

质量-数量性状遗传参数估计的 P_1 、 P_2 、DH 联合分析方法^x

胡中立¹ 章志宏¹ 章元明²

(¹武汉大学生命科学院, 湖北武汉, 430072; ²四川畜牧兽医学院动物科学系, 四川荣昌, 632460)

提 要 本文提出利用亲本 P_1 、 P_2 及其 DH 三个群体联合分析包括两个位点主基因控制的质量-数量性状遗传的统计方法, 共建立了五类共22个遗传模型, 进而, 可采用 AIC 信息准则选择最适模型, 并通过适合性检验对所选择的遗传模型做进一步的检验。

关键词 质量-数量性状; 两位点主基因; DH 群体; 多世代联合分析

The Joint Analysis of P_1 , P_2 , DH Population about Estimation of Genetic Parameters for Qualitative-Quantitative Traits

HU Zhongli¹ ZHANG ZhiHong¹ ZHANG YuanMing²

(¹ College of Life Science, Wuhan University, Wuhan, 430072; ² Sichuan Institute of Animal Husbandry and Veterinary Medicine, Rongchang, 632460)

Abstract The statistical method for qualitative-quantitative traits controlled by two major gene loci by using joint analysis of P_1 , P_2 , DH population was proposed in this paper. Twenty-two genetic models were established, which could be classified into five types. The most suitable genetic model could be selected by using Akaike's Information Criterion and could be further tested by using a set of tests of fitness.

Key words Qualitative-quantitative trait; Two major gene loci; DH population; Joint analysis of multiple generations

质量-数量性状(QQT)是一类由主(效)基因和微(效)基因共同控制的数量性状, 主基因效应较大, 微基因效应较小, 二者甚至可能有很大的差异。在文献[1, 2]中, 王建康和盖钧镒采用混合分布模型, 给出了两组群体中质量-数量性状遗传研究的联合分析方法, 第一组群体为 P_1 、 P_2 、 F_1 、 F_2 、 F_{2+3} 共5个群体, 第二组群体为 P_1 、 P_2 、 F_1 、 F_2 、 B_1 、 B_2 共6个群体, 但在其分析过程中均只涉及到一个主基因和微基因共同控制的质量-数量性状。在文献^[3, 4]中, 王建康、盖钧镒, 以及在文献^[5, 6]中, 胡中立、章志宏分别探讨了利用个别分离世代群体进行 QQT 遗传分析的方法, 尤其是文献^[5, 6]中包含对两对主基因和微基因共同控制的 QQT 的遗传分析, 所论分离群体分别为 F_2 、BC、DH 或 RL。本文特别将上述分析原理与方法拓展到综合利用 P_1 、 P_2 、DH (或 RL) 共3个世代群体的联合分析。

x 湖北省自然科学基金资助课题。

收稿日期: 1999206210, 接受日期: 1999212222

1 材料与方法

1.1 遗传材料的构成与遗传模型的基本假定

本文所用的遗传材料包括两个纯系亲本(分别记为 P_1 、 P_2)，以及由二者的杂交后代 F_1 或 F_2 花培植株经加倍而成的双单倍体群体(DH)共3个群体。或者第3个群体也可以是由双亲的杂交后代 F_2 或 S_2 群体内个体连续自交或兄妹交形成的家系即重组近交系群体(RL)。

3个群体中个体的表型值分别用 X_{1k} 、 X_{2k} 、 X_{3k} 表示，样本含量分别用 n_1 、 n_2 、 n_3 表示，并特别用 n 表示每个DH(或RL)家系的观测个体数，模型中的一些基本假定是^[1]：所考虑的性状是二倍体核遗传；不存在细胞质或母体效应；所有配子具有相同的生活力；主基因和多基因之间不存在互作和连锁(而对主基因之间、微基因之间有无互作与连锁没有限制)；分离世代DH或RL群体中个体的微基因效应是正态随机变量。

本文所考虑的遗传模型共有5类：一对主基因遗传模型、二对主基因遗传模型、微(效多)基因遗传模型、一对主基因+微基因遗传模型、二对主基因+微基因遗传模型(见表1)。

1.2 主基因作用方式

当所论性状由两个主基因共同控制时，王建康和胡中立等分别对其效应(作用方式)的描述方式是各不相同的，前者采用的是经典数量遗传学的加性2显性2上位性遗传模型(王建康等, 1998; Gai等, 1998)，后者采用的是普通遗传学的概念(术语)，例如协调加性、累加作用、显性上位、隐性上位、重叠作用、互补作用、抑制作用等。在本文中我们仍沿用后者。

1.3 多世代联合分析的基本原理与过程

王建康等(1998)提出的多世代联合分析，其分析内容和过程包括：遗传模型的构造；模型参数的估计(即极大似然估计的计算)；遗传模型的选择与检验(分别采用AIC准则、似然比检验、适合性检验)；遗传参数的估计。本研究中我们仍据此思路开展有关的分析。

