

DOI: 10.1360/yc-007-1317

植物基因组中的连锁不平衡

王荣焕, 王天宇, 黎裕

中国农业科学院作物科学研究所, 北京 100081

摘要: 在植物基因组学研究领域, 连锁不平衡(linkage disequilibrium, LD)分析是近年来一个研究亮点和热点。基于 LD 的作图方法不仅是新基因发掘的有效途径, 而且也是联系结构基因组学和表型组学的一座桥梁。自 2001 年基于 LD 的作图方法在植物中的成功运用至今, 已有大量关于植物基因组中 LD 结构及 LD 作图的研究报道。文章系统介绍了 LD 的基本理论及其在 LD 作图、单倍型多样性分析、单倍型标签 SNP 的开发和群体遗传学等研究中的应用, 并就近年来关于 LD 与群体结构、基因转换和上位效应及 G×E 互作等方面的研究热点和发展趋势进行了探讨。当前, 世界各国基因争夺大战日趋激烈。中国是基因资源大国, 但还不是基因大国。植物基因组中 LD 研究热潮的兴起及 LD 研究的进一步深入, 必将大大推动植物基因组学的快速发展, 特别是加速从作物种质资源中发掘新基因的进程。

关键词: 连锁不平衡; LD 结构; LD 作图

Linkage disequilibrium in plant genomes

WANG Rong-Huan, WANG Tian-Yu, LI Yu

Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China

Abstract: Linkage disequilibrium (LD) is one of the most recently focused interests in the field of plant genomics. LD mapping is an effective approach to discovering novel genes and a bridge for connecting structural genomics to phenomics. LD mapping was first applied in plants in 2001. Since then, researches on the structure and extent of LD and LD mapping have been reported in a wide range of plant species. The basic theory of LD and its application in LD mapping, haplotype diversity analysis, htSNP identification and population genetics were reviewed in this paper. And advances of LD research in plants including influences of population structure, gene conversion, epistasis and G×E interactions, and future prospects were also presented. China has abundant germplasm resources, but gene discovery lags behind. Intensive researches on LD will certainly accelerate rapid development of plant genomics, especially the progress of gene discovery based on germplasm resources in China.

Keywords: linkage disequilibrium (LD); LD structure; LD mapping

我国现收集保存作物种质资源共 38 万份, 居世界第一位, 是名副其实的基因资源大国, 但还不是基因大国^[1]。抢占基因种质资源制高点引发了基因争夺这场没有硝烟的世界大战。如何从我国丰富的

收稿日期: 2007-04-09; 修回日期: 2007-06-01

基金项目: 国家高技术研究发展计划项目(863 计划) (编号: 2006AA10Z188) 和国家自然科学基金项目(编号: 30571133) 资助[Supported by the Hi-Tech Research and Development Program of China (863 Program) (No. 2006AA10Z188) and the National Natural Science Foundation of China (No.30571133)]

作者简介: 王荣焕(1980-), 女, 河北衡水人, 在读博士, 研究方向: 基于基因组学的作物种质资源研究。Tel: 010-62186652; E-mail: ronghuanwang@126.com

通讯作者: 黎裕(1966-), 男, 四川仪陇人, 研究员, 博士, 研究方向: 玉米种质资源和基因组学研究。Tel: 010-62131196; E-mail: yuli@mail.caas.net.cn

种质资源中挖掘优异基因, 尽快实现我国由种质资源优势向基因资源优势的转变是新时期摆在我们面前的一个迫切需要解决的问题。随着植物基因组学的迅猛发展, 新的基因发掘方法不断涌现^[1]。近年来, 连锁不平衡(Linkage disequilibrium, LD)作图方法已被证明不仅是基因发掘, 而且也是等位基因发掘的有效手段^[2]。

