

韩俊杰,王昊龙,李卫华. 关联分析及其在不同分子标记中的应用综述[J]. 江苏农业科学,2016,44(2):13-17.
doi:10.15889/j.issn.1002-1302.2016.02.004

关联分析及其在不同分子标记中的应用综述

韩俊杰,王昊龙,李卫华

(石河子大学农学院/新疆生产建设兵团绿洲生态农业重点实验室,新疆石河子 832003)

摘要:关联分析技术在植物中的应用较晚,近些年才在植物数量性状和植物遗传育种研究中得以应用。关联分析以连锁不平衡作图(LD mapping)为基础,用来鉴定植物性状与目标基因之间的关系。关联分析有很明显的应用优势,它不仅不需要特殊的构图群体,而且还能够同时对多个基因位点进行检测,精度非常高。由于其良好的实用性,关联分析技术已经成为国际基因组学的研究热点。介绍了 LD 的定义和度量方法,综述了基于分子标记的关联分析在植物遗传中的应用,并对其在遗传学中的应用和发展作了展望。

关键词:植物;分子标记;关联分析;连锁不平衡;遗传学;应用

中图分类号: Q78 **文献标志码:** A **文章编号:** 1002-1302(2016)02-0013-05

关联分析又称连锁不平衡作图(linkage disequilibrium mapping, LD mapping),是建立在连锁不平衡的基础上用于鉴定群体内目标性状与遗传标记相关关系的方法^[1]。关联分析最早在人类疾病的研究中得以应用,并取得了丰硕的成果,目前在人类不同致病基因方面已有许多成功的报道^[2]。在动物中的应用也相对较多,如在奶牛、鱼、羊、猪、鸡中的应用等。然而,由于目前对植物基因组 LD 结构的认识还不够全面,导致关联分析在植物中的应用相对较少^[3]。但是,随着生物信息学、统计学和基因组学技术的快速发展,关联分析也已成功应用到许多植物上,如玉米、小麦、拟南芥、大麦、大豆等。近年来,许多植物的全基因组测序已经完成,这极大地促

进了基因组学的发展,特别是近年来大量单核苷酸多态性(single nucleotide polymorphisms, SNP)标记的开发,关联分析在植物关键基因的发掘上更是得到了广泛的应用,可为植物分子育种提供标记,在国际上已经成为热点研究方向。

关联分析通过群体上的进化历史和重组,可以在序列水平上来研究复杂的数量性状,因此与其他分析手段相比,关联分析有其独特的优势:(1)所需时间短,用自然群体作为材料,不需要专门构建,且有较大的选择范围,相对于数量性状基因座(QTL)至少需要 2 年以上的群体能节约大量的时间;(2)检测范围大,同一座位的不同基因可以用来同时检测,而 QTL 作图只能检测来自双亲的 2 个等位基因;(3)准确度高, QTL 作图精度一般在 10~30cM 左右,很少能在基因水平上被标记,而关联分析可达到单基因的水平^[4-5],如 Gebhardt 等利用关联分析方法将马铃薯抗晚疫病基因 *RI* 定位在 0.2cM 的基因组区段内^[6]。在植物相关研究中,由于其杂交方式和重复类型都可以得到人为有效控制,因此关联分析在植物中的应用比在动物中的吸引力更高。

作物很多重要的农艺性状(如株高、产量、营养品质和抗

收稿日期:2015-01-05

基金项目:国家自然科学基金(编号:31260357)。

作者简介:韩俊杰(1989—),男,河南开封人,硕士,主要从事小麦品质遗传改良研究。E-mail:hanjunjie1208@sina.com。

通信作者:李卫华,博士,教授,主要从事小麦品质改良和分子机理研究。E-mail:lw_h_agr@shzu.edu.cn。

[33] Fukuoka H, Yamaguchi H, Nunome T, et al. Accumulation, functional annotation, and comparative analysis of expressed sequence tags in eggplant (*Solanum melongena* L.), the third Pole of the genus *Solanum* species after tomato and potato [J]. *Gene*, 2010, 450 (1/2): 76-84.

[34] Yu Y, Liang Y, Lv M, et al. Genome-wide identification and characterization of polygalacturonase genes in *Cucumis sativus* and *Citrullus lanatus* [J]. *Plant Physiology and Biochemistry*, 2014, 74: 263-275.

[35] Qi J, Liu X, Shen D, et al. A genomic variation map provides insights into the genetic basis of cucumber domestication and diversity [J]. *Nature Genetics*, 2013, 45 (12): 1510-1515.

