

# 复杂数量性状的遗传分析及功能作图

黄中文

(河南科技学院生命科技学院, 河南 新乡 453003)

**摘要:** 大多数重要的生物学、生化和农学性状是数量性状, 受 1 个基因网络控制, 且易受环境因素影响。近 20 年来, 随着统计学和分子生物学的迅猛发展, 对数量性状的研究也逐渐深入, 提出了很多研究方法。文中综述了 QTLs 的检测方法进展, 评价了各种方法的优劣, 着重介绍了目前国际上最新提出的多维数量性状的研究方法——功能作图的思想。

**关键词:** 复杂数量性状; 动态性状; 功能作图

**中图分类号:** Q-332      **文献标识码:** A      **文章编号:** 1004-3268(2006)11-0019-04

作物中大多数重要的农艺、生化、经济性状是典型的数量性状(QTLs), 是受众多基因和环境因素所构成的 1 个基因网络控制, 在这个基因网络中, 每个基因的效应很小<sup>[1]</sup>。由于这个原因, 对这些所谓的数量性状或称之为复杂数量性状的研究成为植物数量遗传学中重要的课题。长期以来, 只能借助于经典数理统计方法, 将复杂的基因网络作为一个整体, 用平均值和方差来描述数量性状的遗传特征, 无法对构成基因网络中的每个效应微小的基因进行单独分析, 不了解单个基因的效应、基因间的互作效应, 从而限制了植物育种中对数量性状的遗传操作能力。20 世纪 80 年代以来迅速发展的分子标记技术为深入研究数量性状的遗传机制提供了可能, 统计学与分子生物学相结合, 世界各国科学家相继开展了数量性状的遗传分析和 QTLs 定位研究, 提出了多种 QTLs 定位方法和遗传模型, 对复杂数量性状的遗传剖析能力也在逐渐提高。

最近几年, QTLs 定位所依赖的计算技术和统计学均得以迅猛发展, 新提出的 QTLs 分析方法, 可以研究之前不能分析的基因的各种互作效应; 作图群体也从 F<sub>2</sub>, RIL, DH 等扩展到永久 F<sub>2</sub> 和用系谱有清楚记载的品种群体。文中介绍了 QTLs 定位的各种方法, 重点介绍动态发育性状的遗传分析方法, 即功能作图。

## 1 作图群体

构建一个特定的群体, 并获得该群体的标记基

因型数据和数量性状表型数据是开展 QTLs 研究工作的前提。根据分离群体的特点, 作图群体分为临时性分离群体、永久性分离群体和近等基因系群体三大类。临时性分离群体包括单交组合产生的 F<sub>2</sub> 及其衍生的 F<sub>3</sub>~F<sub>4</sub> 家系, 以及回交群体, 它的显著特征是群体中的每一个个体其后代均可发生分离。除非自交不亲和植物, F<sub>2</sub> 容易配制, 而且提供给遗传分析的信息最为丰富, 可以同时估计加性效应和显性效应。但是由于 F<sub>2</sub> 植株存在分离, 很难进行多年多点研究。由于临时性群体容易获得, 早期 QTLs 研究常用这类群体。永久性分离群体主要包括重组自交系群体(Recombinant inbred lines, RIL)和加倍单倍体群体(Doubled haploid lines, DH), 它的显著特征是群体中每一个体后代稳定, 不发生分离, 可以连续不断地提供家系内遗传上一致的种子。因此, 可以进行重复区组试验, 把区组效应、重复效应和随机误差最小化或分解开来, 增加检测 QTLs 的准确性和效力。但是, 构建 RIL 群体需要很长的时间。构建 DH 群体, 受基因型限制, 难度较大, 而且不能估计显性效应。临时性分离群体和永久性分离群体在 QTLs 定位方面存在共同的局限性, 即对 QTLs 的精确定位能力不强, 对效应小的 QTLs 检测能力有限, 增加群体大小和分子标记数目也只能在一定程度上缓解这种局限性。近等基因系(Near isogenic lines, NIL)基本特征是整套染色体组的绝大部分区域完全相同, 只在少数几个甚至 1 个区段彼此存在差异。因此, 它能使基因组中