另一方面, 揭示植物基因型和表型之间的内在联系也是植物基因组学研究的重要内容。在过去的几十年中, 人们分别从植物的形态学特征、生理生化特性和分子机理等不同水平上进行了大量多层次研究和不懈努力。随着人们认知的深入, 探索表型和基因型之间内在联系的共识加速了植物生命现象的深入研究。植物基因组中多态性位点间的 LD 是联系结构基因组学和表型组学的一座桥梁。LD 为植物基因组中新基因的发掘及揭示特定基因型和表型之间的内在联系提供了一个全新的契机。

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, 引起了近年来科研工作者对 LD 研究的强烈兴趣。本文将对 LD 的基础理论及其在植物基因组学中的应用、目前研究热点及展望进行全面介绍。

1 连锁不平衡的理论基础

1.1 连锁不平衡的概念

连锁不平衡(linkage disequilibrium, LD)亦被称为配子相不平衡(gametic phase disequilibrium)、配子不平衡(gametic disequilibrium)或等位基因关联(allelic association), 指的是一个群体内不同座位等位基因之间的非随机关联, 包括两个标记间或两个基因/QTL间或一个基因/QTL与一个标记座位间的非随机关联^[3]。

连锁不平衡与连锁是相关但完全不同的两个概念。连锁不平衡指的是群体内等位基因之间的相关, 而连锁指的是位于同一条染色体上的基因联合传递的现象。紧密连锁可导致较高的 LD 水平, 但这种 LD 纯粹是由突变产生的等位基因出现后紧密连锁座位间所有重组事件的结果。

1.2 连锁不平衡的度量

所有 LD 统计的是实际观测到的单倍型频率与随机分离时单倍型的期望频率之间的差异。LD 的计算依研究座位的性质和数目而异。

对于只有两个等位基因的座位如 SNP 和 AFLP, 通常用 r^2 和 D' 来估计两个座位之间的 LD 水平^[4]。假设有两个连锁的座位 A 和 B , 其等位基因分别为 A 、 a 和 B 、 b , 4 个等位基因的频率分别为 π_A 、 π_a 、 π_B 、 π_b , 4 种单倍型 AB 、 aB 、 Ab 和 ab 的频率分别为 π_{AB} 、 π_{aB} 、 π_{Ab} 和 π_{ab} 。那么, 实际观测到的单倍型频率与期望单倍型频率之间的差异 D 的计算公式为:

$$D_{ab} = (\pi_{AB} - \pi_A \pi_B)$$

r^2 的计算公式为

$$r^2 = \frac{(D_{ab})^2}{\pi_A \pi_a \pi_B \pi_b}$$

D' 的计算公式为

$$|D'| = \frac{(D_{ab})^2}{\min(\pi_A \pi_b, \pi_a \pi_B)} \text{ for } D_{ab} < 0$$

$$|D'| = \frac{(D_{ab})^2}{\min(\pi_A \pi_B, \pi_a \pi_b)} \text{ for } D_{ab} > 0$$

r^2 和 D' 反映了 LD 的不同方面。 r^2 包括了重组史和突变史, 而 D' 仅包括重组史。 D' 能更准确地估测重组差异, 但样本较小时发现低频率 4 种等位基因组合的可能性大大减小, 因此 D' 不适宜小样本研究中的应用。 r^2 可以提供标记是否能与 QTL 相关的信息, 因此 LD 作图中通常采用 r^2 来表示群体的 LD 水平。

对于 SSRs 和 RFLPs 等有多等位基因的座位, 一种转化形式的 D' 是应用最广泛的衡量两个多等位基因座位间 LD 水平的值。在实际应用中, 我们经常需要计算的是有两个或多个等位基因的两座位间的 LD 水平, 但当构建全基因组的 LD 图谱时就需要考虑多个座位间的 LD 水平。多个座位间 LD 水平的计算包括 bottom-up 和 top-down 两种方法^[5], 这里不再详述。

r^2 和 D' 是两个座位间 LD 的度量。对于基因组内某区域的 LD 分布状况, 通常用两种形象化的方式来表示: LD 衰退图和 LD 矩阵。LD 衰退图是以位点间的 LD 对遗传距离作图来表示一个区域内的 LD 分布情况, 这种表示方法也便于对不同物种中的 LD 水平进行比较。LD 矩阵是某基因内或某染色体上多态性位点间 LD 的线性排列。