2 结果及说明

2.1 关于遗传模型的构造

限于篇幅，此处仅选较复杂的“二对主基因+微基因混合遗传模型”为例进行简要的说明。此时， P_1 、 P_2 群体的分布为单一正态分布，DH群体是4个正态分布的混合，但混合分布中所包含的成份分布与 P_1 、 P_2 的分布不同。记

$$P_1: X_{1r} \sim N(L_1, R^2)$$

$$P_2: X_{2r} \sim N(L_2, R^2)$$

$$DH: X_{3r} \sim \frac{1}{2}(1-r)N(L_{31}, R^2) + \frac{1}{2}(1-r)N(L_{34}, R^2) + \frac{1}{2}rN(L_{32}, R^2) + \frac{1}{2}rN(L_{33}, R^2)$$

若两个主基因位点之间不存在连锁，即在似然方程中设置 $r = 0.5$ ，此时即为“二对独立主基因+微基因混合遗传模型”(记为Bm 20)。

而若仍设置 $r = 0.5$ ，并进一步考虑“主基因协调加性作用+微基因加性遗传模型”(记为Bm 21)，则基因效应与模型参数之间的关系为：

$$L_1 = m - d_a - d_b - [d] \quad L_2 = m + d_a + d_b + [d] \quad L_{31} = m - d_a - d_b \quad L_{32} = m - d_a + d_b \\ L_{33} = m + d_a - d_b \quad L_{34} = m + d_a + d_b \quad R^2 = D + R^2\sigma_n$$

式中, m 为群体平均数, d_a 和 d_b 分别为两个主基因座位的加性效应, $[d]$ 和 D 分别为微基因的加性效应和加性方差。对一阶统计量施加的约束条件为:

$$L_1 + L_2 = L_{31} + L_{34} = L_{32} + L_{33}$$

基因效应的估计为:

$$m = \frac{1}{2}(L_2 + L_1); d_a = \frac{1}{2}(L_2 + L_1 - L_{32} - L_{31}); d_b = \frac{1}{2}(L_{32} - L_{31}); [d] = L_{31} - L_1$$

再, 特别对于DH(或RL)群体而言, 因 $V_g = D = R^2 - R^2\sigma_n$, $V_e = R^2\sigma_n$, 而在“主基因协调加性作用”时 $V_G = d_a^2 + d_b^2$ 及 $V_p = V_G + V_g + V_e$, 从而可以估计各种遗传力、主—微基因效应相对比率、主基因座位效应相对比率和微基因的有效因子数($kg = [d]^2\sigma_D$)等遗传参数^[5, 7]。

值得说明的是, 若主基因的作用方式不是协调加性, 则其 V_G 的表达形式(及估计方法)就略不同, 具体详见文献^[5]。按照同样的原理, 在一定的约束条件下, 可以构造“主基因累加作用+ 微基因加性遗传模型”(记为Bm 22)、“主基因显性上位作用+ 微基因加性遗传模型”(记为Bm 23)、“主基因隐性上位作用+ 微基因加性遗传模型”(记为Bm 24)、“主基因重叠作用+ 微基因加性遗传模型”(记为Bm 25)、“主基因互补作用+ 微基因加性遗传模型”(记为Bm 26)、“主基因抑制作用+ 微基因加性遗传模型”(记为Bm 27), 有关结果列于表1之中(见表1中“二对主基因”部分内容; 而表1中“二对主基因(GW 法)”部分内容将在文末说明)。

表1 各种遗传模型及其所含的参数

Table 1 Various genetic models and its parameters

主基因状态 Major gene	微基因状态 Minor gene					
	无微基因 No minor gene			有微基因 Exist minor gene		
	模型代号 Model	一阶参数 First order parameter	二阶参数 Second order parameter	模型代号 Model	一阶参数 First order parameter	二阶参数 Second order parameter
无主基因 No major gene				m 20	$L_1L_2L_3$	R^2R^2
一对主基因 One major gene	A 20	L_1L_2	R^2	Am 20	$L_1L_2L_{31}L_{32}$	R^2R^2
	A 21	$m d$	R^2	Am 21	$m d [d]$	R^2R^2
	B 20	$L_1L_2L_{32}L_{33}$	R^2	Bm 20	$L_1L_2L_{31}L_{32}L_{33}L_{34}$	R^2R^2
	B 21	$m d_a d_b$	R^2	Bm 21	$m d_a d_b [d]$	R^2R^2
	B 22	$m d$	R^2	Bm 22	$m d [d]$	R^2R^2
二对主基因 Two major gene	B 23	$m d_a d_b$	R^2	Bm 23	$m d_a d_b [d]$	R^2R^2
	B 24	$m d_a d_b$	R^2	Bm 24	$m d_a d_b [d]$	R^2R^2
	B 25	$m i$	R^2	Bm 25	$m i [d]$	R^2R^2
	B 26	$m i$	R^2	Bm 26	$m i [d]$	R^2R^2
	B 27	$m i$	R^2	Bm 27	$m i [d]$	R^2R^2
二对主基因 (GW 法) Two major gene (GW method)	GB 20	$L_1L_2L_{31}L_{32}$	R^2	GBm 20	$L_1L_2L_{31}L_{32}L_{33}L_{34}$	R^2R^2
	GB 21	$m d_a d_b i$	R^2	GBm 21	$m d_a d_b i [d]$	R^2R^2
	GB 22	$m d_a d_b$	R^2	GBm 22	$m d_a d_b i [d]$	R^2R^2
	GB 23	$m d$	R^2	GBm 23	$m d [d]$	R^2R^2