收稿日期: 2006-05-29

基金项目: 河南科技学院研究生重点支持项目(02130)

作者简介: 黄中文(1971-), 男, 河南潢川人, 在读博士研究生, 主要从事生物统计学和生物技术作物育种中的应用研究。

只存在 1 个或几个 QTLs 分离, 消除其他背景干扰和消去主效 QTLs 对微效 QTLs 效应的掩盖作用, 对 QTLs 的精确定位很有帮助。尽管近等基因系构建时间长、工作量大, 但近等基因系对 QTLs 分解和精确定位的独特优越性使其成为分离和克隆 QTLs 的理想群体。特别是现在, 近等基因系的构建和应用已经成为 QTLs 的粗定位工作之后的重点研究内容。

近几年, 用于 QTLs 作图又提出了永久 F<sub>2</sub> 群体 (Immortalized F<sub>2</sub> Population, IF<sub>2</sub>) 和用系谱和性状有清楚记载的品种群体<sup>[2]</sup>。

异交植物构建永久 F<sub>2</sub> 群体的方法是: 将 F<sub>2</sub> 的每个植株在 F<sub>3</sub> 代以及以后世代中进行自由随机授粉, 因为 F<sub>2</sub> 基因型在每个 F<sub>2</sub> 家系中得到维持, 因此, 可以消除用 F<sub>2</sub> 世代进行分子作图的障碍。

应用 IF<sub>2</sub> 植株花粉产生混合种子保持了 F<sub>2</sub> 群体的基因型组成, 不会有大的遗传漂变, 这样就克服了单纯 F<sub>2</sub> 群体作图的缺陷, IF<sub>2</sub> 群体作图允许多点 (相对于两点) 方法进行, 这样共显性分子数据可以得到充分的利用。

对于自花植物, 永久 F<sub>2</sub> 群体的组建方法是<sup>[3]</sup>: 将 RIL 或 DH 群体各家系分成 2 组, 进行组间随机交配, 获得系间交配的 F<sub>1</sub> 品系。这些 F<sub>1</sub> 品系组成的群体, 与 F<sub>2</sub> 具有相同的遗传结构, 每年配制同样的组合满足重复试验的需要, 使得群体的遗传结构得以长期保持。通过 RIL 或 DH 群体系间随机交配获得的 F<sub>1</sub> 构建的永久 F<sub>2</sub> 群体, 既具有信息量大, 可以估计显性效应以及与显性有关的上位性效应的优点, 又能为多单位合作研究或多环境研究 QE 互作研究源源不断提供大量试验材料, 有助于进一步揭示复杂数量性状的遗传结构。由 RIL 群体系间随机交配构建的永久 F<sub>2</sub> 群体有利于鉴别紧密连锁的标记和 QTLs。然而, 永久 F<sub>2</sub> 群体在实施上仍有以下困难: (1) 杂交组合配制工作量大, 难度高, 很多组合难于得到足够的种子, 造成数据缺失; (2) 不同 RIL 或 DH 系的抽穗期很不一致, 对于大量配组来说, 很难做到完全随机。这些因素会导致构建的永久 F<sub>2</sub> 群体往往偏离正常的理论比, 从而导致 QTLs 位置效应的估计出现偏差。

Yuan Min Zhang 等<sup>[2]</sup> 提出用系谱和性状有清楚记载的自然群体作为 QTLs 作图群体的新方法。Yuan Min Zhang 用一套具 70 年记载的玉米自交系系谱, 该系谱包括 404 个自交系, 其中 103 个是基础亲本。通过 Monte Carlo 模拟估算 IBD (Identity-

