

· 专题论坛 ·

关联分析及其在植物遗传学研究中的应用

谭贤杰^{1,2}, 吴子恺¹, 程伟东², 王天宇³, 黎裕^{3*}

¹广西大学农学院, 南宁 530005; ²广西农业科学院玉米研究所, 南宁 530227

³中国农业科学院作物科学研究所, 北京 100081

摘要 植物的很多重要经济性状均属于复杂性状。基于连锁分析的QTL作图是研究复杂性状的有效手段, 但其尚存在一定的局限性。随着现代生物学的发展, 一种基于连锁不平衡的新剖分复杂性状方法——关联分析法, 开始应用于植物遗传学研究。与QTL作图法相比, 应用关联分析法具有不需要构建特殊的群体, 可同时对多个等位基因进行分析, 定位QTL精度可达到单基因水平等优势。该文介绍了关联分析方法学的基础和特性, 简述了其在植物遗传学研究中的进展情况, 并对其未来发展和在植物遗传学研究中的应用进行了展望。

关键词 关联分析, 单倍型, 连锁不平衡, QTL

谭贤杰, 吴子恺, 程伟东, 王天宇, 黎裕 (2011). 关联分析及其在植物遗传学研究中的应用. 植物学报 46, 108–118.

自20世纪80年代以后, 随着RFLP(restriction fragment length polymorphism)和SSR(simple sequence repeat)等分子标记技术在植物遗传学研究中的广泛应用, 利用数量性状位点(quantitative trait locus, QTL)定位目标基因并对其进行图位克隆成为植物遗传学研究的一项重大突破。从QTL定位到基因克隆一般包括初步定位、精细定位、染色体步移和候选基因功能鉴定等步骤。利用QTL-图位克隆法, 现已克隆并阐释了一些重要基因, 如番茄(*Lycopersicon esculentum*)的 *fw2.2*(Frary et al., 2000)和 *Lin5*(Briggs et al., 2007)、玉米(*Zea mays*)的 *tga1*(Wang et al., 2005)和 *Vgt1*(Salvi et al., 2007)以及水稻(*Oryza sativa*)的 *hd1*(Yano et al., 2000)和 *GN1a*(Ashikari et al., 2005)等。但是, 在应用QTL-图位克隆研究中发现该方法存在一定的局限性: (1) 杂交不亲和和物种或难于进行杂交操作的物种很难获得所需的杂交组合或后代; (2) QTL研究群体只能是少数的性状和等位基因(一般为2个); (3) 构建相关群体常常需要多年时间, 构建达到基因克隆水平的次级群体往往需耗时5–10年(Doerge, 2002; Holland, 2007); (4) 连锁分析的QTL定位精确度低。初级群体对QTL定位一般可达到10–30 cM水平, 次级群体可达近1 cM,

但其区段所包含的核苷酸碱基往往也将近百万对(莫惠栋和顾世梁, 2000)。因此, 通过连锁分析克隆基因不仅操作繁琐, 而且耗时耗力。鉴于此, 科学家们开始探讨新的研究方法以克服这些障碍, 其中一个新方法就是基于连锁不平衡的关联分析方法。关联分析, 亦称关联作图, 其分析不需要构建特殊的群体且可同时对多个性状进行分析, 对QTL定位的精度可达到单基因水平。关联分析除了能够定位QTL, 还可用于鉴定不同等位基因引起的表型变异和开发功能标记(Meuwissen and Goddard, 2000; Palaisa et al., 2003)。鉴于关联分析本身存在的优势, 目前关联分析已广泛应用于多种植物的研究, 如玉米的花期(Thornsberry et al., 2001)、籽粒的淀粉含量(Wilson et al., 2004)和维生素A原(Harjes et al., 2008); 小麦(*Triticum aestivum*)的籽粒大小和研磨品质(Bresegghello and Sorrells, 2006); 拟南芥(*Arabidopsis thaliana*)的开花期(Olsen et al., 2004)和抗病性(Aranzana et al., 2005)等。目前关联分析已成为植物遗传学研究的热点。

本文将介绍关联分析方法学的基础及特性, 简述其在植物遗传学研究中的应用进展, 探讨关联分析法的未来发展和在植物研究中的应用前景。

收稿日期: 2010-05-17; 接受日期: 2010-10-13

基金项目: 973 计划(No.2011CB100100, No.2009CB118401)和 863 计划(No.2006AA10Z188)

* 通讯作者。E-mail: yuli@mail.caas.net.cn

1 关联分析研究方法学

1.1 关联分析的基础——连锁不平衡

连锁不平衡(linkage disequilibrium, LD)亦称为配子相不平衡(gametic phase disequilibrium)、配子不平衡(gametic disequilibrium)或等位基因关联(allelic association),是指群体内不同座位等位基因(可以是标记亦可是基因/QTL间与标记)间的非随机关联(Gaut and Long, 2003)。同一染色体或不同染色体的基因座之间均可呈现连锁不平衡。群体内存在的LD均是由突变产生的等位基因出现后座位间所有重组事件累积的结果。位点间连锁越紧密,其LD水平越高。 D 、 r^2 、 D' 、 D^* 、 F 、 Q^* 和 δ 等参数曾用于度量连锁不平衡,目前常用的参数是 r^2 (squared allele-frequency correlations)和 D' (standardized disequilibrium coefficients)(Delvin and Risch, 1995; Jorde, 2000; Flint-Garcia et al., 2003)。LD在染色体上的分布一般用LD衰减散点图和LD配对检测的矩阵图来描述。前者可以观测LD随遗传或物理距离的衰减速率,后者可以直接观测同一染色体的基因座或基因的多态性位点之间LD的线性排列(Flint-Garcia et al., 2003; Gaut and Long, 2003)。

突变和重组是影响LD最重要的因素。突变是LD形成的原因,新突变的产生可打破原有LD,形成新的LD。多态位点间的重组也可打破LD,无连锁和自由交配的重组使位点间等位基因处于连锁平衡状态。群体中的LD是突变、重组和其它因素影响累积的结果。此外,物种交配方式、染色体位置、群体大小、自然与人工选择、遗传漂变和基因转换等也是影响LD的因素(Gupta et al., 2005; Oraguzie et al., 2007)。

1.2 连锁不平衡与关联分析

在群体中,个体等位基因差异是表型差异的根本原因。连锁分析是利用标记位点与引起表型差异的位点(QTL)之间的重组来定位QTL。关联分析则是利用引起表型差异的位点与标记之间的连锁不平衡来定位QTL:在自然群体的基因组中存在数目庞大的多态性,由于连锁的存在及群体形成过程中突变、重组和选择等因素的影响,多态位点的等位基因间存在广泛的非随机关联,亦即连锁不平衡状态。多个基因座的等位基因间的LD形成了一系列的单倍型(haplotype),

