

云南省自然科学基金资助项目

杨德主编

ANALYSIS OF QUANTITATIVE GENETICS
FOR RICE HYBRID OF DIAN TYPE

滇型籼杂交水稻

数量遗传分析



云南科技出版社

云南省自然科学基金资助项目

滇型杂交水稻数量遗传分析

**ANALYSIS OF QUANTITATIVE GENETICS
FOR RICE HYBRID OF DIAN TYPE**

杨 德 主编

云南科技出版社
·昆明·

图书在版编目(CIP)数据

滇型杂交水稻数量遗传分析/杨德主编. —昆明:云南科技出版社,2001.12

ISBN 7-5416-1631-1

I. 滇... II. 杨... III. 水稻, 杂种—遗传—分析—云南省 IV. S511.035.1

中国版本图书馆 CIP 数据核字(2001)第 096826 号

云南科技出版社出版发行

(昆明市环城西路 609 号云南新闻出版大楼 邮政编码:650034)

昆明市五华区教育委员会印刷厂印刷 全国新华书店经销

开本:787mm×1092mm 1/16 印张:21 字数:480 千字

2002 年 12 月第 1 版 2002 年 12 月第 1 次印刷

印数:1~1000 册 定价:40.00 元

序 言

滇型杂交水稻是我国最早研究的杂交水稻之一,有其自己的特点,重点在于粳型杂种优势利用的难度较大。现在,优势问题已经解决,“三系”早已实现配套,制种技术已过了关,尚等解决的问题还有米质、抗性等,更重要的是还有理论基础尚不够完善。多年来云南省滇型杂交稻中心及云南农业大学稻作所主要精力放在组合选育上,而在理论上相对薄弱。

令人可喜的是杨德教授在云南省自然科学基金的资助下,带领了一批年轻的学者及研究生,应用现代分子生物技术和新的遗传分析方法,对滇型杂交水稻的遗传基础作了较为深入的研究。其中活跃着一批年轻人,他们理论基础扎实,思维敏捷,闪烁着未来希望的火花。看来后继有人,令人欣慰。

本论文集的出版,标志着他们已取得阶段性成果。我很高兴为该书作序,希望他们再接再厉,更加努力工作,争取更大的成绩,结出丰硕的果实。

李铮友

2001年10月

前言

我国水稻生产历史悠久,据新石器时代遗址河南舞阳贾湖、浙江余姚河姆渡及浙江桐乡罗家角考证,中国已有 8 千多年的稻谷栽培历史。在中国,水稻乃粮食作物之首,水稻产量一般占全国粮食产量的 40% 左右。20 世纪 50 年代末的矮秆水稻和 70 年代初的杂交水稻选育推广是水稻育种的两次重大突破。

云南山地、高原、盆地、河谷相间分布,地形地貌气候复杂,“一山分四季,十里不同天”。从海拔 76 米的河口县到海拔 2700 米的宁蒗县永宁镇均有稻作分布。云南丰富的稻种资源一直为国内外水稻学家所重视。1965 年李铮友教授在云南保山水稻田中发现雄性不育株,1969 年育成粳稻红帽樱不育系,1973 年实现滇型杂交粳稻三系配套。以李铮友教授为首的云南农业大学稻作研究所选育的一系列滇型杂交水稻独具特色,在我国水稻杂种优势利用中占有重要地位。

1990 年前后,美国、英国、法国、德国、丹麦、俄罗斯、日本、韩国、澳大利亚、中国等许多国家纷纷开展人类基因组计划。水稻由于基因组较小,其基因组计划研究也迅速发展,中国籼稻基因组工作已取得可喜的进展。从生物遗传性状遗传机制来看,一般可分为受寡基因控制的质量性状和受多基因控制的数量性状两大类。植物的产量及动物的体重等一些重要的经济性状、人类的高血压等一些常见病性状都属数量性状。数量性状遗传研究由于其机制复杂而发展缓慢,一个世纪以来数量遗传学家一直采用多基因假说及数理统计方法探索生物发育结果的外部数量性状的多基因整体效应及行为。这种方法无法鉴别出个别数量性状基因(Quantitative Trait Loci, QTL)。近 10 多年来,借助主效和微效基因的混合遗传模型和分子标记技术可以进行数量性状基因作图(Mapping QTL),它不仅可估计出个别主效基因的效应,还可从细胞及分子水平上确定它们的候选基因位置,但是生物数量性状基因研究还将是一条漫长而艰巨的道路。