by-descent) 矩阵, 定位了玉米的开花性状的 QTLs。

## 2 QTLs 定位方法

### 2.1 单一标记分析法

单标记分析法就是通过方差分析、回归分析或似然比检验, 比较不同标记基因型数量性状均值的差异<sup>[4]</sup>。如存在显著差异, 则说明控制该数量性状的 QTLs 与标记有连锁。由于单一标记分析法不需要完整的分子标记连锁图谱, 因而早期的 QTLs 定位研究多采用这种方法<sup>[5]</sup>。方差分析法和简单线性回归法只能检测标记与 QTLs 的连锁, 无法估计它们之间的重组率。传统的单个标记分析方法存在许多缺点, 例如: (1) 不能确定标记是与 1 个 QTLs 连锁还是与几个 QTLs 连锁; (2) 无法确切估计 QTLs 的可能位置; (3) 由于遗传效应与重组率混合在一起, 导致低估了 QTLs 的遗传效应 ( $m_1 - m_0 = (1-2r) \times a$ ); (4) 容易出现假阳性; (5) 检测效率不高, 所需的个体数较多。

### 2.2 区间作图法 (IM)

鉴于传统的单标记分析方法存在的问题, Lander 和 Bostein 等<sup>[6]</sup> 提出了基于 2 个侧邻标记的区间作图法。其遗传假设是, 数量性状遗传变异只受 1 对基因控制, 表型变异受遗传效应 (固定效应) 和剩余误差 (随机效应) 控制, 不存在基因型与环境的互作。区间作图法可以估算 QTLs 加性和显性效应值。与单标记分析法相比, 区间作图法具有以下特点: 能从支撑区间推断 QTLs 的可能位置; 可利用标记连锁图在全染色体组系统地搜索 QTLs, 如果 1 条染色体上只有 1 个 QTLs, 则 QTLs 的位置和效应估计趋于渐进无偏; QTLs 检测所需的个体数大大减少。但 IM 也存在不足: 回归效应为固定效应; 无法估算 Q × E 互作、无法检测复杂的遗传效应 (如上位效应等); 当相邻 QTLs 相距较近时, 由于其作图精度不高, QTLs 间相互干扰导致出现 Ghost QTLs; 1 次只应用 2 个标记进行检查, 效率很低。

### 2.3 复合区间作图法 (CIM)

为了解决区间作图法存在的问题, Rodolphe 和 Lefort<sup>[7]</sup> 提出了一种利用整个基因组上的标记进行全局检测的多标记模型。笔者证明, 染色体上不同类型效应的参数分解是相互独立的, 与一个标记相关联的效应估计值只与侧邻标记的同类型效应相关。进而提出用区间作图法对 1 条染色体进行检测的同时, 在模型中保留其他染色体上的标记, 以减少剩余误差。但是他们的模型不能提供 QTLs 数目、

位置和效应的准确估计,特别是当多个 QTLs 之间无连锁时,该法估计的精度和效率会降低。

Zeng<sup>[8]</sup> 提出的结合了区间作图和多元回归的一种 QTLs 是作图方法(CIM)。其遗传假定是数量性状受多基因控制。该方法中拟合了其他遗传标记,即在对某一特定标记区间进行检测时,将与其他 QTLs 连锁的标记也拟合在模型中以控制背景遗传效应。CIM 主要优点是:由于仍采用 QTLs 似然图来显示 QTLs 的可能位置及显著程度,从而保证了 IM 作图法的优点;假如不存在上位性和 QTLs 与环境互作, QTLs 的位置和效应的估计是渐进无偏的;以所选择的多个标记为条件(即进行的是区间检测),在较大程度上控制了背景遗传效应,从而提高了作图的精度和效率。存在的不足是:由于将两侧标记用作区间作图,对相邻标记区间的 QTLs 估计会引起偏离;同 IM 一样,将回归效应视为固定效应,不能分析基因型与环境的互作及复杂的遗传效应(如上位效应等);当标记密度过大时,很难选择标记的条件因子。

#### 2.4 QTLs 定位的混合线性模型方法(MCIM)

近年,一些学者采用两向方差分析法和多元回归方法分析了基于遗传标记的数量性状的上位性。1998 年,朱军<sup>[9]</sup> 提出用随机效应的预测方法获得基因型效应及基因型×环境互作效应的预测值,然后再用区间作图法或复合区间作图法分别进行遗传主效应及基因型×环境互作效应的 QTLs 定位分析。该方法的遗传假定是数量性状受多基因控制,它将群体均值及 QTLs 的各项遗传效应看作为固定效应,而将环境、QTLs 与环境、分子标记等效应看作为随机效应。由于 MCIM 将效应值估计和定位分析相结合,既可为无偏地分析 QTLs 与环境的互作效应,又提高了作图的精度和效率。此外该模型可以扩展到分析具有加×加、加×显、显×显上位的各种遗传主效应及其与环境互作效应的 QTLs。利用这些效应值的估计,可预测基于 QTLs 主效应的普通杂种优势和基于 QTLs 与环境互作效应的互作杂种优势,因而其具有广阔的应用前景。