[36] Shirasawa K, Fukuoka H, Matsunaga H, et al. Genome-wide association studies using single nucleotide polymorphism markers developed by re-sequencing of the genomes of cultivated tomato [J]. *DNA Research; an International Journal for Rapid Publication of*

Reports on Genes and Genomes, 2013, 20 (6): 593-603.

[37] Zenoni S, Ferrarini A, Giacomelli E, et al. Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq [J]. *Plant Physiology*, 2010, 152 (4): 1787-1795.

[38] Chagné D, Gasic K, Crowhurst R N, et al. Development of a set of SNP markers present in expressed genes of the Apple [J]. *Genomics*, 2008, 92 (5): 353-358.

[39] Schnable P S, Ware D, Fulton R S, et al. The B73 maize genome: complexity, diversity, and dynamics [J]. *Science*, 2009, 326 (5956): 1112-1115.

[40] Paterson A H, Bowers J E, Chapman B A. Ancient polyploidization predating divergence of the cereals, and its consequences for comparative genomics [J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2004, 101 (26): 9903-9908.

病性等)都属于数量性状,了解这些数量性状的遗传规律对作物的改良有重要意义。目前,人们主要是通过全基因组图谱和 QTL 定位来了解植物数量性状,而关联分析是 QTL 定位的有效手段,在分子育种中发挥了重要作用。近年来,随着新一代分子标记——SNPs 标记技术的出现以及植物基因组测序技术日趋成熟,植物遗传学研究已经进入基因组研究时代。如今,在许多重要作物中都已经应用到关联分析技术^[7-8],不仅应用到模式植物、三大粮食作物,还在甘蔗、大豆、马铃薯、甜菜、番茄和向日葵等其他植物中得到大量的应用。

1 关联分析的统计学意义

1.1 LD 连锁不平衡

Jennings 于 21 世纪初期提出了 LD 这一概念^[9],这是自然选择中的一种现象。在某一群体中,如果 2 个等位基因同时出现的概率与期望值不相同时,那么它们就处于 LD 状态。连锁不平衡和遗传连锁是不相同的,遗传连锁指的是由于物理作用而导致的不同位点不分离的现象,而连锁不平衡则是等位基因间的作用。当 2 个基因位点紧密连锁时,就会导致 LD 值处在比较高的水平。关联分析是基于 LD 的,对 LD 结构的了解是关联分析的前提。因此影响 LD 的因素同样影响关联分析的结果,这些因素有选择、突变、重组、遗传漂变、群体混合等。选择、遗传漂变和群体混合会增加 LD 程度,使关联分析的精确性降低;而突变和重组由于多态性位点的产生而破坏 LD 状态,因此突变和重组是影响 LD 的 2 个最主要的因素。标记类型也是 LD 的 1 个影响因素,Remington 等用 SSR 标记和 SNPs 对来自全世界的 102 份常用玉米育种自交系全基因组标记的 LD 大小进行了研究,结果表明 SSR 标记检测的 LD 水平明显偏高,这说明在反映群体演化中 SSR 标记更有说服力^[10]。另外,授粉方式也是影响 LD 的因素,一般而言异花授粉重组率高于自花授粉,其 LD 水平相较于自花授粉作物要低,所以异化授粉作物作关联分析的效果要好于自花授粉作物。

1.2 LD 度量

LD 统计的是 2 个不同位点在实际观测中同时出现的频率与随机分离时期望同时出现的频率之间的差异,即 D 。假设有 2 个连锁的位点 A 、 B ,则 A 、 a 和 B 、 b 是其 4 种等位基因,频率分别为 φ_A 、 φ_a 、 φ_B 、 φ_b ; AB 、 aB 、 Ab 、 ab 的频率分别为 φ_{AB} 、 φ_{aB} 、 φ_{Ab} 、 φ_{ab} 。那么 D 的计算公式为: $D_{ab} = \varphi_{AB} - \varphi_A \varphi_B$ 。 $D=0$ 时,2 个基因座位处于连锁平衡状态; $D=1$ 时,2 个基因座完全关联; $0 < D < 1$ 时,2 个基因座处于 LD 状态^[11]。如果只有 2 个等位基因位点,通常用 D' 、 γ^2 来估计 2 个座位之间的 LD 水平。 γ^2 、 D' 的计算公式如下:

$$\gamma^2 = D^2 / (\varphi_A \varphi_a \varphi_B \varphi_b); \quad (1)$$

$$D' = D^2 / \min(\varphi_A \varphi_b, \varphi_a \varphi_B) \quad (D < 0);$$

或:

$$D' = D^2 / \min(\varphi_A \varphi_B, \varphi_a \varphi_b) \quad (D > 0)。 \quad (2)$$