为了使水稻育种迈上新的台阶,目前国外一些研究机构已开始水稻超高产育种计划,国内在袁隆平院士率领下也开始了中国的水稻超高产育种计划。云南农业大学稻作研究所选育的滇型杂交组合“榆杂 29”1994 年在宾川县 1/15 公顷的小面积上创造了每公顷 16628.25 千克的世界粳稻高产记录。它的超高产、大穗特性受到育种学家的青睐。然而,滇型杂交水稻的基础研究一直滞后于育种实践。以滇型杂交水稻为研究材料,采用现代的混合遗传模型和分子标记技术,开展有关超高产、大穗特性的数量性状遗传研究,它将使滇型杂交水稻的研究深入到分子水平,为有关超高产、大穗、雄性不育等数量性状基因研究奠定基础,进而促进杂交水稻超高产遗传育种研究的发展。

本论文集收集了云南省自然科学基金重点项目《滇型杂交水稻分子数量遗传基础研究》课题组研究人员有关的学术论文。5 年多来,以中青年为主的课题组成员潜心研究,在混合遗传模型与数量性状基因定位统计方法、分子标记技术与应用、育种试验与遗传统计分析等方面获得了一些可喜的研究进展,在此奉献给同行同仁们,以供参考和借鉴。

本书及《滇型杂交水稻分子数量遗传基础研究》项目得到云南省自然科学基金重点项

目的资助,得到云南省科技厅基础研究处,云南农业大学,云南农业大学科技处、稻作研究所、农业与生物技术学院、园林园艺学院等有关部门的支持,得到徐宝明、李村生、薛启荣、刘家培、黄仁跃、刘光维、陈海如、李文祥等领导同志的关心,得到李铮友、魏蓉城、龚明、师常俊、张树华、谭学林、王荔等教授的帮助,在此一并表示感谢。

因编者水平有限,不妥之处在所难免,希望广大的同仁与专家不吝赐教。

杨 德

2001年11月11日

主 编:杨 德

副主编:陈升位 徐绍忠 毛孝强

王晓春 朱海山 谭学林

编辑委员会:

王 群 王海宁 王晓春 王朝佐 尹天水

毛孝强 张雪梅 朱海山 陈永燕 杨 德

林 谦 罗以贵 赵 雁 徐绍忠 谭学林

顾 问:李铮友 徐宝明 薛启荣

目 录

第一部分 混合遗传模型与数量性状基因定位统计方法

三倍体胚乳性状混合遗传模型和图形分析方法

..... 毛孝强 林谦 陈升位 杨德(3)

多重区间作图法在 DH 和 RIL 群体中的拓广及应用

..... 林谦 毛孝强 陈升位 杨德(39)

滇 I 型两个杂交粳稻组合主要数量性状的混合遗传分析

..... 王朝佐 杨德 张雪梅(66)

滇 I 型杂交粳稻高产组合穗部性状的混合遗传分析

..... 陈升位 毛孝强 徐绍忠 杨德(94)

烤烟农艺性状的混合遗传模型和双列杂交的遗传分析

..... 尹天水 杨德 王毅 杨跃 罗以贵(98)

第二部分 分子标记技术与应用

滇 I 型杂交粳稻保持系与恢复系着粒密度 QTLs 的 SSR 标记

..... 陈升位 杨德 张琼仙 徐绍忠(157)

滇 I 型杂交粳稻胞质雄性不育恢复性遗传及 RAPD 标记分析

..... 曹燕 赵雁 王晓春 杨德(162)

滇 I 型杂交粳稻穗部性状的混合遗传及 RAPD 标记分析

..... 陈升位 杨德 张琼仙 徐绍忠(188)