2 连锁不平衡的影响因素

通常, 随机匹配群体中, 在没有选择、突变或迁移因素的影响时, 多态性位点处于连锁平衡状态^[5], 相反, 连锁、选择和群体混合将增加LD的水平。群体中的LD水平是许多遗传因素和非遗传因素综合作用的结果。突变可导致新的多态性产生, 而重组则可通过重新组合序列变异而削弱染色体内部的LD, LD的程度与重组率成反比^[6]。植物中影响LD水平的因素主要包括:

2.1 杂交类型

不同杂交类型植物间的LD水平存在很大差异。在拟南芥^[7,8]、水稻^[9]、大麦^[10]和大豆^[11]等自交物种中, 个体绝大多数为纯合子, 虽然重组仍然发生但不再对LD产生任何影响, 即其有效重组率较低, 因此这些物种在很长的物理距离内(可达几百kb)存在LD^[8,12]。与自交物种相比, 异交物种如玉米中有效重组率高, 重组导致连锁的位点彼此独立存在, 从而削弱染色体内部的LD, 因此异交物种中的LD迅速衰减^[13]。

有两个值得注意的问题是: 某些自交物种如大麦^[14]虽然是由同质个体组成, 但在群体水平却具有很高的遗传多样性。另外, 自然进化进程和人为介入可改变某物种的杂交类型。如栽培大豆的异交率为1%, 而其祖先的异交率高达13%^[15]。异交率的改变将影响群体的LD水平。

2.2 群体特性

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的一个明显特性是群体依赖性。即使来自同一物种的不同群体也可能有明显不同的LD特性。影响群体大小的瓶颈效应将致使仅仅少数等位基因组合能够传递到后代中, 而低频率多态性则丧失掉, 因此其LD水平大大增加。但在没有其他减轻因素(比如群体亚结构)的情况下, 这种影响是短期的^[16]。选择的群体不同, 其LD水平显著不同。多样性较高的群体包括更多不同来源的研究个体, 因此其LD水平较低; 而当所用群体来源有限时, 其LD将维持在一个较高水平。如玉米中, 地方品种在600 bp范围内存在LD衰减^[17]; 不同育种自交系在2 000 bp范围内存在LD衰减^[18]; 而骨干自交系在100 000 bp范围内存在LD^[13]。这些

结论的差异来自所用群体的差异。另外, 群体混合可以通过引进不同祖先来源和等位基因频率的染色体而影响群体的LD水平。

2.3 选择和驯化

对某物种的正向选择和驯化可增加其LD水平。对某特定等位基因的强烈选择(即座位特异的瓶颈效应)限制了该座位周围的遗传多样性, 因此导致所选择基因周围区域的LD水平增加^[19]。人工选择对多样性和LD影响的一个显著例子是玉米基因组中的*y1*座位^[20]。玉米胚乳有黄色和白色两种, 其祖先大刍草的胚乳为白色。黄色胚乳因含有较高的类胡萝卜素, 营养价值高, 因此后来育种家才开始了对黄色胚乳的选择。*Y1*是与玉米黄色胚乳有关的编码八氢番茄红素合酶的显性等位基因, 其上调作用导致黄色胚乳类胡萝卜素含量大大提高。对许多黄色和白色胚乳玉米品种此座位的序列分析发现, 由于选择的作用, 黄色等位基因*Y1*比白色等位基因*y1*的多样性低19倍, 且距其500 kb的范围内均受到选择引起的多样性降低的影响。

3 连锁不平衡的应用

3.1 LD作图

LD的最主要应用是进行LD作图来发掘和定位QTL^[21]、发掘基因内引起表型变异的多态性和进行功能标记的开发^[20,22]。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^[18]。