D' 、 γ^2 代表的意义是不同的,它们的取值范围都是 0~1。 D' 仅包括重组史,对于一些稀有的等位基因来说, D' 可能比较大,有较高的 LD 水平。在小样本研究中,由于这 4 种等位基因相互组合的可能性较小, D' 不能准确反映群体 LD 水平。 γ^2 包括重组史、突变史, γ^2 可以反映标记是否与 QTL 有关,

因此在关联分析中通常用 γ^2 来表示群体的 LD 水平^[12]。而在实际的统计过程中,发生的频率低于 0.05 或 0.1 的变异基本上都是可以忽略不计的。

由于不同作物群体的 LD 程度是不同的,而这恰恰是影响关联分析有效性的主要因素。对 LD 结构有效利用,变异位点与表型性状关联,作出更加精确的高分辨率图谱,在生物基因组的研究上意义重大。在关联分析研究中,LD 弱的群体找到与目的基因连锁的标记相对较为容易,易于实现精确作图;LD 强的群体,找到与目标基因紧密连锁的标记较为困难,不易于实现精确作图。

2 关联分析的研究方法

目前关联分析常用的研究方法有 2 种:全基因组扫描和候选基因途径。全基因组扫描是在标记水平上对一些突变位点进行扫描,这些突变位点可能造成表型变异;而候选基因途径则是通过统计分析将那些与目标性状相关的等位基因从种质资源中挖掘出来,是在基因序列水平上进行的。

2.1 全基因组扫描的关联分析

在全基因组范围内对分子标记进行扫描,将所得分子标记数据与表型性状进行关联分析。本方法需要的标记密度很高,如大量的简单序列重复 (SSR)、SNP 标记等,但是不需要对研究对象再作任何假设,直接对全基因组的 DNA 变异进行研究。但是全基因组关联分析花费巨大、耗时较多,目前完成起来困难重重。

对野生甜菜的研究是全基因组关联分析的最早应用^[13]。在试验中发现,甜菜抽薹基因 B 与 440 个扩增片段长度多态性 (AFLP) 标记中的 2 个显著相关,并且存在 1 个与 B 基因紧密连锁的标记。这一结果表明,关联分析可以用来发掘与目的基因连锁的分子标记。此后越来越多的人通过关联分析的手段在全基因组水平上对不同作物的不同性状进行分析,以期发掘出与性状紧密连锁的标记位点。Huang 等对水稻进行全基因组关联分析,在 950 个品种中分别检测出 32、10 个与开花期、产量性状显著相关的位点^[14]。Weng 等用 41 101 个 SNP 对 284 个玉米自交系进行全基因组关联分析,鉴定出 105 个与株高相关的遗传位点,并借助关联分析进行精心作图,实现了抗玉米丝黑穗病主效位点精细定位^[15]。张焕欣等同样用 41 101 个 SNP 标记对 203 份玉米自交系进行全基因组关联分析,发现 9 个与穗行数显著关联的 SNP^[16]。连锁分析、关联分析都可以进行 QTL 定位,而连锁分析检测到的 QTL 数目一般少于关联分析;2 种方法检测到的 QTL 在位置上具有相当一部分具有一致性。这也表明,进行全基因组的关联分析是对植物数量性状研究的一条有效途径。由于全基因组关联分析可以同时很多个基因位点进行检测,并且不需要对特定的基因位点进行假设,因此全基因组关联分析可以大大减少所需成本,加快植物分子育种的研究进程。

2.2 候选基因的关联分析

候选基因的关联分析对性状的表型鉴定数据和候选基因的多态性进行关联分析,将对目标性状有贡献的等位基因从基因水平上发掘出来,一般涉及候选基因的功能预测。这需要对目标基因有一定的了解,目前应用比较多,特别是在多基因控制的性状 (如抗病性、抗逆性) 研究中具有十分重要的