滇 I 型杂交粳稻抗稻瘟病性的 RAPD 标记筛选

..... 陈永燕 杨德 王群 张亚平(217)

非放射性 ECL 核酸标记在 RFLP 中的初步应用

..... 陈升位 曹燕 王晓春 詹开旺 杨德(228)

滇 I 型两个杂交粳稻组合保持系与恢复系的 RAPD 分析

..... 陈升位 毛孝强 王晓春 杨德 谢昌华(232)

利用 SSR 标记对滇 I 型杂交粳稻每穗总粒数和实粒数 QTL 的候选基因定位

..... 敬科举 陈升位 杨德(239)

第三部分 育种试验与遗传统计分析

SGL 设计与一阶差分近邻分析在育种试验中的应用

..... 王群 陈永燕 杨德 徐绍忠(247)

水稻品系试验中遗传方差的刀切法估计

..... 王海宁 常进荣 王建雄 徐绍忠 杨德 王群(256)

作物育种品系试验的三种近邻分析方法比较	徐绍忠 杨德 廖晓虹(261)
不完全区组设计与随机完全区组设计的比较	王群 年伟 杨德(271)
NCⅢ设计的 BASIC 语言程序及其应用	尹天水 陆俊江 陈升位 杨德(274)
滇型杂交粳稻寻杂 36 组合 DH 群体数量性状遗传分析	张雪梅 杨德 王朝佐 谭学林 王荔(283)
云南粳稻育成品种(系)主要数量性状遗传参数分析	徐绍忠 杨德 陈升位(303)
云南水稻品质性状遗传参数分析	徐绍忠 杨亚同 王海宁 杨德(309)
云南软米育成品种品质的分析与鉴定	徐绍忠 毛孝强 杨丽 杨德(315)
三个烤烟品种太空诱变 M_1 代的农艺性状变异	罗以贵 尹天水 杨跃 师跃金 杨德(321)

第一部分

混合遗传模型与数量性状
基因定位统计方法

三倍体胚乳性状混合遗传模型和图形分析方法

毛孝强¹ 林谦² 陈升位¹ 杨德³

(1 云南农业大学农学与生物技术学院, 2 云南农业大学基础与信息工程学院,

3 云南农业大学园林园艺学院, 昆明 650201)

摘要:在现代统计分析方法中,当试验数据可以看成是由两个或多个正态分布按一定的比例混合而形成的一个总体中的样本时,利用混合分布作为其数学模型,已经得到了广泛的研究和应用。本文根据三倍体胚乳性状遗传表达的特征,利用混合分布理论,结合莫惠栋(1994)提出的三倍体质量-数量性状的遗传分析,以及王建康、盖钧镒(1997)提出的二倍体的主基因-多基因混合遗传模型和图形分析方法,研究了受三倍体遗传控制的胚乳性状的主基因-多基因混合遗传模型和图形分析方法。所获得的主要结果是:

1. 提出了三倍体胚乳性状主基因-多基因混合遗传模型:

$$p = m + g + c + e$$

其中 p 为表型值, m 为群体平均数, g 为主基因效应, c 为多基因效应, e 为环境效应。 P_1 、 P_2 、 F_1 、和反交 F_1' 世代群体的分布为单一正态分布:

$$P_1: x_{1i} \sim N(\mu_1, \sigma^2)$$

$$P_2: x_{2i} \sim N(\mu_2, \sigma^2)$$

$$F_1: x_{3i} \sim N(\mu_3, \sigma^2)$$

$$F_1': x_{4i} \sim N(\mu_4, \sigma^2)$$

B_1 、 B_1' 、 B_2 、 B_2' 为两个正态分布的混合:

$$B_1: x_{5i} \sim (1/2)N(\mu_1, \sigma^2) + (1/2)N(\mu_4, \sigma^2)$$

$$B_1': x_{6i} \sim (1/2)N(\mu_1, \sigma^2) + (1/2)N(\mu_3, \sigma^2)$$

$$B_2: x_{7i} \sim (1/2)N(\mu_2, \sigma^2) + (1/2)N(\mu_3, \sigma^2)$$

$$B_2': x_{8i} \sim (1/2)N(\mu_2, \sigma^2) + (1/2)N(\mu_4, \sigma^2)$$

$F_2 (= F_2')$ 为四个正态分布的混合:

$$F_2 \text{ 或 } F_2': x_{9i} \sim (1/4)N(\mu_1, \sigma^2) + (1/4)N(\mu_2, \sigma^2) + (1/4)N(\mu_3, \sigma^2) + (1/4)N(\mu_4, \sigma^2)$$