LD作图包括基于全基因组扫描和基于候选基因的两种策略。针对特定群体, LD作图策略的选择及其分辨率取决于目标群体内的LD程度及等位基因或单倍型的频率^[23]。如果目标群体在很长的物理距离内存在LD, 那么此群体适宜采用基于全基因组扫描的策略; 相反, 如果目标群体的LD在很短的物理距离内迅速衰退, 则此群体适宜采用基于候选基因的高分辨率LD作图策略。最近, 在玉米^[8]和大麦^[14]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中进行了成功运用^[31], 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^[21]、胚乳颜色^[20]、淀粉合成^[24]、maysin和绿原酸含量^[25]、饲用品质^[26]等性状变异之间的关系进行了研究。尽管如此, 我们仍不能完全放弃传统的连锁作图方法。对于遗传多样性较低的物种, 即使最理想的种质收集也不能包含足够多的多样性以完全解决LD作图中统计能力降低的问题, 在这种情况下连锁分析比LD作图更具优越性^[27]。而某些情况下, LD作图与传统QTL作图的整合更有利于对数量性状的深入剖析。2001年Wu等^[28]提出了一种在异交物种中整合连锁和LD优势的新的作图策略, 并且这一新策略在QTL精细作图中更为有效^[29]。

3.2 单倍型分析和标签 SNP 的开发

LD的一个重要应用是发掘基因内的单倍型区块和由不同等位基因组合所确定的单倍型类型。单倍型^[30]指基因组内处于LD状态的一组紧密连锁的等位基因, 其不易受重组的影响, 而是作为一个整体或一个单元遗传。几个多态性位点可以组成特定的单倍型, 并且较低的单倍型多样性有利于仅用少数的单倍型标签SNP(htSNP)或标签SNP(tSNP)来区分不同的单倍型。单倍型可被用来进行群体内单倍型多样性的分析、htSNP或tSNP的开发及基于单倍型的LD作图。

Malysheve-Otto等^[31]对 493 份栽培大麦材料中胚乳特异 β -淀粉酶基因*Bmy1*的单倍型多样性进行了研究, 共发现了6种单倍型类型, 其中2种单倍型是在此次研究中新发现的。并且该研究表明处于LD状态的3个SNP位点(分别对应*Bmy1*蛋白质序列的115、233和347位置)可以作为单倍型标签SNP来区分493份材料中91.6%的多样性和4种单倍型。Olsen等^[32]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 A^Q、HAP A^S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 A^Q单倍型而言, HAP A^S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的LD作图对QTL进行精细作图的成功运用。

3.3 群体遗传学研究

LD一直是群体遗传学的研究热点, LD在研究群体进化方面具有重要作用。选择是物种的主要进化动力, 而驯化是选择的一种方式。中性理论认为群体进化过程中物种内和物种间的大部分多态性是中性的。但驯化物种所经历的“驯化瓶颈效应”即对目标基因的人工选择导致了某些座位遗传多样性的降低和其周围区域LD水平的增加, 并且如果被选择表型由存在上位效应的两个非连锁的座位所控制, 那么选择会导致这两个物理距离上不相邻座位间的LD。某物种座位间多样性降低的程度和LD变化模式有利于我们更好地了解其驯化历史。

玉米是在10 000年前由大刍草驯化而来。2005年, 来自*Science*的报道表明玉米中2%~4%的基因在驯化过程中经历了人工选择^[33]。玉米与其祖先大刍草形态的显著区别是人类在玉米驯化过程中对5个基因的选择所造成的^[34], 其中*tb1*是起关键作用的基因。对玉米和大刍草中*tb1*基因的多样性研究表明^[35], 其启动子区域受到强烈的选择作用, 此基因的多样性大大降低并引起了等位基因的关联。与*Y1*基因不同的是, 对*tb1*基因启动子区域的选择并没有影响到该基因上游163 kb基因组区域内的多样性。虽然此163 kb范围内并不包含任何基因, 但是距其启动子上游最近的一个基因也是选择中性座位^[36]。另外, 受选择的影响玉米中的*su1*座位在7 000 bp范围内存在显著的LD, 而其他基因仅在2 000 bp范围内存在LD^[20]。