作用。

候选基因关联分析在植物上的应用起步较晚。2001 年, Thornsberry 等第 1 次将关联分析应用到对玉米基因的研究中^[17], 这意味着关联分析可用于验证基因功能和发掘新基因, 为植物数量性状研究提供了新的思路。基于候选基因关联分析在玉米上的研究或许能作为典型代表。Tian 等为鉴定影响玉米花期性状的遗传位点, 选用含 5 000 个重组自交系 (RIL) 的玉米巢式作图群体作关联分析, 成功鉴定出其候选基因^[18]。Salvi 等对参与玉米开花期的基因 *Vgt1* 也进行了基于候选基因策略的关联分析^[19-20]。在研究与淀粉代谢有关的关键酶基因及其核苷酸 LD 的基础上, Wilson 等选择 6 个基因的不同区域进行关联分析, 发现有 4 个与玉米籽粒各成分和淀粉糊化特性的一些指标呈显著相关^[21]。于永涛运用候选基因关联分析法, 选出了 94 份能代表我国玉米核心种质的自交系, 采用关联分析方法鉴定候选基因 *rab17* 并成功检测到 6 个与表型性状相关的位点^[22]。Ravel 等以控制高分子谷蛋白亚基的 2 个 QTL 位点为候选基因, 对 113 份小麦进行关联分析, 发现其中 1 个 QTL 位点与所测性状存在显著关联^[23]。Zheng 等使用 7 个与小麦品质相关的候选基因对 96 份小麦栽培种和高代品系进行关联分析发现, 个体亲缘关系分析可以起到改进小麦品质性状关联分析的效果^[24]。另外, 基于候选基因的关联分析在大麦、高粱、棉花、番茄等作物中也多见报道。

对于全基因组序列已经获得的材料, 可以先通过 QTL 把候选基因在基因序列上的位置缩小, 然后利用生物信息学手段排除多的冗余序列, 再对候选基因进行关联分析, 这样就可以比较快速地找到目标性状的候选基因。

3 基于分子标记的关联分析及在植物研究中的应用

全基因组关联分析需要大量的分子标记, 随着许多物种测序工作的逐步完成, 各种标记类型的大量开发, 关联分析在鉴定植物数量性状上将表现出更大的应用潜力。目前, 已有很多分子标记在植物性状关联分析中得到成功应用。

3.1 AFLP 标记与关联分析

在植物研究中, 首次运用全基因组扫描方法进行关联分析的是 2001 年 Hansen 等对野生甜菜生长习性的研究^[13]。2004 年, Kraakman 等对春大麦品种产量特征作关联分析时, 发现所用的 236 个 AFLP 标记中有 8 个与产量相关联, 有 5 个与产量稳定性相关联^[25]。在 2006 年, Kraakman 等又利用 148 份春大麦栽培种的部分性状与 AFLP 标记进行 LD 作图, 找到与性状对应的标记位点, 证明 LD 作图是寻找标记位点的另一种重要途径^[26]。王蕾等以 215 份芝麻核心种质为材料进行芝麻素、芝麻酚林与 SSR、AFLP、SRAP 标记的关联分析, 发现 33 个标记位点与供试群体的芝麻素、芝麻酚林极显著关联, 与芝麻酚林极显著关联的 8 个位点中有 1 个 AFLP 标记^[27]。杨鑫雷等以陆地棉和海岛棉为材料, 设计 20 对 AFLP 引物, 共发现有 125 个位点, 用这些位点与不同的农艺性状进行关联, 发现有 15 个位点与不同的性状有着显著相关性, 其中 6 个 AFLP 位点可以用于分子标记辅助育种^[28]。AFLP 标记用到的选择引物较少, 并且可以在短时间内检测大量的位点, 作为一种高密度标记系统, 结合关联分析应用

时, 可以作为全基因组作图和基因定位的一种补充手段, 依然具有很大的应用价值。

3.2 限制性片段长度多态性 (RFLP) 标记与关联分析

由于 RFLP 分子标记有限, 因此其应用也是有限的, 只能作为一种初级工具使用。Beer 等利用 64 个燕麦地方品种分析 13 个 QTL 区域的 RFLP 位点与表型的相关性, 在未考虑群体结构的前提下检测到了较多的显著关联^[29]。RFLP 虽然应用有限, 但是其结果稳定、重复性好, 在关联分析中应用时, 仍然可以作为分子标记辅助育种的手段。而且在鉴定外源染色体、确定易位系的易位断点、定位外源基因和水稻籼粳分类及度量籼粳分化程度等方面也有很好的应用。

