

遗传标记与数量性状基因间连锁关系的分析^①

姜长鉴 顾兴友

(江苏农学院农学系, 扬州 225009)

摘要 本文讨论标记基因与数量性状主基因连锁关系的一般分析方法, 包括重组值的估计和有关遗传假设的测验。并以我们水稻遗传试验中两个具有互补和重叠作用的卷叶基因和一个矮秆基因试验结果的分析为例作了较详细的示范。

关键词 基因连锁, 重组值, 数量性状基因, 遗传标记

Linkage Analysis between Genetic Markers and a Major Gene for a Quantitative Trait

Jiang Changjian Gu Xingyou

(Department of Agronomy, Jiangsu Agricultural College, Yangzhou 225009)

遗传标记基因和质量性状基因间的连锁分析通常以多项分布函数为基础, 重组值的极大似然估计在很多情况下有明确解⁽¹⁾。遗传标记基因和数量性状主基因的每一基因型组合, 由于微效基因和环境条件的作用, 在分离世代呈现(或经转换后近似于)连续的正态分布, 不同基因型组合则呈具有不同平均数和相同方差的多个正态分布的混合^(4-6,9), 相应连锁关系的分析必须以混合模型为基础, 借助数值方法求解。本文详细讨论混合模型重组值估计和有关假设测验的原理和方法, 并以我们水稻遗传试验中一组资料的分析作示范说明其应用。

1 统计模型及分析方法

设一杂交组合的分离群体包括标记性状的分离, 标记性状由单个(或几个)标记基因决定。同时有一个数量性状, 以 y 表示, 表现有主基因分离。设 p_{ik} 为标记性状基因与数量性状主基因型的分离比, $i=1, \dots, a$, $k=1, \dots, b$, i 表示标记性状的表现型, k 表示数量性状主基因的表现型, 如 F_2 群体 $k=2$ 或 3 相应于一个位点的完全或不完全显性模型。标记性状基因与数量性状主基因连锁关系的统计分析即为对分离比 p_{ik} 的分析。设微基因和环境效应服从正态分布, 分离群体标记性状与数量性状的联合分布为混合分布^(5,6,9), 其似然函数:

$$f(y_j) = \sum_i \sum_k p_{ik} f_k(y_j), \quad j=1, \dots, n \quad (1)$$

n 为样本容量, f_k 表示平均数为 μ_k 、方差为 σ_k^2 的正态分布。

混合模型的分析必须采用似然函数法。

1.1 分离比的估计

设个体 j 属于标记性状的第 i 类型, 数量性状的主基因型虽然不能由观察值 y 确定, 但其属于第 k 类型的可能性以条件概率估计为:

$$w_{ikj} = p_{ik} f_k(y_j) / \sum_i \sum_k p_{ik} f_k(y_j) \quad (2)$$

^①国家教委优秀青年教师基金资助项目。

由此可估计分离比:

$$\hat{p}_{ik} = \sum_{j=1}^n w_{ikj} / n \quad (3)$$

但(2)、(3)的计算依赖于 p_{ik} 本身以及其他参数, 分离比估计必须采用 EM 循环迭代法^[3], EM 的每一循环分为两步。第一步由 p_{ik} 、 μ_k 、 σ_k^2 的一组(可由目测给出的)初始值, 代入(2), 计得 w_{ikj} 。第二步由(3)式计算 \hat{p}_{ik} , 同时估计其他参数:

$$\hat{\mu}_k = \sum_{i,j} w_{ikj} y_j / \sum_{i,j} w_{ikj} \quad (4)$$

$$\hat{\sigma}_k^2 = \sum_{i,j} w_{ikj} (y_j - \hat{\mu}_k)^2 / \sum_{i,j} w_{ikj} \quad (5)$$

以(3)~(5)的结果作为相应的参数值进入下一循环。理论研究结果已经证明, 以此循环迭代, 各参数的估计值将逐步收敛。收敛时的值即为参数的极大似然估计值^[3]。

1.2 标记性状基因与数量性状主基因的独立性测验

标记基因与数量性状主基因独立分离的统计假设可表示为:

$$H_0: p_{ik} = p_{i.} p_{.k}, \quad i = 1, \dots, a, \quad k = 1, \dots, b, \quad H_1: \text{不全等} \quad (6)$$

$p_{i.}$ 为标记性状的观察比, $p_{.k}$ 为数量性状主基因型的分离比。(6)式的假设与质量性状的有关假设完全相同, 但其测验必须采用似然比法。

似然比测验首先根据 H_0 和 H_1 分别估计参数, 与 H_1 相应的参数由(3)~(5)迭代估计。 H_0 涉及数量性状基因的合并分离比, 迭代估计为:

$$\hat{p}_{.k} = \sum_{i,j} w_{ikj} / n \quad (7)$$

循环迭代收敛后, 将估计值代入似然函数, 计算似然值:

$$L = \sum_j \ln \hat{f}(y_j) \quad (8)$$

设与 H_0 和 H_1 相应的似然值分别为 L_0 和 L_1 , $2(L_1 - L_0)$ 即为似然比, 用于对(6)的 χ^2 测验, 自由度为 $(a-1)(b-1)$ 。