由于基因进化史的差异所导致的LD在基因组中的间断分布也是定位与适应性有关基因和驯化过程的有力工具^[37]。正向选择或平衡选择对某些座位曾经产生了巨大影响。如大麦中经历正向选择的*hmb-1*基因^[10]和*adh1*基因^[38, 39]、拟南芥中与适应性有关的*FR1*基因^[8, 40]和经历平衡选择的*CLV2*和*RPS5*座位^[41, 42]均是由于基因进化史差异所导致的基因内LD水平的变化。

4 连锁不平衡的研究热点及展望

4.1 群体结构与LD作图

运用LD作图解析性状需要具有不同LD水平的群体, 包括适用于全基因组扫描的、具有较高LD水平、但低分辨率的群体和适用于基于候选基因的、LD衰退迅速、分辨率较高的作图群体^[43]。现存的种

质群体基本能满足这种需要, 但群体结构的存在限制了LD作图的成功运用。这是因为许多重要作物有着复杂的进化和育种历史, 再加上大部分野生植物有限的基因漂流导致了种质材料中存在着复杂的群体结构, 从而使关联分析更加复杂^[44]。

群体结构的存在和亚群内等位基因频率的不均匀分布将导致多态性位点和表型的假阳性关联。在玉米中, 对 *D8*^[45,46] 和 *sh2*^[24] 序列多样性分别与开花时间和淀粉总量的关联研究中都发现了群体结构对关联结果的影响, 即当考虑群体结构时, 有些不考虑群体结构时发现的关联就不存在了。近年来, 统计方法的发展使运用独立标记检测和校正群体结构来恰当解释关联分析的结果成为了可能。

然而值得注意的是: 群体结构的校正实际上还可能引起基因与表型关联的丧失^[25]。这是因为(1)所检测到的多态性是没有功能的多态性, 并且这种关联是由群体结构引起的; (2)所检测到的多态性确实是功能多态性, 只是其多态性的分布与群体结构恰相吻合。分析群体结构对表型性状的影响对于防止假阳性关联是非常有必要的。如果所构建的关联群体中群体结构解释太多比例的表型变异, 那么用此关联群体检测单个基因效应的功效甚微。在这种情况下, 就需要用其他群体对所检测到的多态性进行再次评价。

4.2 重组与 LD

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^[47]。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* 基因内部的重组率比整个基因组的平均重组水平要高出 100 倍^[48]。

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^[49]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。Haubold等^[50]发现, 拟南芥中 170 kb 基因组范围内的 14 个座位中 90% 的重组是由基因转变引起的。Morrell等^[51]研究发现, 来自大麦、玉米和果蝇的共 27 个座位中有 13 个座位存在基因转变现象, 并且由基因转变而引起的重组是由交换所引起重组的两倍。Plagnol等^[52]的研究表明 96 份拟南芥材料中 1347 个片段的基因转变与交换的比率大约是 1。理论上, 相距很近的两个座位是处于完全连锁和完全LD状态的, 但基因转变现象的存在将导致两个座位间的不完全LD。

4.3 上位互作和 G×E 互作与 LD

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所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^[53]使存在上位互作座位的发掘成为了可能, 而上位互作导致的相关座位间LD的存在为更好地研究上位效应及基因型与环境的互作提供了一条新的思路。Lou等^[54]提出了一种新的在自然群体中鉴定加性、显性和上位性效应座位的统计方法, 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。

在植物基因组学研究领域, LD是近年来的一个研究亮点和热点。尽管人们对植物中的LD有了初步了解并已有多项成功运用, 但实际上植物中的LD研究刚刚起步。LD作图与连锁作图的有效整合、强大统计模型的开发和运用、单倍型标签SNP(htSNP)的发掘及植物基因组LD图谱的构建^[3]等将是今后植物中LD研究的重要方向。总之, 植物基因组中LD研究热潮的兴起, 必将大大推动植物基因组学的飞跃发展及我国丰富的植物种质资源中新基因发掘的研究进程。

参考文献(References):

- [1] JIA Ji-Zeng, LI Yu. Plant genomics and gene discovery in germplasm resources. *Scientia Agricultura Sinica*, 2004, 37(11): 1585–1592.