3.3 SSR 标记与关联分析

SSR 标记在关联分析中的应用较多, 在不同的作物中都有大量应用, 并取得了丰硕的研究成果。2006 年, Flavio 等为了验证已开发的 18 个 SSR 标记和籽粒性状之间的关系, 采用不同的小麦品种进行关联分析, 发现其中有 3 个标记与籽粒宽度存在显著关联^[30]。Pariaseaux 等利用全基因组扫描, 对与玉米不同性状相关的 QTL 进行图谱定位查询, 将 1 266 个玉米品种自交系分为 A~I 9 个组, 并对不同的组进行 7 年多地点试验, 用 SSR 标记对相关性状进行关联分析, 发现与株高、黑穗病抗性、籽粒含水量相关联的 SSR 标记分别有 37、24、44 个^[31]。Malysheva-Otto 等对 953 个大麦栽培种进行研究, 用 48 个 SSR 标记分析了它们在染色体上的 LD 水平, 发现 LD 衰减的距离在 1~150cM, 这表明其标记位点之间连锁不紧密, LD 水平较高^[32]。刘新伦等对小麦 99 个 SSR 标记作关联分析发现, 有 16 个标记与麦长管蚜抗性相关联^[33]。孙晓棠等对 456 份水稻材料的 144 个 SSR 标记与纹枯病抗性进行关联分析表明, 有 13 个标记位点与纹枯病抗性显著关联, 并发现 3 个新的抗病相关位点^[34]。范虎等利用 SSR 标记对大豆不同的农艺性状进行关联分析, 共获得 51 个与籽粒和花期等性状相关联的位点, 其中与 QTL 定位一致的有 16 个^[35]。左巧美等研究发现, 有 40 个 SSR 位点与大豆生育期性状关联, 与光温反应值相关的位点有 5 个^[36]。同样是对大豆相关性状的关联分析, 王欢等研究发现, 有 33 个 SSR 标记位点与大豆花荚脱落性显著相关^[37]。陈氏秋江等对不同地区的 540 份水稻品种用 262 个 SSR 标记进行基因组变异扫描检测到与粒长、粒宽、粒厚显著相关的标记位点分别有 45、7、11 个^[38]。贺道华等利用 132 个 SSR 标记, 对 92 份栽培棉花进行全基因组扫描, 获得了 21 个与棉花纤维品质相关的 SSR 位点^[39]。邵冰欣等利用 SSR 分子标记对 134 份陆地棉栽培种进行检测, 共检测出 148 个多态性位点, 涉及 246 个等位变异, 关联分析共发现与棉花耐盐性状相关的 SSR 分子标记位点有 8 个^[40]。这些研究表明, 关联分析在探究标记与相关性状的关联性、发掘功能位点的研究中具有重要作用。

3.4 SNP 标记与关联分析

单核苷酸多态性是由单碱基的转化、颠换, 以及单碱基的插入和缺失等改变引起的^[41], 作为第 3 代标记类型, 它比 RFLP、SSR 标记等更有优势, 是目前所有标记类型中精确度最高的标记类型。随着大量 SNP 标记的开发, 基于 SNP 标记的关联分析成为研究植物数量性状的重要手段。

Christian 等利用 56 110 个 SNP 标记对 289 份玉米自交

系的表型和生理性状进行全基因组关联分析,最后找到 15 个 SNP 标记与生理性状显著关联^[42]。Aranzana 等为了研究拟南芥的花期和抗病性与 SNP 位点的相关性,利用覆盖全基因组的 2 553 个 SNP 标记作关联分析,结果表明有 4 个基因与目标性状存在关联^[43]。姚玉莹对水稻作全基因组关联分析,发现在检测到的 48 个 SNPs 中有 6 个与水稻种质资源农艺性状显著相关^[44]。Beló 等鉴定出了与油酸含量相关的主效位点^[45],采用全基因组扫描的方法,对 553 份优良自交系的 8 950 个 SNP 位点与油酸性状进行关联分析表明,1 个油酸去饱和酶基因 *fad2* 与油酸含量相关联。任凤阳用玉米 10 条染色体上的 35 171 个 SNP 标记对 172 份玉米自交系在干旱胁迫条件下的相对发芽率、相对活力指数以及综合指标进行关联分析,发现与 3 个指标相关联的标记有 114 个,其中 20 个 SNP 标记与相对发芽率关联,53 个 SNP 标记与相对活力指数关联,41 个 SNP 标记与综合指标关联^[46]。Hao 等对 191 个大豆地方品种的 209 个单倍型和 1 536 个 SNP 位点分析,得到与产量性状显著相关的 19 个 SNP 标记位点^[47]。姜晓东等通过对 58 份大麦品种中 *Amy6-4* 基因核苷酸多态性与 α -淀粉酶活性的关联分析发现,*Amy6-4* 基因的 7 个 SNP 位点及其构成的单倍型均与酶活性无关联性^[48]。