单倍型的大小取决于LD的衰减水平。LD的衰减水平越高,则形成的单倍型越小。根据单倍型可把群体内个体区分为不同类型或亚群。由于存在引起表型变异的等位基因,使得不同的单倍型群体具有表型上的差异,分析不同单倍型群体与表型变异的协变性(关联),就可把引起表型变异的位点定位到相对应的单倍型上。因此,分析标记与引起表型变异位点(QTL)的关联性,根据分子标记的信息即可定位QTL在染色体上的位置。如果所分析的分子标记恰为引起表型变异的位点,这种关联称之为直接关联;如果通过标记与QTL形成单倍型定位QTL,则称之为间接关联。间接关联定位QTL的精度与物种中的LD衰减大小密切相关,LD衰减速度慢则定位粗略,LD衰减速度快则定位精细。与常规QTL定位相比,在玉米中利用关联分析定位精度可提高5 000倍(Remington et al., 2001)。

1.3 关联分析的特性

与基于连锁分析的QTL相比,关联分析具有以下优势。(1) 关联分析利用的是自然群体,构建群体不需要控制材料的交配方式。构建常规QTL作图群体时需要控制实验群体的交配方式,通常需要2年或更长,特别是构建精细定位的次级群体可能会耗时数年。(2) 关联分析所用群体有更为广泛的遗传基础,可同时对同一基因座的多个等位基因进行分析,而绝大部分常规QTL作图所用群体通常为两亲本杂交重组后代,其基因座一般只涉及2个等位基因。(3) 关联分析作图定位更为精确,可以达到单基因水平。关联分析利用的是自然群体在长期进化过程中所累积的重组信息,因此具有更高的分辨率,可实现对QTL的精细定位,甚至可直接定位到基因本身;常规QTL作图则受重组发生率的影响,一般分辨率较低,通常初级群体能够将基因定位到10–30 cM的基因组区内,次级群体可将基因定位到1 cM区段内(Doerge, 2002; Holland, 2007)。

1.4 关联分析的研究策略

关联分析是利用标记与QTL等位基因间的LD来定位QTL,当选取的标记数量多到足以覆盖全基因组片段时,即可定位到所有影响表型的QTL,此种定位QTL的策略称为基于全基因组扫描的关联分析。全基因组扫描方法所需标记的数目取决于物种的基因组大小

和LD水平。物种基因组大小相同时, LD衰减速度慢的物种所需标记少, 但由于标记与目标基因在物理距离较远的情况下亦可出现高的LD, 故其定位精度比衰减速度快的物种低。鉴于物种的基因组碱基序列通常数以千万计甚至更多, 全基因组扫描所需检测标记数量极为庞大。据估计, 若保证对绝大部分重要的基因均实现作图, 人类约需要检测70 000个标记, 玉米地方品种群体则需750 000个, 优良玉米自交系群体的LD衰减速度慢, 约需50 000个, 基因组较小且LD衰减速度较慢的拟南芥约需2 000个标记(Flint-Garcia et al., 2003)。因此, 目前全基因组扫描方法仅应用于基因组信息丰度较高且标记易于获得的物种。LD较高的物种或群体, 应用较少的标记即可实现全基因组扫描。自花授粉的物种, 经历瓶颈效应和强烈人工选择的群体仅包含所有群体中少部分的等位基因, 故可用于用全基因组扫描法进行分析(Hastbacka et al., 1992; Rafalski, 2002; Rostoks et al., 2006)。在植物研究中, 亦可采取此法对F₂代分离群体进行全基因组扫描。由于F₂代分离群体亲缘关系极高且LD水平很高, 因此, 应用少量标记即可实现对群体的全基因组扫描。另外, 鉴于每个位点只有2个等位基因, 统计分析等位基因的效应和等位基因之间的上位性比采用自然群体功效更高(Flint-Garcia et al., 2003)。

有些基因对表型有决定性的影响, 这种基因则是主效基因或质量性状基因。有时主效基因单个碱基的差异亦可决定表型。因此, 对可能影响表型性状的基因组部分区段进行关联分析, 不需要过多的基因型分析工作即可定位目的基因, 这种策略称为基于候选基因的关联分析。应用全基因组扫描方式研究LD衰减速度快的物种时, 标记与QTL处于LD状态的概率较低, 定位到目标基因的几率很小, 因此, 采用候选基因法对这类物种进行研究更为有效。此外, 利用候选基因关联分析法可鉴定到位于该区段中影响表型的多态性, 并可估计其效应, 因此应用候选基因关联分析可对特定基因的等位变异是否控制目标性状进行验证, 进而挖掘出优异的等位基因(Flint-Garcia et al., 2003)。候选基因法所需标记数量较少且成本较低, 并可对目的基因进行功能鉴定, 因而在植物遗传学研究中较为常用。为了提高候选基因关联分析的目的性和效率, 选择候选基因(特别是关键生理生化途径中的重要功能基因、前期QTL研究定位区域所含的基因

和近缘物种研究中表明效应较大的同源基因)时, 往往需要利用基因组测序、比较基因组学、转录组学、QTL和反向遗传学研究所提供的信息。

1.5 关联分析中的假阳性及其消除

关联分析中, 群体中的LD将受到遗传漂变、群体分层和自然选择等诸多因素的影响, 因此在进行关联分析时, 一些非原因等位基因亦可与QTL形成LD, 从而表现出与性状关联, 此种现象称为伪关联或假阳性(Lander and Kruglyak, 1995)。在上述因素中, 群体分层(population stratification)被认为是引起假阳性的最主要因素(Cardon and Palmer, 2003)。群体分层是指群体内存在等位基因频率不同的亚群体, 这些亚群体的产生可能是因为有共同的祖先或者经受了相同的环境和人工选择等因素所致(Hey and Machado, 2003)。人类遗传学研究表明, 在人群中普遍存在分层现象(Rosenberg et al., 2002)。在对植物的研究中, Flint-Garcia等(2005)研究了玉米群体中分层对表型性状的影响, 结果表明, 群体分层解释了表型性状中约9.3%的变异。群体分层效应与目标性状基因效应协同干扰了关联分析, 形成假阳性, 为了减轻并消除关联分析中主要由群体分层引起的假阳性, 研究者们发展了一系列方法, 如传递不平衡法(transmission disequilibrium test, TDT) (Spielman et al., 1993)、基因组对照(genome control)法(Devlin and Roeder, 1999)、结构关联法(Pritchard and Rosenberg, 1999; Pritchard et al., 2000)、结构关联(Q)+亲缘关系(K)混合模型法(Yu et al., 2006)、主成分分析法(principal component analysis, PCA) (Price et al., 2006; Patterson et al., 2006)、多维标度法(multidimensional scaling, MDS) (Purcell et al., 2007; Li and Yu, 2008)和非计量多维标度法(nonmetric multidimensional scaling, NMDS) (Zhu and Yu, 2009)等。传递不平衡法是基于家系分析的研究方法, 其余方法均是基于群体的关联分析法。基于群体的关联分析法均利用随机且均匀分布于基因组的标记信息来估计群体内部个体间的遗传关系或作统计假设检验, 从而在关联分析时去除群体分层时引起的假阳性。为了检测这些方法在单独应用和联合应用时的有效性, Yu等(2006)和Zhao等(2007)对这些方法进行了模拟数据和实际数据的比较验证, 结果表明, Q+K混合模型和P(PCA)+