注: 表中A 代表有一对主基因, B 代表有两对主基因; 模型代号中m 代表有微基因, 而一阶参数中m 代表群体平均值。A 20与A 21之间、GB 20与GB 21之间无本质差异, 因为二者都无约束条件, 分别列出只是为了记述上的对应。

GB 20代表独立遗传模型, GB 21代表加性2上位性遗传模型, GB 22代表加性遗传模型, GB 23代表等加性遗传模型。

——参见文末说明。

表中“二对主基因”栏和“二对主基因(GW 法)”栏中的i 并不同义, 读者并不难加以区分。

(在对 r 不施加约束的情况下, 可以建立与两主基因座位独立遗传类似的混合遗传模型, 并可特分别记为 B_{m-r-0} 、 B_{m-r-1} 、 \dots 、 B_{m-r-7} , 表1中没有具体列出)。

2.2 关于后续的分析过程

对表1所列遗传模型的各种参数仍采用极大似然估计, 而有关遗传模型的选择与检测遗传参数的估计等均按王建康等(1998)的方法进行。

2.3 若干补充说明

其一, DH 群体及 RL 群体是两类较特别的群体, 由于各家系均为纯合基因型, 其中没有显性效应的作用, 能够很方便的用于加性效应的分析; 且因分析时均采用家系均值, 所以分析结果更少的受环境影响。而也正因此, 从理论上讲, 采用DH 或 RL 群体还可能分析三个甚至四个主基因座位控制的QQ T (当然主要限于几种分离类型较简单的情形: 例如对于三对和四对主基因座位的QQ T, DH 家系的分离比在重叠作用时分别为 $1\div 7$ 、 $1\div 15$; 互补作用时分别为 $7\div 1$ 、 $15\div 1$; 累加作用时分别为 $1\div 3\div 3\div 1$ 、 $1\div 4\div 6\div 4\div 1$)。

其二, 本文中列出的一条遗传假定: “主基因和微基因之间不存在互作和连锁”, 而特别对于主基因之间的连锁与否而言, 自然界中确实有许多QQ T 的主基因之间不存在连锁, 例如油菜种子芥酸含量(主基因为二对独立遗传)、硫甙含量(主基因为三对独立遗传), 水稻胞质雄性不育恢复基因(主基因为二对独立遗传)等等, 这样也就保证了本文所示的基于DH 群体的分析理论与方法能直接用于处理RL 群体数据(当且仅当所论性状的所有基因座位之间不存在连锁时, DH 和 RL 群体才具有相同的遗传结构^[8])。

其三, 与胡中立等的前文(1998)相比, 本文因同时采用了 P_1 、 P_2 、DH (或 RL) 三个世代的数据资料, 从而使得分析结果更加准确可靠。

其四, 与王建康等人(1998)的多世代联合分析方法相比, 本文所示的方法除了所适用的世代群体不同之外, 还有两点不同: 即所适用的主基因对数不同(由一对扩展到二对)和所采用的遗传学概念(及对应的模型)不同。事实上, 如果两个主基因座位间的关系仍然采用经典数量遗传学中的加性 2 显性 2 上位性效应模型加以描述, 则有表1中“二对主基因(GW 法)”部分的内容与结果。显然采用这种模型, 即使估计了互作效应之值, 但也并不知道两主基因究竟是何种基因作用方式。

参 考 文 献

- 1 王建康, 盖钧镒 作物学报, 1998, 24(6): 651~ 659
- 2 Gai J Y, J K Wang *Theor Appl Genet*, 1998, 97: 1162~ 1168
- 3 王建康, 盖钧镒 遗传学报, 1997, 24(5): 432~ 440
- 4 盖钧镒, 王建康 作物学报, 1998, 24(4): 402~ 409
- 5 胡中立, 章志宏 武汉大学学报, 1998, 44(6): 784~ 788
- 6 胡中立, 章志宏, 宋运淳 生物数学学报, 1999, 14(3): 348~ 354
- 7 胡中立, 张修富, 章显光等 武汉大学学报, 1991, 37(1): 115~ 123
- 8 胡中立, 周发松, 何瑞锋等 武汉大学学报, 1996, 42(2): 249~ 251