genetic studies of complex human diseases. *Science*, 1996,273:1516-1517.
- [27] Meuwissen T H E, Karlsen A, Lien S. Fine mapping of a quantitative trait locus for twinning rate using combined linkage and linkage disequilibrium mapping. *Genetics*, 2002,161:373-379.
- [28] Olsen KM, Halldorsdottir SS, Stinchcombe JR, et al.. Linkage disequilibrium mapping of *Arabidopsis CRY2* flowering time alleles. *Genetics*,2004,167:1361-1369.
- [29] Tenaillon MI, Sawkins MC, Anderson LK, et a. Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Genetics*, 2002,162:1401-1413.
- [30] Dooner HK, Weck E, Favreau MA. A molecular genetic analysis of insertion mutations in the bronze locus in maize. *Mol Gen Genet*, 1985,200,240-246.
- [31] Drouaud J, Camilleri C, Bourguignon P, et al. Variation in crossing-over rates across chromosome 4 of *Arabidopsis thaliana* reveals the presence of meiotic recombination "hot spots". *Genome Res*,2006, 16:106-114.
- [32] Carlborg Ö, Haley CS. Epistasis: too often neglected in complex trait studies. *Nat Rev Genet*,2004,5:618-625.
- [33] Lou XY, Casella G, Littell RC, et al. A haplotype-based algorithm for multilocus linkage disequilibrium mapping of quantitative trait loci with epistasis. *Genetics*,2003,163:1533-1548.