K(Kinship)混合模型均能够较好地捕获由群体分层引起的假阳性。随后, Stich和Melchinger(2009)研究认为Q+K混合模型的功效较高, 但是他们建议最好用KT代替K, KT为基于REML(restricted maximum likelihood)的估计个体间状态等同(而不是血缘等同估计下)的亲缘关系矩阵。除了上面提到的几种方法外, 对应分析(Epstein et al., 2007)和EMMA(efficient mixed-model association) (Kang et al., 2008)等方法也可以用来估计群体结构。这些方法的有效性尚有待进一步验证。随着研究的不断深入, 将会涌现出更为有效地减少假阳性、假阴性和更高功效的分析方法。

2 关联分析在植物遗传学研究中的应用

Thornsberry等(2001)首次将关联分析方法引入植物研究领域。迄今为止, 应用关联分析研究的植物已有10多种(表1)。在这些研究中, 大部分均利用了候选基因法, 选择的候选基因主要是生理生化途径中重要的功能基因、QTL研究定位区域所含的基因和近缘物种研究中表明效应较大的同源基因。其中以生理生化途径中重要功能基因为候选基因的研究中, 最典型的是Wilson等(2004)对玉米籽粒淀粉代谢和Harjes等(2008)对玉米维生素A代谢的研究。Wilson等(2004)分析了玉米淀粉代谢途径中的*sh1*、*sh2*、*bt2*、*wxl*、*ae1*和*sul* 6个关键酶基因与代谢产物的关联, 发现这6个基因中有4个基因与籽粒成分和淀粉糊化特性的一些指标存在显著相关。Harjes等(2008)对玉米维生素A代谢关键酶基因*LCYE*的研究表明, *LCYE*基因内存在4个影响 α -胡萝卜素与 β -胡萝卜素含量差异(达3倍以上)的多态性位点。Szalma等(2005)对玉米自交系群体中与可凝性球蛋白(*maysin*)和绿原酸(*chlorogenic acid*, *CGA*)积累有关的4个相关位点基因*p*、*a1*、*c2*和*whp1*进行了关联分析, 第1次在关联分析中阐明了上位性效应的重要作用。其分析表明, 作为主效QTL的*p*位点对*a1*、*c2*和*whp1*有上位性效应, 只有在*p*的功能性等位基因存在时, 才能检测到*c2*和*whp1*基因序列变异及*a1*启动子区域的2个序列多态性与*maysin*和*CGA*积累间的关联。Olsen等(2004)、Sköt等(2005)和Saïdou等(2009)对不同作物开花期的相关基因进行了关联分析, 其候选基因的选择均参考了同源基因的相关研究。这充分说明了前期QTL研究、

同源基因和生理生化基础信息对候选基因关联分析是不可或缺的。利用这些信息可以提高实验的目的性, 降低基因型分析的成本。Thornsberry等(2001)利用一个由92份玉米自交系构成的关联分析群体研究了*Dwarf8*基因多态性与开花期的关联, 结果表明, *Dwarf8*基因的9个基因多态性位点与玉米开花期的变化呈显著关联。然而, Andersen等(2005)使用另一个由71份欧洲优异玉米自交系构成的群体重新对*Dwarf8*基因的序列多样性与开花期和株高进行关联分析验证。结果发现, 在不考虑群体结构的情况下只检测出6个SNP与开花期相关, 如考虑群体结构则只有1个Indel与株高相关。针对*Dwarf8*基因, Camus-Kulandaivelu等(2006)选用了代表美国和欧洲遗传多样性的375份玉米自交系和275份地方品种, 对其开花期进行关联分析, 得出的结论与Thornsberry等的相同。这一系列对同一目标基因*Dwarf8*的研究说明, 群体分层是关联分析中不可忽略的因素, 特别是对经历了长期人工选择和改良过程的作物(可能存在复杂的亲缘关系)进行关联分析研究时, 需要充分考虑群体的分层效应。

目前, 鉴于标记数量和基因型分析技术的限制, 对大部分物种实现全基因组扫描还存在一定的困难。近年来仅有一些利用较少的标记开展关联分析的研究报道。例如, Hansen等(2001)分析了覆盖全基因组的440个AFLP标记与控制甜菜(*Beta vulgaris*)生长习性的*B*基因的关系, 发现有2个标记与*B*基因之间的连锁程度很高, 而在前期连锁分析中这2个标记中有1个与*B*基因存在紧密连锁, 说明通过关联分析完全可以寻找到与目标基因紧密连锁的分子标记。随后, Sköt等(2005)对多年生黑麦草(*Lolium perenne*)的抽穗期性状、Breseghello和Sorrells(2006)对小麦的籽粒大小和碾磨品质性状、Malosetti等(2007)对马铃薯(*Solanum tuberosum*)的晚疫病抗性以及Agrama等(2007)对水稻的产量与相关性状进行了类似的研究。结果均表明, 与性状关联的标记位点和前期QTL有很好的一致性。Aranzana等(2005)利用覆盖拟南芥全基因组的2 553个SNP标记进行了花期和抗病性的关联分析, 并对4个已知功能的基因进行了检测以验证全基因组扫描的有效性。结果表明, 有4个基因与目标性状存在关联, 但也发现, 应用GC法和SA法分析并未有效减少假阳性。Zhao等(2007)对此研究数据重新应