### 3 功能作图(Functional mapping)

传统的 QTLs 统计作图方法尽管很实用,但是却忽略了性状形成的发育特性。因此,性状的遗传控制也应该是一个独立变量的函数。现代发育遗传研究表明,基因的表达有一定的时序性和选择性。借助分子标记连锁图和 QTLs 定位技术,在分子水

平研究特定发育时段内基因的表达,有助于揭示单个基因在性状发育过程中的作用模式,阐明性状的发育遗传规律和进化机制。

传统的 QTLs 分析往往是对数量性状在测定末期的最终表型值进行定位,这种静态的 QTLs 分析方法不能充分反映发育过程中基因的真实作用方式,也不能精确、完整地定位显著影响植株性状的众多 QTLs。因而,基于最终表型值定位的 QTLs 分子标记辅助选择,其有效性还有待商榷。结合 QTLs 作图和发育分析方法对数量性状进行动态 QTLs 作图,能检测到更多的数量性状位点,对分子标记辅助选择、数量性状的遗传操作有重要的参考价值。

动态性状的遗传分析给统计学提出了一个挑战。复杂性状的全基因组作图方法,为解决这一问题提供了通用的统计框架,美国 Florida 大学统计系 Rongling Wu 提出了一种全新的作图方法,这就是功能作图<sup>[10]</sup>。功能作图将动态性状的数学特性与 QTLs 作图理论组装在一起。功能作图能够分解发育性状的遗传控制因子,还可以利用高通量的 SNP 数据去定位 QTNs 数量性状核苷酸。

#### 3.1 遗传控制的动态模式

各种动态性状的表现形式是不同的。因此,这些动态性状的遗传控制模式在某一特定时间点的表现也不同<sup>[11]</sup>。例如,有些 QTLs 是永久表达的,而有些 QTLs 则在早期发育阶段开启,有些在特定的发育阶段开启。

#### 3.2 功能作图的优势

功能作图用数学模型将基因的活动(或环境影响)与发育联系在一起,建立了几个数学公式去描述一个表型的发育和阐明得到的生长进程模式的特征。例如,从描述株高、大小和体重的 S 型生长曲线可以推导出一系列生长公式<sup>[11]</sup>。最近,West 等<sup>[12]</sup> 解释了为什么一个生物体的生长遵循 S 型曲线,其基本规律是由于代谢能量会在保持现有组织和生成新的生物量之间进行分配。

功能作图将数学公式整合到 QTLs 作图的统计框架之中,估计决定一个性状发育进程的特定基因型的参数,而不是去直接估计基因在所有时间点的基因效应。由于结合了生物学规律(用数学模型定义的),而且需要估计的遗传参数更少,因此,功能作图增加了检测到显著 QTLs 的效力。另外,通过

功能作图对数学公式的估计, 会加深我们对控制发育变化的遗传、生化和生理途径的理解。这些信息对药物的治疗和预防(例如, 基因设计治疗)、选择育种和 MAS 是有用的。

### 3.3 功能作图的前景

随时间或其他独立变化而变化的性状在农业、生物和医学研究中是很重要的。正是这个原因, 对这些所谓的纵向的或动态的性状的遗传分析, 已经成为统计学和遗传学研究的焦点, 目的是在基因型水平上预测遗传控制的动态改变<sup>[13]</sup>。

功能作图的框架建立在模拟全基因组分布的 QTLs 和 QTNs 之间的遗传互作。功能作图整合了全基因组的遗传信息, 通过统计方法, 例如: 模型选择或收敛估计<sup>[14]</sup>, 对一个动态性状所有可能的基因所构成的基因互作网络作完全的刻画。