- 贾继增, 黎裕. 植物基因组学与种质资源新基因发掘. 中国农业科学, 2004, 37(11): 1585–1592.
- [2] Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D. *Dwarf8* polymorphisms associate with variation in flowering time. *Nat Genet*, 2001, 28: 286–289. [\[DOI\]](#)
- [3] 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. [\[DOI\]](#)
- [4] Flint-Garcia SA. Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol*, 2003, 54: 357–374. [\[DOI\]](#)
- [5] Falconer DS, Mackay TFC. Introduction to Quantitative Genetics (Fourth Edition). Harlow: Adison Welsey Longman Limited, 1996.
- [6] Nachman MW. Variation in recombination rate across the genome: evidence and implications. *Curr Opin Genet Dev*, 2002, 12(6): 657–663. [\[DOI\]](#)
- [7] Shepard KA, Purugganan MD. Molecular population genetics of the *Arabidopsis CLAVATA2* region: The genomic scale of variation and selection in a selfing species. *Genetics*, 2003, 163: 1083–1095.
- [8] Nordborg M, Borevitz JO, Bergelson J, Berry CC, Chory J, Hagenblad J, Kreitman M, Maloof JN, Noyes T, Oefner PJ, Stahl EA, Weigel D. The extent of linkage disequilibrium in *Arabidopsis thaliana*. *Nat Genet*, 2002, 30: 190–193. [\[DOI\]](#)
- [9] Garris AJ, McCouch SR, Kresovich S. Population structure and its effect on haplotype diversity and linkage disequilibrium surrounding the *xa5* locus of rice (*Oryza sativa* L.). *Genetics*, 2003, 165: 759–769.
- [10] Caldwell KS, Russell J, Langridge P, Powell W. Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics*, 2006, 172: 557–567. [\[DOI\]](#)
- [11] Zhu YL, Song QJ, Hyten DL, Van Tassell CP, Matukumalli LK, Grimm DR, Hyatt SM, Fickus EW, Young ND, Cregan PB. Single nucleotide polymorphisms in soybean. *Genetics*, 2003, 163: 1123–1134.
- [12] Morrell PL, Toleno DM, Lundy KE, Clegg MT. Low levels of linkage disequilibrium in wild barley (*Hordeum vulgare* ssp *spontaneum*) despite high rates of self-fertilization. *Proc Natl Acad Sci USA*, 2005, 102(7): 2442–2447. [\[DOI\]](#)
- [13] Nordborg M. Linkage disequilibrium, gene trees and selfing: an ancestral recombination graph with partial self-fertilization. *Genetics*, 2000, 154: 923–929.
- [14] Caldwell KS, Russell J, Langridge P, Powell W. Extreme population-dependent linkage disequilibrium detected in an inbreeding plant species, *Hordeum vulgare*. *Genetics*, 2006, 172: 557–567. [\[DOI\]](#)
- [15] Fujita R, Ohara M, Okazaki K, Shimamoto Y. The extent of natural cross-pollination in wild soybean (*Glycine soja*). *J Hered*, 1997, 88: 124–128.
- [16] Wall JD, Andolfatto P, Przeworski M. Testing models of selection and demography in *Drosophila simulans*. *Genetics*, 2002, 162:203–216.
- [17] Tenailon MI, Sawkins MC, Long AD, Gaut RL, Doebley JF, Gaut BS. Patterns of DNA sequence polymorphism along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Proc Natl Acad Sci USA*, 2001, 98(16): 9161–9166. [\[DOI\]](#)
- [18] Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES. Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc Natl Acad Sci USA*, 2001, 98(20):11479–11484. [\[DOI\]](#)
- [19] Przeworski M. The signature of positive selection at randomly chosen loci. *Genetics*, 2002, 160: 1179–1189.
- [20] Palaisa KA, Morgante M, Williams M, Rafalski A. Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. *Plant Cell*, 2003, 15: 1795–1806. [\[DOI\]](#)
- [21] Meuwissen THE, Goddard ME. Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics*, 2000, 155: 421–430.
- [22] Palaisa K, Morgante M, Tingey S, Rafalski A. Long-range patterns of diversity and linkage disequilibrium surrounding the maize *Y1* gene are indicative of an asymmetric selective sweep. *Proc Natl Acad Sci USA*, 2004, 101(26): 9885–9890. [\[DOI\]](#)
- [23] Rafalski A. Applications of single nucleotide polymorphisms in crop genetics and breeding. *Curr Opin Plant Biol*, 2002, 5: 94–100. [\[DOI\]](#)
- [24] Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM, Buckler IV ES. Dissection of maize kernel composition and starch production by candidate gene association. *Plant Cell*, 2004, 16: 2719–2733. [\[DOI\]](#)
- [25] Szalma SJ, Buckler IV ES, Snook ME, McMullen MD. Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks. *Theor Appl Genet*, 2005, 110: 1324–1333. [\[DOI\]](#)
- [26] Andersen JR, Zein I, Wenzel G, Krutzfeldt B, Eder J, Ouzunova M, Lübberstedt T. 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. [\[DOI\]](#)
- [27] Flint-Garcia SA, ThUILlet AC, Yu JM, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES. Maize association population: a high-resolution platform for quantitative trait locus dissection. *Plant J*, 2005, 44: 1054–1064. [\[DOI\]](#)
- [28] Wu RL, Ma CX, Casella G. Joint linkage and linkage disequilibrium mapping of quantitative trait loci in natural populations. *Genetics*, 2002, 160: 779–792.
- [29] Wu RL, Zeng ZB. Joint linkage and linkage disequilibrium