在植物遗传研究中,各种类型的分子标记不断被发现,特别是 SNP 高密度图谱构建,对目标 QTL 的定位越来越精细。关联分析在 QTL 定位中的应用也日趋成熟,已经由最初应用于候选基因的功能验证发展为全基因组关联分析,成为挖掘作物数量性状新基因的强力工具。如在玉米的研究中,美国农业部建立了代表全球玉米种质多样性的巢式关联图谱群体,并通过全基因组关联分析鉴定了 1 批与农艺性状相关的 QTL。目前,基于分子标记的关联分析已经得到了十分广泛的应用,通过关联分析可以找到与性状相关的位点、实现标记的精确定位,也能够对候选基因进行多态性分析、验证候选基因功能。利用连锁分析作图,一般其 QTL 位点目的基因之间的图距都比较大,而关联分析鉴定标记则可以达到单基因水平,使精度得到很大的提升,在分子标记辅助育种(molecular assisted selection, MAS)中可极大地提高选择的目的性和准确性,进而提高育种效率。分子标记在关联分析中的应用价值巨大,可以明确目的基因与目标性状之间的内在关系、鉴定评价多个基因等。通过关联分析对大量等位基因分析,不但有助于了解不同基因变异对基因功能的影响,同时还可以通过基因变异发现最适合的等位基因,为深入了解植物数量性状、作物数量性状的遗传改良提供新的手段。

另外,关联分析在功能基因的验证上也得到了很好的应用。由于植物的很多性状都属于数量性状,由多个途径共同控制,很难通过转化的方法予以验证。例如著名的“黄金水稻”,通过把控制类胡萝卜素合成的 4 个基因同时转入水稻,才获得了能合成类胡萝卜素的黄金水稻。在对目的基因的代谢网络不清楚的情况下,可以用关联分析技术来验证,例如 Palaisa 对维生素 A 的 2 个限速酶基因 *YI*、*PSY2* 在黄色玉米(含类胡萝卜素)、白色玉米(不含类胡萝卜素)中进行分析发现,*YI* 基因是控制类胡萝卜素的关键基因,*PSY2* 则可能是假基因^[49]。

4 存在的问题与展望

对植物基因组的研究目前正由简单性状到复杂性状过渡,传统的 QTL 已经逐渐失去优势,建立在“自然群体”上的关联分析则成为研究的热点。但是,关联分析仍然存在一些问题,如关联分析数据量太大,而统计处理的方法还有待提高;群体结构的影响;另外,对于遗传多样性不高的群体,关联分析作图并不理想。然而,随着各种生物学手段和生物技术得到飞速发展、基因型分析技术和高通量测序技术的不断提高,关联分析必将在植物遗传学研究中发挥更为重要的作用。与此同时,关联分析在植物种质资源的遗传改良和种质创新领域也有重大应用价值,关联分析的应用大大加快了种质资源的研究进程(如核心种质资源库的构建)。关联分析技术虽然已有多项成功运用的报道,但是在植物上的应用还是相对较少,因此在植物遗传领域具有巨大的应用价值和发展潜力。

参考文献:

- [1] Flint - Garcia S A, Thornberry J M. Structure of linkage disequilibrium in plants[J]. Annual Review of Plant Biology, 2003, 54 (1): 357 - 374.
- [2] Gibbs R A, Belmont J W, Hardenbol P, et al. The international Hap-Map project[J]. Nature, 2003, 426 (6968): 789 - 796.
- [3] Flint - Garcia S A, Anne - Céline T, Yu J, et al. Maize association population: a high - resolution platform for quantitative trait locus dissection[J]. The Plant Journal: for Cell and Molecular Biology, 2005, 44 (6): 1054 - 1064.
- [4] Flint - Garcia S A, Thornberry J M. Structure of linkage disequilibrium in plants[J]. Annual Review of Plant Biology, 2003, 54 (4): 357 - 374.
- [5] Yu J, Buckler E S. Genetic association mapping and genome organization of maize[J]. Current Opinion in Biotechnology, 2006, 17 (2): 155 - 160.
- [6] Gebhardt C, Ballvora A, Walkemeier B, et al. Assessing genetic potential in germplasm collections of crop plants by marker - trait association: a case study for potatoes with quantitative variation of resistance to late blight and maturity type[J]. Molecular Breeding, 2004, 13 (1): 93 - 102.
- [7] Brescaghello F, Sorrells M E. Association analysis as a strategy for improvement of quantitative traits in plants[J]. Crop Science, 2006, 46 (3): 1323 - 1330.
- [8] Gupta P K, Rustgi S, Kulwal P L. Linkage disequilibrium and association studies in higher plants: Present status and future prospects [J]. Plant Molecular Biology, 2005, 57 (4): 461 - 485.
- [9] Jennings H S. The numerical results of diverse systems of breeding, with respect to two pairs of characters, linked or independent, with special relation to the effects of linkage [J]. Genetics, 1917, 2 (2): 97.
- [10] Remington D L, Thornsberry J M, Matsuoka Y, et al. Structure of linkage disequilibrium and phenotypic associations in the maize genome[J]. Proceedings of the National Academy of Sciences of the United States of America, 2001, 98 (20): 11479 - 11484.
- [11] Devlin B, Risch N. A comparison of linkage disequilibrium measures for fine - scale mapping[J]. Genomics, 1995, 29 (2): 311 - 322.