1.3 标记基因与数量性状主基因间重组值的估计

否定假设(6)的 H_0 , 表明基因间可能连锁, 但连锁关系的确定必须估计重组值 r , 并测验假设:

$$H_0: p_{ik} = p_{i.} p_{k/i}(r), \quad i = 1, \dots, a, \quad k = 1, \dots, b, \quad H_1: \text{不全等} \quad (9)$$

$p_{i.}$ 仍为标记性状的观察比, $p_{k/i}(r)$ 表示标记性状第 i 类表现型中数量性状主基因的分离比, 为标记基因与数量性状基因间交换率 r 的函数。最常见的形式为两个连锁基因的理论比, 两类配子组成三类个体: 非重组体、单重组体、双重组体, 相对比例为 $(1-r)^2 : 2r(1-r) : r^2$ (Jensen^[4], Luo 和 Kearsey^[5], Luo 和 Woolliams^[6])。当存在完全显隐性时, 相应的比例项合并。(9)式重组值 r 的估计须在 EM 循环中和其他参数同时进行。

如果第 ik 组的次数 $\sum w_{ikj}$ 可由观察得到, 其抽样变异则服从多项分布, 似然函数有对数式:

$$Q(r) = K + \sum_{i,k,j} w_{ikj} \ln(p_{i.} p_{k/i}(r)) \quad (10)$$

K 表示与组合数有关的常数项。将(2)式的 w 代入(10), 可用极大似然法估计 r 。莫惠栋曾求得两个基因均为完全显隐性时 r 的极大似然估计^[1], 但 r 的估计一般无明确解, 同样必须借助数值方法(如黄金切割法^[7])。因

此, (9) 式 H_0 中 r 的极大似然估计仍采用 EM 循环迭代法, 每次迭代, 在估计其他参数的同时据 (10) 估计 r . H_1 的参数估计不变, 假设 (9) 似然比测验的自由度为 $[a(b-1)-1]$.

1.4 标记性状基因与数量性状主基因分离比的联合测验

标记性状与数量性状主基因分离比的联合测验, 相当于假设 (6) 和 (9) 式 H_0 的 p_{i_1} 为孟德尔的理论比, 似然比测验的自由度将分别增加到 $b(a-1)$ 和 $ab-2$. 此外还可根据标记性状的遗传组成, 对不同表现型内数量性状主基因的分离比逐一测验. 方法类似, 只要根据 H_0 的内容将 (7) 改变为部分组的合并估计, 同时对似然比测验的自由度作相应变化. 因此, 不管标记性状的基因型与表现型间的关系如何复杂, 与数量性状主基因的连锁关系均可照此分析.

2 实例分析

为了明确水稻品种流岗卷叶梗所带的卷叶基因与卷叶标记基因系 I89020 所带卷叶基因的等位关系, 将两品种杂交, F_2 代考察卷叶性状. 因伴有矮秆基因的分, 于成熟前同时考察株高. 两个性状 1824 株的分组资料列于表 1. I89020 所带卷叶基因已定名为 RI_3-rI_3 , 位于第 12 染色体上, 表现完全隐性. 有关组合的研究结果初步表明, 流岗卷叶梗带有一个不等位的卷叶基因, 表现不完全显性, 暂记为 RL_1-rl_1 . 初步确定的卷叶基因型与表现型的关系列于表 2. 现分析两个卷叶基因与矮秆基因的连锁关系, 进一步证实两个卷叶基因间的关系, 同时估计卷叶基因与矮秆基因间的重组值.

表 1 水稻流岗卷叶梗(LGJ)×I89020 F_2 群体卷叶和株高的次数分布表¹⁾

株 高 (cm)	次 数	卷 叶 分 组		
		平 展 叶	中 度 卷 叶	高 度 卷 叶
55~	26	1	1	24
65~	108	4	1	103
75~	251	3	9	239
85~	94	3	9	82
95~	71	2	32	37
105~	239	29	93	117
115~	504	110	250	144
125~	444	165	244	35
135~	85	53	30	2
145~	2	2	0	0
合 计	1824	372	669	783

1): 双亲株高分别为 116 cm 和 79 cm.

由 (2)~(5) 估计与 (9) 式的 H_1 相应的分离比以及不同矮秆基因型的株高和方差, 结果列于表 2 每格的上一行. 设 (9) 式的 H_0 表示两个卷叶基因相互独立, RI_3-rI_3 与矮秆基因 $D-d$ 连锁, 具重组值 r , 据此可导出分离比 p_{i_k} 与 r 的相互关系, 亦列于表 2, 由此可确定函数 (10). 由 (2)、(10)、(4)、(5) 求出与 (9) 式的 H_0 相应的参数估计值, 列于表 2 每格的下一行. 将两组估计值分别代入 (8), 并计算得似然比为 9.20, 用于自由度为 4 的 χ^2 测验, 结果不显著, 但与显著标准相差不大. 以下对表 1 资料进一步作 4 个独立的测验, 包括两个卷叶基因间, 以及与矮秆基因间的连锁关系的测验. 各个测验的统计假设以及测验结果均列于表 3. 由表 3 的测验结果可以看出, 联合测验的似然比偏大主要由矮秆基因的分离比与一对基因分离的理论比的差异造成, RI_1-rl_1 与 $D-d$ 间相互独立, 而 RI_3-rI_3 与 $D-d$ 间紧密连锁是毫无疑问的.