- rium mapping in natural populations. *Genetics*, 2001, 157: 899–909.
- [30] Wang QH, Dooner H. Remarkable variation in maize genome structure inferred from haplotype diversity at the *bz* locus. *Proc Natl Acad Sci USA*, 2006, 103(47): 17644–17649. [\[DOI\]](#)
- [31] Malysheva-Otto LV, Röder. Haplotype diversity in the endosperm specific β -amylase gene *Bmy1* of cultivated barley (*Hordeum vulgare* L.). *Mol Breeding*, 2006, 18: 143–156. [\[DOI\]](#)
- [32] Olsen KM, Halldorsdottir SS, Stinchcombe JR, Weinig C, Schmitt J, Purugganan MD. Linkage disequilibrium mapping of *Arabidopsis* *CRY2* flowering time alleles. *Genetics*, 2004, 167: 1361–1369. [\[DOI\]](#)
- [33] Wright SI, Irie Vroh Bi, Schroeder SG, Yamasaki M, Doebley JF, McMullen MD, Gaut BS. The effects of artificial selection on the maize genome. *Science*, 2005, 308: 1310–1314. [\[DOI\]](#)
- [34] Doebley J, Stec A, Wendel J, Edwards M. Genetic and morphological analysis of a maize-teosinte F2 population: Implications for the origin of maize. *Proc Natl Acad Sci USA*, 1990, 87: 9888–9892. [\[DOI\]](#)
- [35] Wang RL, Stec A, Hey J, Lukens L, Doebley J. The limits of selection during maize domestication. *Nature*, 1999, 398: 236–239. [\[DOI\]](#)
- [36] Clark RM, Linton E, Messing J, Doebley JF. Pattern of diversity in the genomic region near the maize domestication gene *tb1*. *Proc Natl Acad Sci USA*, 2004, 101(3): 700–707. [\[DOI\]](#)
- [37] Kraakman, ATW, Niks RE, Van den Berg P, Stam P, Van Eeuwijk FA. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics*, 2004, 168: 435–446. [\[DOI\]](#)
- [38] Lin JZ, Brown AHD, Clegg MT. Heterogeneous geographic patterns of nucleotide sequence diversity between two alcohol dehydrogenase genes in wild barley (*Hordeum vulgare* subspecies *spontaneum*). *Proc Natl Acad Sci USA*, 2001, 98(2): 531–536. [\[DOI\]](#)
- [39] Lin JZ, Morrell PL, Clegg MT. The influence of linkage and inbreeding on patterns of nucleotide sequence diversity at duplicate alcohol dehydrogenase loci in wild barley (*Hordeum vulgare* ssp. *spontaneum*). *Genetics*, 2002, 162: 2007–2015.
- [40] Johanson U, West J, Lister C, Scott M, Amasino R, Dean C. Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. *Science*, 2000, 290: 344–347. [\[DOI\]](#)
- [41] Tian DC, Araki H, Stahl E, Bergelson J, Kreitman M. Signature of balancing selection in *Arabidopsis*. *Proc Natl Acad Sci USA*, 2002, 99(17): 11525–11530. [\[DOI\]](#)