表1 关联分析在植物研究中的实例

Table 1 Examples of association mapping studies in various plant species

种名	群体	样本数量	背景标记数量	性状	参考文献
玉米 (<i>Zea mays</i>)	自交系	92	141	开花期	Thornsberry et al., 2001
	优良自交系	71	55	开花期	Andersen et al., 2005
	自交系和地方品种	375+275	55	开花期	Camus-Kulandaivelu et al., 2006
	自交系	95	192	开花期	Salvi et al., 2007
	自交系	102	47	籽粒成分, 淀粉黏性	Wilson et al., 2004
	自交系	86	141	可凝性球蛋白与绿原酸含量	Szalma et al., 2005
	优良自交系	75		籽粒颜色	Palaisa et al., 2003
	自交系	57		甜度	Tracy et al., 2006
	优良自交系	553	8 950	油酸含量	Belo et al., 2008
	自交系	282	553	胡萝卜素含量	Harjes et al., 2008
	自交系	282		铅毒抗性	Krill et al., 2010
大刍草 (<i>Zea diploperennis</i>)	野生材料	817	123	花期、植株、花器官和籽粒性状	Weber et al., 2008
甜菜 (<i>Beta vulgaris</i>)	自然种质	106	440	生长习性	Hansen et al., 2001
拟南芥 (<i>Arabidopsis thaliana</i>)	不同生态型材料	95	104	开花期	Olsen et al., 2004
	不同生态型材料	95	2 553	抗病性, 开花期	Aranzana et al., 2005; Zhao et al., 2007
	不同生态型材料	96		分枝数	Ehrenreich et al., 2007
	不同生态型材料	275		开花期	Ehrenreich et al., 2009
	MAGIC群体	527	1 260	生育期性状等	Kover et al., 2009
高粱 (<i>Sorghum vulgare</i>)	核心种质	95+96	250 000	生育期等107个性状	Atwell et al., 2010
	自交系	377	47	花期、株高等8个性状	Casa et al., 2008
小麦 (<i>Triticum aestivum</i>)	自交系	107	98	抽穗期等26个性状	Shehzad et al., 2009
	栽培品种	95	95	籽粒大小, 研磨品质	Breseghele and Sorrells, 2006
大麦 (<i>Hordeum vulgare</i>)	栽培品种	44	91	颖枯病抗性	Tommasini et al., 2007
	栽培品种	148	139	抽穗期, 锈叶病、黄矮病毒病抗性, 穗毛长和浆片大小	Kraakman et al., 2004
马铃薯 (<i>Solanum tuberosum</i>)	地方品种	429	129	开花期	Cockram et al., 2008
	栽培品种	600		晚疫病抗性	Gebhardt et al., 2004
	栽培品种	123	49	晚疫病抗性	Malosetti et al., 2007
水稻 (<i>Oryza sativa</i>)	地方品种	105		米饭黏性表型	Olsen and Purugganan, 2002
	地方品种	577	577	淀粉质量	Bao et al., 2006
	地方品种	103	123	产量和相关性状	Agrama et al., 2007
	地方品种	90	218	小穗花性状	Yan et al., 2009b
	地方品种	170	132	抽穗期, 株高, 穗长	Wen et al., 2009
	栽培品种	70		米饭品质	Tian et al., 2009
	地方品种	293	179	籽粒外形	Iwata et al., 2010
火炬松 (<i>Pinus taeda</i>)	自然群体	32	21	木质性状	Gonzalez-Martinez et al., 2006
	自交系	435	288	微纤丝角度, 纤维素含量	Gonzalez-Martinez et al., 2007
甘蔗 (<i>Saccharum officinarum</i>)	无性系	154	2 209	抗病性	Wei et al., 2006
桉树 (<i>Eucalyptus</i> spp.)	自然群体	290	35	微纤丝角度	Thumma et al., 2005
黑麦草 (<i>Lolium perenne</i>)	自然种质	26	589	抽穗期	Skøt et al., 2005
	自然种质	96	506	开花期, 水溶性糖	Skøt et al., 2007
大豆 (<i>Glycine max</i>)		48	150	籽粒蛋白质含量	Jun et al., 2008
珍珠粟 (<i>Pennisetum glaucum</i>)	自交系	90		开花期, 穗长, 茎粗	Saïdou et al., 2009

MAGIC: 多亲本高级世代互交系 MAGIC: Multiparent advanced generation inter-cross

用P+K和Q+K等模型进行了分析,结果表明,Q+K混合模型在维持统计功效的同时可有效降低假阳性,但伴随产生了一定的假阴性。尽管无法完全消除假阳性和假阴性,全基因组扫描关联分析仍然是一个鉴定QTL的有力工具。Belo等(2008)利用8 590个SNP标记全基因组扫描影响玉米籽粒的QTL,最终定位到*fad2*位点,并验证了*fad2*基因激活区的核苷酸替换可能会影响基因的表达从而影响表型。这是第1个利用全基因组扫描关联分析定位到目标基因的报道。此外,研究者们还对实验群体进行了探讨。如Kover等(2009)利用1 026个SNP标记对一个多亲本高级世代互交系(multiparent advanced generation intercross, MAGIC)群体(为19个拟南芥材料相互杂交构建的群体)进行了全基因组扫描,分析结果证实,利用此群体定位的QTL与已知的QTL极为接近,说明应用此类等位基因有限群体进行全基因组扫描功效更高且更加精确。最近,Atwell等(2010)使用包含250 000个SNP的基因芯片对拟南芥107个性状进行了全基因组扫描,鉴定到众多主效位点,但是目前还难于解释其中不少基因位点的效应,原因是多个位点的效应混杂在一起,受群体结构的影响很难分辨是否存在真正的关联。然而,一些前期研究发现的候选基因位点显示出了强烈的关联,这些基因位点可作为下一步研究的首选目标。Atwell等(2010)的研究是一个真正意义上的全基因组扫描实例,尽管还存在假阳性等问题,但却证实了关联分析可以有效鉴定QTL,显示了全基因组扫描在植物遗传学研究中的巨大潜力。

3 展望

自2001年关联分析引入植物遗传学研究领域以来,关联分析备受关注,其应用正处于快速发展时期。近年来,随着新统计分析方法的发展,关联分析在减少假阳性的同时提高了分析功效,使得关联分析方法学日趋完善。伴随基因型分析技术的发展,特别是高通量测序技术的发展,关联分析必将在植物遗传学研究中发挥更为重要的作用。