2. 在一个主基因位点的情形下,建立了三倍体胚乳性状在各分离世代(F_2 、 B_1 、 B_1' 、 B_2 、 B_2')中的混合分布密度函数;在主基因遗传率分别为 0.90、0.80、0.70、0.60、0.50、0.40 和第一显性度 r_1 与第二显性度 r_2 分别为 0.0、0.5、1.0、1.5 的情况下,建立了三倍体胚乳性状在各分离世代中的表型分布图共 640 幅。提出了三倍体胚乳性状的主-多基因混合遗传图形分析方法为:建立数量性状在分离世代的次数分布图,根据不同世代的次数分布特征并对照附图,寻找与附图最相似的一行,从而对性状的遗传作初步的判断。主基因的对数及显隐性由图例说明可以得到,主基因和多基因在不同分离世代的遗传率由图中右上角的文字可以得到。

3. 利用混合分布理论和 AIC 信息准则,采用极大似然法估计的 EM 算法、分离分析法和适合性测验,给出了利用单个世代(以 F_2 世代为例)进行主基因-多基因混合遗传分析和估计遗传参数的方法。根据不同成分分布数下对混合群体的分解结果和 AIC 准则可以判断主基因的存在,进一步可以确定主基因的显隐性、遗传效应、主基因遗传率等参数。

4. 采用 Visual Basic 语言,开发出了 Windows 环境下的可视化混合遗传分析软件。以 F_2 世代的 Monte Carlo 模拟数据分析为例,演示了利用该软件进行混合遗传分析的基本方法。

关键词:三倍体胚乳性状 混合遗传模型 图形分析 遗传参数 极大似然法估计 EM 算法

1 前言

1.1 混合模型的发展和研究概况

在现代统计分析方法中,当试验数据可以看成是由两个或多个正态分布按一定的比例混合而形成的一个总体中的样本时,利用混合分布(Mixtures of distributions),或是有限正态混合分布(Finite mixture of normal distributions),作为其数学模型(混合模型 Mixture model)进行研究和应用^[1,2]。对混合分布的研究最早可追溯到 Newcomb(1886)^[3]。但是,要对一个有限混合分布所包含的成分进行分解,是一件十分困难的事情。Pearson(1894)提出一个矩法估计的分解法,试图分解由两个单变量不等方差混合而成的正态分布,这要求解一个九次方程^[2];Cohen(1967)提出了一个迭代过程,这一过程包含求解有唯一负根的一个三次方程,从而简化了 Pearson(1894)的方法;但是 Tan 和 Chang(1972)以及 Fryer 和 Robertson(1972)证明了对于混合问题,矩法估计不如极大似然法估计精确。Fowlkes(1979)在利用测绘诊断程序对混合分布成分的检测研究中指出,在 20 世纪前后,混合分布的分解主要采用图形技术(graphical techniques)。这一技术的开拓者有 Harding(1948)、Cassie(1954)、Bhattacharya(1967)以及 Wilk 和 Gnanadesikan(1968)等人。Tarter 和 Silver(1975)根据双变量正态密度函数的性质,提出了一个图形分析程序,Chhikara 和 Register(1979)基于计算机辅助方法发展了一项数值分类技术^[3]。