- [42] Shepard KA, Purugganan MD. Molecular population genetics of the *Arabidopsis* *CLAVATA2* region: The genomic scale of variation and selection in a selfing species. *Genetics*, 2003, 163: 1083–1095.
- [43] Rafalski A, Morgante M. Corn and humans: recombination and linkage disequilibrium in two genomes of similar size. *Trends Genet*, 2004, 20(2): 103–111. [\[DOI\]](#)
- [44] Sharbel TF, Haubold B, Mitchell-Olds T. Genetic isolation by distance in *Arabidopsis thaliana*: biogeography and postglacial colonization of Europe. *Mol Ecol*, 2000, 9: 2109–2118. [\[DOI\]](#)
- [45] Andersen JR, Schrag T, Melchinger AE, Zein I, Lübberstedt T. Validation of *Dwarf8* polymorphisms associated with flowering time in elite European inbred lines of maize (*Zea mays* L.). *Theor Appl Genet*, 2005, 111: 206–217. [\[DOI\]](#)
- [46] Camus-Kulandaivelu L, Veyrieras JB, Madur D, Combes V, Fourmann M, Barraud S, Dubreuil P, Gouesnard B, Manicacci D, Charcosset A. Maize adaptation to temperate climate: relationship with population structure and polymorphism in the *Dwarf8* gene. *Genetics*, 2006, 172: 2449–2463. [\[DOI\]](#)
- [47] Tenailon MI, Sawkins MC, Anderson LK, Stach SM, Doebley J, Gaut BS. Patterns of diversity and recombination along chromosome 1 of maize (*Zea mays* ssp. *mays* L.). *Genetics*, 2002, 162: 1401–1413.
- [48] 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. [\[DOI\]](#)
- [49] Drouaud J, Camilleri C, Bourguignon P, Canaguier A, Bérard A, Vezon D, Giancola S, Brunel D, Colot V, Prum B, Quesneville H, Mézard C. 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. [\[DOI\]](#)
- [50] Harbold B, Kroymann J, Ratzka A, Mitchell-Olds T, Wiehe T. Recombination and gene conversion in a 170-kb genomic region of *Arabidopsis thaliana*. *Genetics*, 2002, 161: 1269–1278.
- [51] Morrell PL, Toleno DM, Lundy KE, Clegg MT. Estimating the contribution of mutation, recombination and gene conversion in the generation of haplotypic diversity. *Genetics*, 2006, 173: 1705–1723. [\[DOI\]](#)
- [52] Plagnol V, Padhukasahasram B, Wall JD, Marjoram P, Nordborg M. Relative influences of crossing over and gene conversion on the pattern of linkage disequilibrium in *Arabidopsis thaliana*. *Genetics*, 2006, 172: 2441–2448. [\[DOI\]](#)
- [53] Carlborg Ö, Haley CS. Epistasis: too often neglected in complex trait studies. *Nat Rev Genet*, 2004, 5: 618–625. [\[DOI\]](#)
- [54] Lou XY, Casella G, Littell RC, Yang MCK, Johnson JA, Wu R. A haplotype-based algorithm for multilocus linkage disequilibrium mapping of quantitative trait loci with epistasis. *Genetics*, 2003, 163: 1533–1548.