3.1 基因型分析技术的发展

进入21世纪后,测序技术和基因芯片技术得到迅速发展。在测序技术方面,以454-GSFLX、Illumina

Genome Analyzer、Solid System和SMRT为代表的高通量测序技术极大地提高了测序通量,同时降低了成本,这为大规模测序提供了极大的便利。目前已完成了多种植物的基因组测序。此外,数十种植物的基因组正在测序中(<http://www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantList.html>)。在基因芯片技术方面,Affymetrix和Illumina等公司已能够提供多个物种的数以百万标记的基因芯片,用于SNP分析和表达分析。高通量测序和基因芯片等生物技术的发展促进了基因组学、转录组学、比较基因组学、进化基因组学和反向遗传学等学科的飞速发展,使得对基因多态性、遗传定位、表达、调控和功能的研究更加深入全面,从而在候选基因策略研究中减少了实验风险,提高了实验的目的性。利用高通量测序技术对基因组测序、重测序和表达序列测序可为关联分析候选基因的选择和基于全基因组扫描的基因芯片开发提供信息。目前,高通量测序技术和基因芯片技术已在大麦(*Hordeum vulgare*)(Hamblin et al., 2010)、玉米(Yan et al., 2009a)、葡萄(*Vitis vinifera*)(Myles et al., 2010)和拟南芥(Atwell et al., 2010)中得到应用。伴随高通量测序技术和基因芯片技术的高速发展,全基因组扫描关联分析方法将会在植物遗传学研究中广泛应用。

3.2 关联分析的方法学发展

自从关联分析引入植物学领域以来,关联分析方法学得到了快速发展,但还有许多方面尚需完善。首先,鉴于群体分层会引起假阳性,研究者尚需探索对群体结构剖分更为精确的模型及特殊的关联作图群体,以减少假阳性并提高功效。如Yu等(2008)发展了巢式关联作图(nested association mapping, NAM)群体,该群体由25个代表性玉米自交系分别与B73杂交后形成的重组近交系构成,并从理论上论证了该作图群体的高功效性。之后不久该策略便被应用于高粱(*Sorghum vulgare*)的研究中(Casa et al., 2008)。类似的报道还有Stich等(2008)发展了AMMSP(association mapping in multiple segregating populations)群体;Kover等(2009)发展了MAGIC(multiparent advanced generation intercross)群体。这些研究策略均是通过分析有限的等位基因数目,增加分析群体容量的方法,在有效地减少假阳性的同时提高统计功效。其次,在全基因组关联分析数理统计方面,基因型标记与表

型数据的关联分析上,目前主要还是采用“两步法”(Stich, 2009),即第1步分析性状平均值或调整平均值,第2步应用这些数据与基因型数据进行分析,这样不可避免地会引入更多的实验误差。在关联分析计算效率上,由于算法尚不完善,存在运算处理耗时太长的弊病,特别是全基因组扫描关联分析时,数据量极为庞大,耗时甚长(Yu et al., 2006)。针对这两方面的问题,仍需探索有效手段以提高分析的精确度和效率。最后,在利用软件整合、分析数据方面,由于关联分析程序涉及基因型分析、群体结构和亲缘分析、表型鉴定等数据收集和分析过程,故所用软件众多,且其过程极为繁琐,特别是在利用高通量测序和基因芯片技术进行全基因组扫描关联分析时,将会涉及数量巨大的数据、信息处理和分析过程。因此,急需发展更方便高效的程序、软件来整合和分析数据。总之,发展更好的分析方法和算法,实现高功效、低假阳性和提高运算效率将会更有助于关联分析在植物遗传学研究中的广泛应用。

3.3 关联分析在植物遗传学研究中的应用

作物的许多重要农艺性状,如产量、营养品质和抗逆性等多属数量性状,常规育种往往效率不高,利用分子育种技术育种是未来的发展方向。关联分析作为一种高效的QTL鉴定工具,可在分子育种中发挥重要作用。首先,植物育种的本质是优良等位基因的选择与聚合,而关联分析可同时对多个等位基因进行鉴定,筛选到最优等位基因的效率更高,因而关联分析可用于对重要目标基因的等位基因筛选,为优良基因的聚合奠定基础。其次,在分子辅助育种(molecular assisted selection, MAS)中,连锁分析定位到的QTL与目标基因之间遗传图距往往在1 cM以上,且在MAS中易发生目标基因的丢失或连锁累赘,而关联分析鉴定标记则可以达到基因水平,精度为连锁分析的5 000倍以上,在MAS中可极大地提高选择的的目的性和准确性,进而提高育种效率。最后,在分子设计育种方面,关联分析可为基因的功能分析、功能标记开发和反向遗传学研究提供有效信息,从而了解目标基因的位置、结构、遗传效应和功能等全面信息,并在此基础上,通过基因诱变和基因敲除等方式改良目标基因,或通过多个优良基因聚合实现对目标性状的控制。目前,一些关联分析的成果已开始在实践中应用。

如Andersen等(2005)对玉米开花期的分子育种进行了研究;Lubberstedt等(2005)通过对影响青贮玉米消化能力的**bm3**基因进行分析,开发出功能标记,并将其用于青贮玉米的MAS研究中;国际玉米小麦改良中心(CIMMYT)目前正对玉米抗旱有关的候选基因进行关联分析,并在此基础上开发功能标记以进行玉米抗旱的分子育种研究(Xu et al., 2009)。

参考文献

- 莫惠栋, 顾世梁 (2000). 基因组长度的估计方法. 科学通报 45, 1414–1418.
- Agrama HA, Eizenga GC, Yan W (2007). Association mapping of yield and its components in rice cultivars. *Mol Breed* 19, 341–356.
- Andersen JR, Schrag T, Melchinger AE, Zein I, Lübberstedt T (2005). Validation of *Dwarf8* polymorphisms associated with flowering time in elite European inbred lines of maize (*Zea mays* L.). *Theor Appl Genet* 111, 206–217.
- Aranzana MJ, Kim S, Zhao K, Bakker E, Horton M, Jakob K, Lister C, Molitor J, Shindo C, Tang C, Toomajian C, Traw B, Zheng H, Bergelson J, Dean C, Marjoram P, Nordborg M (2005). Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. *PLoS Genet* 1, e60.
- Ashikari M, Sakakibara H, Lin SY, Yamamoto T, Takashi T, Nishimura A, Angeles ER, Qian Q, Kitano H, Matsuoka M (2005). Cytokinin oxidase regulates rice grain production. *Science* 309, 741–745.
- Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT, Jiang R, Mulyati NW, Zhang X, Amer MA, Baxter I, Brachi B, Chory J, Dean C, Debieu M, de Meaux J, Ecker JR, Faure N, Kniskern JM, Jones JD, Michael T, Nemri A, Roux F, Salt DE, Tang C, Todesco M, Traw MB, Weigel D, Marjoram P, Borevitz JO, Bergelson J, Nordborg M (2010). Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465, 627–631.
- Bao JS, Corke H, Sun M (2006). Nucleotide diversity in starch synthase *Ila* and validation of single nucleotide polymorphisms in relation to starch gelatinization temperature and other physicochemical properties in rice (*Oryza sativa* L.). *Theor Appl Genet* 113, 1171–1183.
- Belo A, Zheng PZ, Luck S, Shen B, Meyer DJ, Li BL,