Rao(1948)^[4]首次利用极大似然法估计的原理来对混合分布的成分进行分解,他将 Fisher 得分法(Fisher's method of scoring)用于两个等方差正态分布的混合。随着计算机的出现,人们对混合分布研究的注意力开始转向对分布参数的极大似然法估计上来。Butler(1986)^[5]指出,早在 Pearson(1894)的矩法估计之前,Newcomb(1886)就曾试图利用一种“权重轮回迭代计算方法”的思想来分解混合分布的成分;Jeffreys(1932)首先利用一种最基本的迭代计算方法,对由两个等方差和比例均已知的正态分布的混合,估计出各成分分布的平均数。这就是现在称为 EM(expectation and maximization)算法(Dempster, Laird 和 Rubin 1977)^[6]的雏形。但是直到 Hasselblad(1966,1969)将极大似然法估计用于具有相同方差的 k 个正态分布的混合,随后又应用于来自指数分布族的混合,极大似然法估计才得到进一步的深入研究。Wolfe(1967,1970,1971),Day(1969)研究了多元分布的混合问题,但 Day(1969)集中于对两个具有相同协方差阵的正态总体混合的求解问题。Wolfe 则主要解决了任意个数的正态异质总体以及多元贝努里(Bernouli)分布的混合问题。Dick 和 Bowden

(1973)研究了两个正态分布混合的极大似然法估计(MLE)^[7];Hosmer(1973)研究了小样本两个正态分布混合的MLE问题,并比较了三种不同类型的抽样下正态混合分布的MLE。随着Dempster, Laird和Rubin(1977)提出和发展了EM算法,混合分布似然解的收敛性质有了理论基础,极大似然法估计用于混合模型的拟合得到广泛的应用。Hosmer(1978)研究了正态分布的混合在分类问题上的应用。Boyle和Elston(1979)将混合遗传模型推广到子代基因型的条件分布依赖于参数(称为转移概率)的情形,这些参数包括亲本的主基因基因型、表型值等。Elston(1983)将Elston和Stewart(1973)的模型推广到两个位点的假设检验问题,通过数据变换去掉了正态性假定,并考虑了不同世代间环境效应的相关性。Shoukri和McLachlan(1994)将混合遗传模型用于人类家系血管数据的遗传分析^[8],发现在舒张压和收缩压的变异中,均存在有一个主基因位点,主基因变异可解释总变异的73%和42%。Elston(1984),Hoeschele(1988),Goffinet et al(1990),Knott(1991),Fernando et al(1994)和Shoukri(1994)等考虑到家系内个体间的不独立性、亲子之间的相关以及性状分布的非正态性,推广了混合遗传模型,这些方法广泛应用于人类遗传流行病的研究。Titterton(1985)总结了混合分布在各个方面的应用^[3],并给出每个实例中所采用的分布类型、数据结构和参数估计方法等,应用的领域包括渔业研究、医学、心理学、古生物学、沉积岩学、地理学、植物学、农业科学、动物学等。

在实际问题中,凡数据结构具有以下特征的均可利用混合模型^[1,9]:数据来自一组分布总体的混合,但组成这一混合总体的单个总体以及单个总体占混合总体的比例是未知的。比较典型的实例有:(1)渔业研究中,鱼的体长容易度量,但其性别和年龄不易确定;(2)沉积岩学中,如何从一个颗粒大小的沙样的分布来判定该样本中含有多少不同种类的矿石;(3)医学诊断中,可对一组病人作临床检查,但他们所患疾病属于同一类型或是不同类型是未知的;(4)在RFLP-QTL连锁分析中,每个标记基因型是由三种不同的QTL基因型混合而成,个体的QTL基因型不能确定^[10,11,12]; (5)在主基因-多基因混合遗传模型研究中,当性状在分离世代 F_2 中表现出多个分布混合的特征,需要从中确定主基因的存在以及将主基因的遗传效应从总遗传效应中分离出来^[13,14,15,16]; (6)在一些可度量抗性性状的遗传分析中,利用对混合分布的分解结果代替人为的对性状划分分类标准,可避免 χ^2 测验的误用,如有些研究者将病虫害的危害率简单地分成两组或几组,再假设该性状受几对基因控制,然后通过 χ^2 测验检验适合性。在杂交稻的育性上,简单地将结实率分为50%以上和50%以下两组,再假定育性受一对或两对主基因控制,然后测验适合性,这实际上是人为的先定论后测验。利用混合模型理论可以克服这种先定论后测验的矛盾现象^[13]。