- [12] Balding D J. A tutorial on statistical methods for population association studies[J]. *Nature Reviews Genetics*, 2006, 7(10): 781–791.
- [13] Hansen M, Kraft T, Ganestam S, et al. Linkage disequilibrium mapping of the bolting gene in sea beet using AFLP markers[J]. *Genetical Research*, 2001, 77(1): 61–66.
- [14] Huang X, Zhao Y, Wei X, et al. Genome-wide association study of flowering time and grain yield traits in a worldwide collection of rice germplasm[J]. *Nature Genetics*, 2012, 44(1): 32–39.
- [15] Weng J F, Xie C X, Hao Z F, et al. Genome-wide association study identifies candidate genes that affect plant height in Chinese elite maize (*Zea mays* L.) inbred lines[J]. *PLoS One*, 2011, 6(12): e29229.
- [16] 张焕欣, 翁建峰, 张聪晓, 等. 玉米穗行数全基因组关联分析[J]. *作物学报*, 2014, 40(1): 1–6.
- [17] Thornsberry J M, Goodman M M, Doebley J, et al. Dwarf8 polymorphisms associate with variation in flowering time[J]. *Nature Genetics*, 2001, 28(3): 286–289.
- [18] Tian F, Badbury P J, Brown P J, et al. Genome-wide association study of leaf architecture in the maize nested association mapping population[J]. *Nature Genetics*, 2011, 43(2): 159–162.
- [19] Salvi S, Sponza G, Morgante M, et al. Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2007, 104(27): 11376–11381.
- [20] Ducrocq S, DVEYRIERAS J B M. Key impact of *Vgt1* on flowering time adaptation in maize: evidence from association mapping and ecogeographical information[J]. *Genetics*, 2008, 178(4): 2433–2437.
- [21] Wilson L M, Whitt S R, Ibañez A M, et al. Dissection of maize kernel composition and starch production by candidate association[J]. *The Plant Cell*, 2004, 16(10): 2719–2733.
- [22] 于永涛. 玉米核心自交系群体结构及耐旱相关候选基因 *rab17* 的等位基因多样性分析[D]. 北京: 中国农业科学院, 2006.
- [23] Ravel C, Praud S, Murigneux A, et al. Identification of *Glu-B1-1* as a candidate gene for the quantity of high-molecular-weight glutenin in bread wheat (*Triticum aestivum* L.) by means of an association study[J]. *Theoretical & Applied Genetics*, 2006, 112(4): 738–743.
- [24] Zheng S S, Byrne P F, Bai G H, et al. Association analysis reveals effects of wheat glutenin alleles and rye translocations on dough-mixing properties[J]. *Journal of Cereal Science*, 2009, 50(2): 283–290.
- [25] Kraakman A T W, Niks R E, van den Berg P M M M, et al. Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars[J]. *Genetics*, 2004, 168(1): 435–446.
- [26] Kraakman A T W, Martinez F, Mussiraliyev B, et al. Linkage disequilibrium mapping of morphological, resistance, and other agronomically relevant traits in modern spring barley cultivars[J]. *Molecular Breeding*, 2006, 17(1): 41–58.
- [27] 王蕾, 黎冬华, 齐小琼, 等. 芝麻核心种质芝麻素和芝麻酚林的关联分析[J]. *中国油料作物学报*, 2014, 36(1): 32–37.
- [28] 杨鑫雷, 周晓栋, 刘恒蔚, 等. AFLP 标记与棉花重要农艺性状的关联研究[J]. *棉花学报*, 2013, 25(3): 211–216.
- [29] Beer S C, Siripoonwiwat W, O'Donoghue L S, et al. Associations between molecular markers and quantitative traits in an oat germplasm pool: can we infer linkages? [J]. *Journal of Agricultural Genomics*, 1997, 3: 1–16.
- [30] Flavio B, Sorrells M E. Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars[J]. *Genetics*, 2006, 172(2): 1165–1177.
- [31] Parisseaux B, Bernardo R. In silico mapping of quantitative trait loci in maize[J]. *Theoretical and Applied Genetics*, 2004, 109(3): 508–514.
- [32] Malysheva-Otto L V, Canal M W, Röder M S. Analysis of molecular diversity, population structure and linkage disequilibrium in a worldwide survey of cultivated barley germplasm (*Hordeum vulgare* L.) [J]. *BMC Genetics*, 2006, 7(2): 1–14.
- [33] 刘新伦, 李志超, 王亚娟, 等. 抗麦长管蚜小麦的遗传多样性及 SSR 标记与麦长管蚜抗性的关联分析[J]. *农业生物技术学报*, 2015(3): 291–301.
- [34] 孙晓棠, 卢冬冬, 欧阳林娟, 等. 水稻纹枯病抗性关联分析及抗性等位变异发掘[J]. *作物学报*, 2014, 40(5): 779–787.
- [35] 范虎, 文自翔, 王春娥, 等. 中国野生大豆群体农艺加工性状与 SSR 关联分析和特异材料的遗传构成[J]. *作物学报*, 2013, 39(5): 775–788.
- [36] 左巧美. 大豆生育期的遗传变异、QTL 定位和关联分析[D]. 南京: 南京农业大学, 2011.
- [37] 王欢, 孙霞, 岳岩磊, 等. 东北春大豆花荚脱落性状与 SSR 标记的关联分析[J]. *土壤与作物*, 2014, 3(1): 32–40.
- [38] 陈氏秋江, 党小景, 刘强明, 等. 水稻籽粒性状的 SSR 关联分析[J]. *中国水稻科学*, 2014, 28(3): 243–257.
- [39] 贺道华, 邢宏宣, 赵俊兴, 等. 棉花资源群体结构的推测与纤维品质的关联分析[J]. *西北农林科技大学学报: 自然科学版*, 2011, 39(1): 103–112.
- [40] 邵冰欣, 王红梅, 赵云雷, 等. 陆地棉耐盐性状与 SSR 分子标记的关联分析[C]. 呼和浩特: 中国棉花学会 2014 年年会, 2014.
- [41] Brookes A J. The essence of SNPs[J]. *Gene*, 1999, 234(2): 177–186.
- [42] Christian R, Jan L, Angelika C E, et al. Genome-wide association mapping of leaf metabolic profiles for dissecting complex traits in maize[J]. *Proceedings of the National Academy of Sciences of the United States of America*, 2012, 109(23): 8872–8877.
- [43] Aranzana M J, Kim S, Zhao K Y, et al. Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes[J]. *PLoS Genetics*, 2005, 1(5): e60.
- [44] 姚玉莹. 水稻核心种质抗倒特性及其主要农艺性状的全基因组关联分析[D]. 武汉: 华中农业大学, 2014.
- [45] Beló A, Zheng P Z, Luck S, et al. Whole genome scan detects an allelic variant of *fad2* associated with increased oleic acid levels in maize[J]. *Molecular Genetics & Genomics*, 2008, 279: 1–10.
- [46] 任凤阳. 玉米自交系萌发特性的评价及 SNP 关联分析[D]. 泰安: 山东农业大学, 2014.
- [47] Hao D, Cheng H, Yin Z T, et al. Identification of single nucleotide polymorphisms and haplotypes associated with yield and yield components in soybean (*Glycine max*) landraces across multiple environments[J]. *Theoretical and Applied Genetics*, 2012, 124(3): 447–458.
- [48] 姜晓东, 郭刚刚, 张京. *Amy6-4* 基因遗传多样性及其与 α -淀粉酶活性的关联分析[J]. *作物学报*, 2014, 40(2): 205–213.
- [49] Palaisa K A, Morgante M, Williams M, et al. Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci[J]. *Plant Cell*, 2003, 15(8): 1795–1806.