- Tingey S, Rafalski A** (2008). Whole genome scan detects an allelic variant of *fod2* associated with increased oleic acid levels in maize. *Mol Genet Genomics* **279**, 1–10.
- Breseghele F, Sorrells ME** (2006). Association mapping of kernel size and milling quality in wheat (*Triticum aestivum* L.) cultivars. *Genetics* **172**, 1165–1177.
- Briggs W, McMullen MD, Gaut BS, Doebley J** (2007). Linkage mapping of domestication loci in a large maize-teosinte backcross resource. *Genetics* **177**, 1915–1928.
- Camus-Kulandaivelu L, Veyrieras JB, Madur D, Combes V, Fourmann M, Barraud S, Dubreuil P, Gouesnard B, Manicacci D, Charcosset A** (2006). Maize adaptation to temperate climate: relationship with population structure and polymorphism in the *Dwarf8* gene. *Genetics* **10**, 1534–1572.
- Cardon LR, Palmer JL** (2003). Population stratification and spurious allelic association. *Lancet* **361**, 598–604.
- Casa AM, Pressoira G, Brown PJ, Mitchell SE, Rooney WL, Tuinstra MR, Franks CD, Kresovich S** (2008). Community resources and strategies for association mapping in sorghum. *Crop Sci* **48**, 30–40.
- Cockram J, White J, Leigh FJ, Lea VJ, Chiapparino E, Laurie DA, Mackay IJ, Powell W, O'Sullivan DM** (2008). Association mapping of partitioning loci in barley. *BMC Genet* **18**, 9–16.
- Delvin B, Risch N** (1995). A comparison of linkage disequilibrium measures for fine mapping. *Genomics* **29**, 311–322.
- Devlin B, Roeder K** (1999). Genomic control for association studies. *Biometrics* **55**, 997–1004.
- Doerge RW** (2002). Mapping and analysis of quantitative trait loci in experimental populations. *Nat Rev Genet* **3**, 43–52.
- Ehrenreich IM, Hanzawa Y, Chou L, Roe J, Kover P, Purugganan M** (2009). Candidate gene association mapping of Arabidopsis flowering time. *Genetics* **183**, 325–335.
- Ehrenreich IM, Stafford PA, Purugganan MD** (2007). The genetic architecture of shoot branching in *Arabidopsis thaliana*: a comparative assessment of candidate gene associations vs. quantitative trait locus mapping. *Genetics* **176**, 1223–1236.
- Epstein MP, Allen AS, Satten GA** (2007). A simple and improved correction for population stratification in case-control studies. *Am J Hum Genet* **80**, 912–930.
- Flint-Garcia SA, Thornsberry JM, Buckler ES** (2003). Structure of linkage disequilibrium in plants. *Annu Rev Plant Biol* **54**, 357–374.
- Flint-Garcia SA, Thillet A, Yu J, Pressoir G, Romero SM, Mitchell SE, Doebley J, Kresovich S, Goodman MM, Buckler ES** (2005). Maize association population: a high resolution platform for quantitative trait locus dissection. *Plant J* **44**, 1054–1064.
- Frary A, Nesbitt TC, Grandillo S, Knaap E, Cong B, Liu J, Meller J, Elber R, Alpert KB, Tanksley SD** (2000). *fw2.2*: a quantitative trait locus key to the evolution of tomato fruit size. *Science* **289**, 85–88.
- Gaut BS, Long AD** (2003). The lowdown on linkage disequilibrium. *Plant Cell* **15**, 1502–1506.
- Gebhardt C, Ballvora A, Walkemeier B, Oberhagemann P, Schuler K** (2004). 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. *Mol Breed* **13**, 93–102.
- Gonzalez-Martinez SC, Ersoz E, Brown GR, Wheeler NC, Neale DB** (2006). DNA sequence variation and selection of tag singlenucleotide polymorphisms at candidate genes for drought-stress response in *Pinus taeda* L. *Genetics* **172**, 1915–1926.
- Gonzalez-Martinez SC, Wheeler NC, Ersoz E, Nelson CD, Neale DB** (2007). Association genetics in *Pinus taeda* L. I. Wood property traits. *Genetics* **175**, 399–409.
- Gupta PK, Rustgi S, Kulwal PL** (2005). Linkage disequilibrium and association in higher plants: present status and future prospects. *Plant Mol Biol* **57**, 461–485.
- Hamblin MT, Close TJ, Bhat PR, Chao S, Kling JG, Abraham KJ, Black T, Brooks WS, Cooper B, Griffey CA, Hayes Hole DJ, Horsley RD, Obert DE, Smith KP, Ullrich SE, Muehlbauer GJ, Jannink JL** (2010). Population structure and linkage disequilibrium in US barley germplasm: implications for association mapping. *Crop Sci* **50**, 556–566.
- Hansen M, Kraft T, Ganestam S, Säll T, Nilsson N** (2001). Linkage disequilibrium mapping of the bolting gene in sea beet using AFLP markers. *Genet Res* **77**, 61–66.
- Harjes CE, Rocheford TR, Bai L, Brutnell TP, Kandianis CB, Sowinski SG, Stapleton AE, Vallabhaneni R, Williams M, Wurtzel ET, Yan J, Buckler ES** (2008). Natural genetic variation in lycopene epsilon cyclase tapped for maize biofortification. *Science* **319**, 330–333.
- Hastbacka J, de la Chapelle A, Kaitila I, Sistonen P, Weaver A, Lander E** (1992). Linkage disequilibrium