以上这些数据的统计分析并不十分容易,主要因为^[1,9]:(1)没有一个简单的公式可以表示各种统计参数的估计值,因此需要采用数值分析的方法;(2)混合分布分解的复杂性使得在计算机出现以前,人们只能处理一些最简单的混合问题;(3)从统计分析的某些方面提出的理论上的难题,如混合分布分解的唯一性问题、似然比统计量的性质等。尽管如此,混合模型仍在许多方面得到了广泛的应用^[16]。

1.2 混合模型在植物数量性状(二倍体性状)遗传体系检测研究的发展

关于数量性状遗传的机制,早期认为是受微效多基因(简称多基因, polygene)控制的,

并常假定各位点有相同的遗传效应,后来发现也有受主效基因(简称主基因, major gene)控制的,如矮秆突变体,在一个时期内,认为非此即彼,而对主基因效应和多基因效应大小的相对性、多基因各位点效应的特异性缺乏认识。

多基因假说在解释动物的产奶量和作物产量等性状的遗传上是合适的。但是,近年来,随着现代生物技术和实验手段日新月异的发展,大量的研究,不论是 QTL(quantitative trait loci)作图的结果或是利用新的统计工具对已有试验数据的重新分析,都发现对许多数量性状来说,控制性状遗传的基因在效应大小上有很大的差异,效应较大的基因可以表现出主基因的特性,遗传效应相对较小的基因表现为微效多基因,性状的遗传表现为主基因-多基因混合遗传的模式^[1,12,17,18,19]。

控制质量性状与控制数量性状基因间的区别在于它们所产生的相对于总变异幅度的表型效应的大小上,对于遗传效应较大的一个基因,即使在其它位点也存在分离的情况下,也能产生可识别的不连续性,那么这个基因可以用质量性状方法进行研究;然而,对于具有较小效应、不能产生表型不连续变异的基因,就无法单个地进行研究,用主效基因(主基因)和微效多基因(多基因)来表示具有这种区别的两类基因。在实际问题中还可能存在这种情况:作为一因多效的一个结果,同一个基因对某一个性状可划分为主基因,而对另一性状却表现为多基因^[9]。多基因和主基因并没有本质上的区别,也没有证据表明这两类基因具有不同的遗传性质。

利用主基因作标记,可以了解到连续变异的决定因子也同样位于染色体上,Breese 和 Mather 将果蝇Ⅲ染色体分成六个片段,结果发现每个片段对刚毛数的差异都起作用;Mather 和 Harrison(1949)对选系的染色体分析,又表明Ⅱ染色体上至少有两个多基因,Ⅲ染色体上至少有三个多基因,X染色体上至少有三个多基因,三条染色体上至少存在八个多基因才能解释选系表现出的变异;Wigan, Cooke 和 Mather, Thodat 及其同事们关于果蝇胸侧板刚毛数的研究也有类似的结果;Davies(1971)在一套试验材料中发现影响胸侧板刚毛数的位点至少有 15 个,有 14 或 15 个位点影响腹部刚毛数。

多基因假说为一些连续性状的变异提供了一个恰当的解释。多基因和主基因一样都位于染色体上,是染色体的组成部分,在世代间传递服从相同的遗传规律;但是,当从其它方面去比较主基因和多基因时,会发现多基因效应微小,相似且互补,而主基因有较大的效应;在主基因和多基因的遗传分析中常采用截然不同的两种统计方法,但是主基因和多基因的区别不应与统计方法的区别混为一谈。有时似乎是这样:产生表型分布不连续的基因,实际上也可用它们产生的方差和协方差来处理;同样,一个作用相当小且非专一的基因,若能通过适当的遗传交配设计排除其它基因分离的干扰,精确地控制环境,尽可能地减少非遗传变异,那么也可用孟德尔方法研究它的遗传规律^[20]。