功能作图作为孟德尔遗传、统计学和发育的结合, 优于仅结合统计和遗传的传统的作图方法。功能作图在农学、进化遗传学和医学遗传学研究中有一定的应用价值<sup>[15~17]</sup>。

## 4 小结

研究数量性状的遗传结构和遗传控制模式, 是研究控制数量性状基因的基础。只有搞清楚这些问题, 数量性状才能逐渐在实践中应用。近 20 年来, 相继提出了多种研究方法, 对 QTLs 的检测能力也在逐渐增强, 但仅有几个 QTLs 被克隆。因此, 研究实践需要更精确、更有效力的检测方法。现代统计遗传学一个很重要的发展方向是研究有多个时间序列的动态性状, 因为这更接近生物学本质。从这个意义上论, 功能作图的思想值得关注。

### 参考文献:

[ 1 ] Lynch M, Walsh B. Genetics and Analysis of Quantitative traits[ M ]. Sinauer, Sunderland, Massachusetts 1998.

[ 2 ] YuanMing Zhang, YongCai Mao, ChongQing Xie, *et al.* Mapping quantitative trait loci using naturally occurring genetic variance among commercial inbred lines of maize(*Zea mays* L. )[ J ]. Genetics, 2005, 169: 2267—2275.

[ 3 ] Jinping Hua, Yongzhong Xing, Weiren Wu, *et al.* Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in an elite rice hybrid[ J ]. PNAS, 2004, 100

(4): 2574—2579.

[ 4 ] Jayakar S D. On the detection and estimation of linkage between a locus influencing a quantitative character and a marker locus[ J ]. Biometrics, 1970, 26: 451—464.

[ 5 ] Edwards M D, Stuber C W, Wendel J F. Molecular marker facilitated investigations of quantitative trait loci in maize. I. Numbers, genomic distribution and types of gene action[ J ]. Genetics, 1987, 116: 113—125.

[ 6 ] Lander E S, Botstein D. Mapping medelian factors underlying quantitative traits using RFLP linkage maps [ J ]. Genetics, 1989, 121: 185—199.

[ 7 ] Rodolphe F, Lefort M. A multi-marker model for detecting chromosomal segments displaying QTL activity [ J ]. Genetics, 1993, 134: 1277—1288.

[ 8 ] Zeng Z B. Theoretical basis of separation of multiple linked gene effects on mapping quantitative trait loci [ J ]. PNAS, 1993, 90: 10972—10976.

[ 9 ] 朱军. 复杂数量性状基因定位的混合线性模型方法 [ A ]. 王连铮, 戴景瑞. 全国作物育种学术讨论会论文集 [ C ]. 北京: 中国农业科技出版社, 1998. 11—20.

[ 10 ] Roingling Wu, Min Lin. Functional mapping-how to map and study the genetics architecture of dynamic complex traits[ J ]. Nature Reviews Genetics, 2006, 7 (3): 229—237.

[ 11 ] Ricø S H. The analysis of ontogenetic trajectories: When a change in size or shape is not heterophony [ J ]. Proc Natl Acad Sci, 1997, 94: 907—912.

[ 12 ] West G B, Brown J H, Enquist B J. A general model for ontogenetic growth[ J ]. Nature, 2001, 413: 628—631.

[ 13 ] Meyer K. Random regression to model phenotypic variation in monthly weights of Australian beef cows [ J ]. Livestock Prod Sci, 2000, 65: 19—38.

[ 14 ] Wang H. Bayesian shrinkage estimation of QTL parameters [ J ]. Genetics, 2005, 170: 465—480.

[ 15 ] Arthur W. The emerging conceptual framework of evolutionary developmental biology [ J ]. Nature, 2002, 415: 757—764.

[ 16 ] Dusheck J. It's the ecology, stupid[ J ]. Nature, 2002, 418: 578—579.

[ 17 ] Wu R L, Ma C X, Lou Y X, *et al.* Molecular dissection of allometry, ontogeny and plasticity: A genomic view of developmental biology[ J ]. Bio Science, 2003, 53: 1041—1047.