- mapping in isolated founder populations: diastrophic dysplasia in Finland. *Nat Genet* **2**, 204–211.
- Hey J, Machado CA** (2003). The study of structured populations—new hope for a difficult and divided science. *Nat Rev Genet* **4**, 535–543.
- Holland JB** (2007). Genetic architecture of complex traits in plants. *Curr Opin Plant Biol* **10**, 156–161.
- Iwata H, Ebana K, Uga Y, Hayashi T, Jannink JL** (2010). Genome-wide association study of grain shape variation among *Oryza sativa* L. germplasms based on elliptic Fourier analysis. *Mol Breed* **25**, 203–215.
- Jorde JB** (2000). Linkage disequilibrium and the search for complex disease gene. *Genome Res* **10**, 1435–1444.
- Jun TH, Van K, Kim MY, Lee SH, Walker DR** (2008). Association analysis using SSR markers to find QTL for seed protein content in soybean. *Euphytica* **162**, 179–191.
- Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E** (2008). Efficient control of population structure in model organism association mapping. *Genetics* **178**, 1709–1723.
- Kover PX, Valdar W, Trakalo J, Scarcelli N, Ehrenreich IM, Purugganan MD, Durrant C, Richard M** (2009). A multiparent advanced generation inter-cross to fine-map quantitative traits in *Arabidopsis thaliana*. *PLoS Genet* **5**, e1000551.
- Kraakman ATW, Niks RE, van den Berg PMMM, Stam P, van Eeuwijk FA** (2004). Linkage disequilibrium mapping of yield and yield stability in modern spring barley cultivars. *Genetics* **168**, 435–446.
- Krill AM, Kirst M, Kochian LV, Buckler ES, Hoekenga OA** (2010). Association and linkage analysis of aluminum tolerance genes in maize. *PLoS One* **5**, e9958.
- Lander E, Kruglyak L** (1995). Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nat Genet* **11**, 241–247.
- Li Q, Yu K** (2008). Improved correction for population stratification in genome-wide association studies by identifying hidden population structures. *Genet Epidemiol* **32**, 215–226.
- Lubberstedt T, Zein I, Andersen JR, Wenzel G, Krutzfeldt B, Eder J, Ouzunova M, Chun S** (2005). Development and application of functional markers in maize. *Euphytica* **146**, 101–108.
- Malosetti M, van der Linden CG, Vosman B, van Eeuwijk FA** (2007). A mixed-model approach to association mapping using pedigree information with an illustration of resistance to phytophthora infestans in potato. *Genetics* **175**, 879–889.
- Meuwissen THE, Goddard ME** (2000). Fine mapping of quantitative trait loci using linkage disequilibria with closely linked marker loci. *Genetics* **155**, 421–430.
- Myles S, Chia JM, Hurwitz B, Simon C, Zhong GY, Buckler ES, Ware D** (2010). Rapid genomic characterization of the genus *Vitis*. *PLoS One* **5**, e8219.
- Olsen KM, Halldorsdottir SS, Stinchcombe JR, Weigand C, Schmitt J, Purugganan MD** (2004). Linkage disequilibrium mapping of *Arabidopsis* *CRY2* flowering time alleles. *Genetics* **167**, 1361–1369.
- Olsen KM, Purugganan MD** (2002). Molecular evidence on the origin and evolution of glutinous rice. *Genetics* **162**, 941–950.
- Oraguzie NC, Wilcox PL, Rikkerink EHA, de Silva HN** (2007). Linkage disequilibrium. In: Oraguzie NC, Rikkerink EH, Gardiner SE, de Silva HN, eds. Association Mapping in Plants. New York: Springer Verlag. pp. 11–39.
- Palaisa KA, Morgante M, Williams M, Rafalski A** (2003). Contrasting effects of selection on sequence diversity and linkage disequilibrium at two phytoene synthase loci. *Plant Cell* **15**, 1795–1806.
- Patterson N, Price AL, Reich D** (2006). Population structure and eigenanalysis. *PLoS Genet* **2**, e90.
- Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D** (2006). Principal components analysis corrects for stratification in genome-wide association studies. *Nat Genet* **38**, 904–909.
- Pritchard JK, Rosenberg NA** (1999). Use of unlinked genetic markers to detect population stratification in association studies. *Am J Hum Genet* **65**, 220–228.
- Pritchard JK, Stephens M, Rosenberg NA, Donnelly P** (2000). Association mapping in structured populations. *Am J Hum Genet* **67**, 170–181.
- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC** (2007). PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* **81**, 559–575.
- Rafalski A** (2002). Applications of single nucleotide polymorphisms in crop genetics. *Curr Opin Plant Biol* **5**, 94–100.
- Remington DL, Thornsberry JM, Matsuoka Y, Wilson LM, Whitt SR, Doebley J, Kresovich S, Goodman MM, Buckler ES** (2001). Structure of linkage disequilibrium and phenotypic associations in the maize genome. *Proc*