数量遗传学研究对性状变异起作用的全部基因,即使对性状变异影响最小的基因也和影响最大的基因一样进行考虑,这是因为性状的变异是作为一个整体对待的^[21]。孟德尔方法只强调主基因的差异,由于主基因的效应大且相对稳定不变,使得它们在遗传研究中容易识别和追踪;在基因作用方面,早已公认显性和非等位基因互作是主基因的普遍特性(Bateson, 1909),同样,用生物统计学的方法进行的第一个关于数量变异的遗传分析也表明,控制这种变异的基因至少表现部分显性(Fisher, 1918)。而且此后的无数生物统计

学的研究已经证实这是此类基因的一个共同特性,还证明非等位基因互作也是多基因系统的一个普遍特性^[22]。因此,显性和非等位基因互作都是多基因的固有特性,犹如它们在主基因中一样,但是有一点差别是:主基因的显性和互作在绝大多数情况下是完全的,而数量性状变异中的显性和互作通常似乎是不完全的^[21]。

遗传学上,对主基因的研究朝着基因定位、克隆、转移的方向发展,最终企图利用生物技术的手段获得转基因动植物;对多基因的研究主要通过遗传交配设计和遗传参数估计了解性状的遗传规律,最终指导常规育种实践^[11,16]。这两个方面的研究代表了自 1900 年孟德尔的论文被重新发现之后遗传学在微观和宏观两个方面的发展,微观方面发展的大致阶段是孟德尔遗传学(家系水平)、摩尔根遗传学(细胞水平)、分子遗传学(DNA 水平)和基因工程,宏观方面发展的大致过程是孟德尔遗传学、群体遗传学和数量遗传学,目前这两方面的发展有结合的趋势。

对数量性状基因个数和单个基因位点遗传特征的了解一直是数量遗传学家想要达到的目标。早期,由于无法对单个基因进行识别和追踪,多基因假说将控制数量性状的基因全部归类为多基因,在遗传分析中,将所有的基因作为一个整体来考虑基因效应的大小。近十多年来,随着分子生物学的发展,出现了许多分子水平上的多态性标记,利用这些标记构建出的饱和遗传连锁图谱使得了解单个数量性状基因的遗传成为可能^[10],许多分子遗传学家、生物统计学家和数量遗传学家纷纷在这一领域开展研究,利用分子图谱分析 QTL 的各种理论、模型和方法不断提出^[23,24,25,26],为数量遗传研究注入了新的活力,成为数量遗传研究的一个热点。

大量的试验数据和 QTL 作图结果表明:数量性状基因在遗传效应上有很大差异,动植物育种过程中也早已发现,一些数量性状既受少数主基因的控制,同时又有多基因的修饰作用,控制数量性状的基因体系中既有遗传效应较大的主效基因,又有遗传效应较小的微效多基因。称这种遗传现象为主基因-多基因混合遗传或主-多基因混合遗传(mixed major gene and polygene inheritance or major-polygene mixed inheritance)^[9,27,28]。构建饱和分子标记连锁图谱需要很大的花费和很长的时间,因此在许多物种中,利用适当的统计方法鉴定一些具有较大遗传效应的主基因以用于指导育种实践仍是十分有用的^[29,30]。目前看来,数量性状由主基因和多基因控制的混合遗传模型可能具有相当的普遍性,因而正受到数量遗传学家的重视而形成目前研究的热点^[31]。

主基因-多基因混合遗传模型首先在人类和动物遗传研究中提出并得到发展。早在 1971 和 1973 年 Elston 和 Stewart 就提出“一个主基因和多基因”的遗传模型并用于人类系谱数据的遗传分析,Elston 和 Stewart 认为两个近交系(近等基因系)间,表型差异遗传机制的发现常常由研究亲本、 F_1 和回交世代的表型分布开始,因此描述了一系列数据来自这些世代的遗传模型。在这些模型中,都假定对于一个特定的基因型来说,个体的表现型服从单一的正态分布,对所有基因型来说,有相同的环境方差。利用极大似然法估计^[2,5,32,33]获得模型参数的估计值,然后选择最合适的遗传模型,并将这一方法用于老鼠一组生理性状的遗传分析。模型中的正态性(normality)和方差同质性(homoscedasticity)的假定可以通过选用非参数方法来排除,但是非参数方法只可用于检验一组数据是否与单基因位点假设相符合,如果需要检验两个基因位点假设的适合性,那么对表型分布的形式