分析结果进一步证实, 流岗卷叶梗的卷叶基因独立于 I89020 的矮秆基因, I89020 的卷叶基因与矮秆基因紧密连锁。两品种的基因型可分别表示为 $Rl_1Rl_1Rl_3Rl_3DD$ 和 $rl_1rl_1rl_3rl_3dd$ 。 $Rl_3 - rl_3$ 和 $D - d$ 间的重组值估计为 0.0345。应用 Jackknife^[8] 法估计 r 估计值的偏差和标准误, 矫正的估计值为 0.0324, 标准误为 0.00048。 $Rl_3 - rl_3$ 与 $D - d$ 早期认为是同一基因, 它们的表现型为一因多效, 后确定为紧密连锁, 矮秆基因编号为 d_{33} , 但重组值的估计仍未见报道^[2], 可能是由于株高为连续分布, 分离群体中矮秆基因型分组不明确, 常规方法无法估计其交换率。

表 2 水稻流岗卷叶梗 × I89020 F_2 群体卷叶和株高性状的遗传组成、分离比与重组值的关系
以及与假设(9)的 H_1 (上)和 H_0 (下)相应的参数估计值

卷叶分组及其 基因型	株高分组及其基因型				行总和($p_{i.}$)
	矮秆(dd)		高秆(D_{-})		
	理论比(p_{11})	估计比(\hat{p}_{11})	理论比(p_{22})	估计比(\hat{p}_{22})	
平展叶 $rl_1rl_1Rl_3-$	$\frac{r(2-r)}{16}$	0.0059 0.0042	$\frac{(3-2r+r^2)}{16}$	0.1981 0.1833	0.2040 0.1875
中度卷叶 $Rl_1rl_1Rl_3-$	$\frac{r(2-r)}{8}$	0.0099 0.0085	$\frac{(3-2r+r^2)}{8}$	0.3569 0.3665	0.3668 0.3750
高度卷叶 $Rl_1Rl_1Rl_3- - -rl_3rl_3$	$\frac{(4-6r+3r^2)}{16}$	0.2556 0.2373	$\frac{(3+6r-3r^2)}{16}$	0.1737 0.2002	0.4293 0.4375
列总和 $p_{.j}$	$\frac{1}{4}$	0.2714	$\frac{3}{4}$	0.7286	1.0000
株高平均和方差 μ and σ^2	83.65(67.37) 83.48(64.32)		126.33(83.92) 126.25(85.37)		

表 3 对表 1 资料提出的遗传和统计假设及其测验结果

遗传假设	统计假设 H_0	测验结果
$Rl_1 - rl_1$ 和 $Rl_3 - rl_3$ 独立分离	$(p_{11} + p_{12}) : (p_{21} + p_{22}) : (p_{31} + p_{32}) = 3 : 6 : 7$	$\chi^2 = 3.24$
$D - d$ 自由组合	$(p_{11} + p_{21} + p_{31}) : (p_{12} + p_{22} + p_{32}) = 1 : 3$	$\chi^2 = 3.22$
$Rl_1 - rl_1$ 和 $D - d$ 独立分离	$p_{11} : p_{12} = p_{21} : p_{22}$	$\chi^2 = 0.05$
$Rl_3 - rl_3$ 和 $D - d$ 独立分离	$(p_{11} + p_{21}) : (p_{12} + p_{22}) = p_{31} : p_{32}$	$\chi^2 = 780.75$

参 考 文 献

- (1) 莫惠栋, 1984. 遗传, 6(5): 42—48.
- (2) 松尾孝岭等, 1990. 稻学大成, (第三卷, 遗传篇), 东京: 农文协, 167, 210.
- (3) Dempster A P, Laird N M, Rubin D B, 1977. J. R. Statist. Soc. B., 39: 1—22.
- (4) Jensen J, 1989. Theor. Appl. Genet., 78: 613—618.
- (5) Luo Z W, Kearsey M J, 1989. Heredity, 63: 401—408.
- (6) Luo Z W, Woolliams J A, 1993. Heredity, 70: 245—253.
- (7) Press W H *et al*, 1986. Numerical Recipes. Cambridge University Press, Cambridge.
- (8) Weir B S, 1990. Genetic Data Analysis. Sinauer, Sunderland, Mass, pp. 137.
- (9) Weller J I, 1986. Biometrics, 42: 627—640.

本文于 1994 年 9 月 6 日收到。