- Natl Acad Sci USA* **98**, 11479–11484.
- Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, Feldman MW** (2002). Genetic structure of human populations. *Science* **298**, 2381–2385.
- Rostoks N, Ramsay L, MacKenzie K, Cardle L, Svensson JT, Bhat P, Roose ML, Stein N, Varshney RK, Marshall D, Graner A, Close TJ, Waugh R** (2006). A recent history of artificial outcrossing facilitates whole genome association mapping in elite inbred crop varieties. *Proc Natl Acad Sci USA* **103**, 18656–18661.
- Saïdou AA, Mariac C, Luong V, Pham JL, Bezancon G, Yigouroux Y** (2009). Association studies identify natural variation at *PHYC* linked to flowering time and morphological variation in pearl millet. *Genetics* **182**, 899–910.
- Salvi S, Sponza G, Morgante M, Tomes D, Niu X, Fengler KA, Meeley R, Ananiev EV, Svitashov S, Bruggemann E, Li B, Hainey CF, Radovic S, Zaina G, Rafalski JA, Tingey SV, Miao GH, Phillips RL, Tuberosa R** (2007). Conserved noncoding genomic sequences associated with a flowering-time quantitative trait locus in maize. *Proc Natl Acad Sci USA* **104**, 11376–11381.
- Shehzad T, Iwata H, Okuno K** (2009). Genome-wide association mapping of quantitative traits in sorghum (*Sorghum bicolor* (L.) Moench) by using multiple models. *Breed Sci* **59**, 217–227.
- Sköt L, Humphreys J, Humphreys MO, Thorogood D, Gallagher J, Sanderson R, Armstead IP, Thomas ID** (2007). Association of candidate genes with flowering time and water-soluble carbohydrate content in *Lolium perenne*(L.). *Genetics* **177**, 535–547.
- Sköt L, Humphreys MO, Armstead I, Heywood S, Sköt KP, Sanderson R, Thomas ID, Chorlton KH, Hamilton NRS** (2005). An association mapping approach to identify flowering time genes in natural populations of *Lolium perenne* (L.). *Mol Breed* **15**, 233–245.
- Spielman RS, McGinnis RE, Ewens WJ** (1993). Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus. *Am J Hum Genet* **52**, 506–516.
- Stich B** (2009). Comparison of mating designs for establishing nested association mapping populations in maize and *Arabidopsis thaliana*. *Genetics* **183**, 1525–1534.
- Stich B, Melchinger AE** (2009). Comparison of mixed-model approaches for association mapping in rapeseed, potato, sugar beet, maize, and *Arabidopsis*. *BMC Genomics* **10**, 94–94.
- Stich B, Melchinger AE, Heckenberger M, Möhring J, Schechert A, Piepho HP** (2008). Association mapping in multiple segregating populations of sugar beet (*Beta vulgaris* L.). *Theor Appl Genet* **117**, 1167–1179.
- Szalma SJ, Buckler ES, Snook ME, McMullen MD** (2005). Association analysis of candidate genes for maysin and chlorogenic acid accumulation in maize silks. *Theor Appl Genet* **110**, 1324–1333.
- Thornsberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES** (2001). Dwarf8 polymorphisms associate with variation in flowering time. *Nat Genet* **28**, 286–289.
- Thumma BR, Nolan MF, Evans R, Moran GF** (2005). Polymorphisms in cinnamoyl CoA reductase (CCR) are associated with variation in microfibril angle in *Eucalyptus* spp. *Genetics* **171**, 1257–1265.
- Tian Z, Qian Q, Liu Q, Yan M, Liu X, Yan C, Liu G, Gao Z, Tang S, Zeng D, Wang Y, Yu J, Gu M, Li J** (2009). Allelic diversities in rice starch biosynthesis lead to a diverse array of rice eating and cooking qualities. *Proc Natl Acad Sci USA* **106**, 21760–21765.
- Tommasini U, Schnurbusch T, Fossati D, Mascher F, Keller B** (2007). Association mapping of *Stagonospora nodorum* blotch resistance in modern European winter wheat varieties. *Theor Appl Genet* **115**, 697–708.
- Tracy WF, Whitt SR, Buckler ES** (2006). Recurrent mutation and genome evolution: example of sugary1 and the origin of sweet maize. *Crop Sci* **46**, S49–S54.
- Wang H, Nussbaum-Wagler T, Li B, Zhao Q, Yigouroux Y, Faller M, Bomblies K, Lukens L, Doebley JF** (2005). The origin of the naked grains of maize. *Nature* **436**, 714–719.
- Weber AL, Briggs WH, Rucker J, Baltazar BM, Sanchez-Gonzalez JJ, Feng P, Buckler ES, Doebley J** (2008). The genetic architecture of complex traits in teosinte (*Zea mays* ssp. *parviglumis*): new evidence from association mapping. *Genetics* **180**, 1221–1232.
- Wei XM, Jackson PA, McIntyre CL, Aitken KS, Croft B** (2006). Associations between DNA markers and resistance to diseases in sugarcane and effects of population substructure. *Theor Appl Genet* **114**, 155–164.
- Wen W, Mei H, Feng F, Yu S, Huang ZC, Wu JH, Chen L, Xu XY, Luo LJ** (2009). Population structure and association mapping on chromosome 7 using a diverse panel of Chinese germplasm office (*Oryza sativa* L.). *Theor Appl Genet* **119**, 459–470.
- Wilson LM, Whitt SR, Ibanez AM, Rocheford TR, Goodman MM, Buckler ES** (2004). Dissection of maize kernel

composition and starch production by candidate gene associations. *Plant Cell* **16**, 2719–2733.

Xu Y, Skinner DJ, Wu H, Palacios-Rojas N, Araus JL, Yan J, Gao S, Warburton ML, Crouch JH (2009). Advances in maize genomics and their value for enhancing genetic gains from breeding. *Int J Plant Genomics* **2009**, Article ID 957602.

Yan J, Shah T, Warburton ML, Buckler ES, McMullen MD, Crouch J (2009a). Genetic characterization and linkage disequilibrium estimation of a global maize collection using SNP markers. *PLoS One* **4**, e8451.

Yan WG, Li Y, Agrama HA, Luo D, Gao FY, Lu XJ, Ren GJ (2009b). Association mapping of stigma and spikelet characteristics in rice (*Oryza sativa* L.). *Mol Breed* **24**, 277–292.

Yano M, Katayosea Y, Ashikarib M, Yamanouchi U, Monnac L, Fuseb T, Babac T, Yamamoto K, Ume-haraa Y, Nagamura Y, Sasakia T (2000). *Hd1*, a major

photoperiod sensitivity quantitative trait locus in rice, is closely related to the Arabidopsis flowering time gene *CONSTANS*. *Plant Cell* **12**, 2473–2483.

Yu J, Holland JB, McMullen MD, Buckler ES (2008). Genetic design and statistical power of nested association mapping in maize. *Genetics* **178**, 539–551.

Yu JM, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doe-bley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB, Kresovich S, Buckler ES (2006). A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. *Nat Genet* **38**, 203–208.

Zhao HH, Fernando RL, Dekkers JCM (2007). A power and precision of alternate methods for linkage disequilibrium mapping of quantitative trait loci. *Genetics* **175**, 1975–1986.

Zhu C, Yu J (2009). Nonmetric multidimensional scaling corrects for population structure in whole genome association studies. *Genetics* **182**, 875–888.

Association Analysis and Its Application in Plant Genetic Research

Xianjie Tan^{1,2}, Zikai Wu¹, Weidong Cheng², Tianyu Wang³, Yu Li^{3*}

¹School of Agronomy, Guangxi University, Nanning 530005, China; ²Guangxi Maize Research Institute, Guangxi Academy of Agricultural Sciences, Nanning 530227, China; ³Institute of Crop Science, Chinese Academy of Agricultural Sciences, Beijing 100081, China

Abstract Most of the important economic traits in plant genetics are attributed to complex traits. Quantitative trait loci (QTL) mapping is an important tool for studying complex traits. However, because of its limitations, QTL mapping is used only in certain cases. Association analysis, a new approach for dissecting complex traits based on linkage disequilibrium, can overcome the limitations of QTL mapping for use in plant genetic studies. As compared with QTL mapping and other methods used to dissect complex traits, association analysis does not need specific population construction, multiple alleles can be analyzed synchronously, and mapping resolution is higher. Here, we introduce the methodology and properties of association analysis, outline the progress with such analysis in plant genetics research, and discuss future applications and potential development of association analysis in plant genetics research.

Key words association analysis, haplotypes, linkage disequilibrium, quantitative trait loci

Tan XJ, Wu ZK, Cheng WD, Wang TY, Li Y (2011). Association analysis and its application in plant genetic research. *Chin Bull Bot* **46**, 108–118.

* Author for correspondence. E-mail: yuli@mail.caas.net.cn