还必须做一些假定。Morton 和 MacLean(1974)进一步发展了这一模型,把这种遗传模型称为混合模型(mixed model),并解决了模型中参数的计算问题;Elston(1984)、Hoeschele(1988)、Goffinet et al(1990)、Knott(1991)、Fernando et al(1994)和 Shoukri(1994)等^[8]考虑到家系内个体间的不独立性、亲子之间的相关以及性状分布的非正态性,大大推广了混合遗传模型,在参数估计上均采用极大似然法估计,计算方法采用 EM 算法或改进的 EM 算法^[2,6,7,32],这些方法广泛应用于人类遗传流行病的研究。Elkind 和 Cahaner(1986,1990)提出单基因-多基因(single gene-polygene, SG-PG)遗传模型^[28],其单基因是可鉴定的,即后代中个体的基因型可以明确分类,作者通过两项因子试验设计,不仅估计了单基因和多基因的遗传效应,而且还估计出互作效应的大小和方差。SG-PG 模型仅适用于某个已知的主基因对另一个数量性状也产生效应的情形。一般情况下,主基因即使存在,人们对它的了解也不多,更不可能对后代个体的主基因基因型明确的分类,因此这一模型应用起来有很大的局限性。

莫惠栋分析了一对主基因存在时,主基因-多基因混合遗传性状在各个世代的遗传组成以及遗传参数的估计问题,并把这类性状称为质量-数量性状^[34,35,36],由于 F_2 代的分组趋势不明显,盖钧镒等建议采用后裔测验的方法,然后通过聚类分析确定 F_2 个体的主基因基因型^[16,37,38]。这种方法只有当主基因效应充分大时,才能将 F_2 确切分类,不便处理主基因效应较小,主基因表现为显性或超显性以及主基因对数多于一对的情形,对主基因的加、显(隐)性无法检验。Loisel(1994)研究了在 F_2 世代中探测主基因存在的似然比统计量的渐近性质,由此可对主基因是否存在,主基因是否表现为加性或完全显性进行检验。姜长鉴将 Loisel 的结果用于大麦矮秆突变体与正常秆品系间杂交产生的 F_2 代株高性状的遗传分析^[39]。

王建康和盖钧镒(1997)提出利用混合分布理论在 F_2 群体中鉴定数量性状主基因的存在和估计遗传参数的方法,系统地提出了二倍体的主基因-多基因混合遗传模型和图形分析方法^[13],极大地丰富和完善了二倍体混合遗传模型的数量遗传理论。该方法可以直观地判断一个数量性状的遗传是主基因遗传、多基因遗传、或是主-多基因混合遗传,并能估计主基因对数、主基因的遗传效应、主基因遗传方差和遗传率、多基因效应、多基因遗传方差和遗传率等相应的遗传参数。这些结果虽然可以检验主基因的加显性并估计主基因的加显性效应和主基因的遗传方差,但 F_2 世代受环境影响较大,检测主基因的功效不高,对多基因的存在无法作进一步鉴定^[11,16]。盖钧镒等认为主基因-多基因混合遗传模型为植物数量性状的通用性模型,单纯主基因模型或单纯多基因模型为其特例,由此发展了适合植物数量性状 QTL 体系检测的遗传试验方法和统计分析方法,包括一对主基因-多基因(简称主-多基因)混合遗传模型单世代和多世代联合的分离分析方法^[11,16,23]。多世代的联合分析方法可以对多基因的存在进行鉴定,并且可以判断多基因是否满足加-显性模型,并进一步估计多基因的遗传效应。章元明等^[37]已拓展了两对主基因-多基因混合遗传模型的个别世代和多世代联合的分离分析方法。

在实际应用方面,张雪梅等(1999)^[41]用榆杂 29、寻杂 36 的 DH 群体对穗长、颖花数、结实率的遗传率进行了估计;陈升位(1999)^[42]、曹燕等(1999)^[43]运用主基因-多基因混合遗传模型中 F_2 世代单世代图形分析方法,对滇 I 型杂交水稻榆杂 29、寻杂 36 F_2 群体的