

随机效应模型下标记-QTL 连锁分析方法

殷宗俊^{1,2} 张勤¹

(1. 中国农业大学 动物科学技术学院,北京 100094; 2. 安徽农业大学 动物科技学院,合肥 230036)

摘要 在 QTL 随机模型的框架内,通过对 QTL 方差组分的最大似然估计,进行家畜一般系谱群体的 QTL-标记连锁分析,研究中同时考虑了性状遗传力和 QTL 效应的不同水平对 QTL 定位效果的影响。结果表明,采用 QTL 随机效应模型可以有效地进行标记-QTL 间的连锁分析并可获得较高的检验功效,QTL 位置与效应参数的估计均在合理的范围内;同时,性状遗传力和 QTL 方差贡献对 QTL 定位准确度和功效均有较大的影响,随着性状遗传力和 QTL 方差贡献的增大,参数估计的准确性和检验功效提高。固定性状遗传力的情况下($h^2 = 0.1$),QTL 效应由 0.1~0.5 时,相应的检验功效由 17% 提高到 49%;当遗传力和 QTL 效应均提高到 0.5 时,QTL 检验功效最高达到 88%。

关键词 随机效应模型;连锁分析;QTL 定位;最大似然法

中图分类号 Q 348; S 813.1

文章编号 1007-4333(2005)04-0093-04

文献标识码 A

Linkage analysis between marker and QTL under the frame work of random effect model

Yin Zongjun^{1,2}, Zhang Qin¹

(1. College of Animal Science and Technology, China Agricultural University, Beijing 100094, China;

2. College of Animal Science and Technology, Anhui Agricultural University, Hefei 230036, China)

Abstract Methodology of linkage analysis by means of QTL variance using maximum likelihood based on the framework of a random QTL effect model was present in the paper. The effects of putative QTL were treated as random effect in the model. The factors of influencing QTL mapping (e. g. QTL effect and heritability) were simulated in our study too. The results showed that random QTL model have a good behavior for QTL mapping, and have a higher statistical power. The estimated QTL parameters are close to the expected values with reasonable standard errors in our study. In addition, the accuracy of QTL mapping depended on the effect of putative quantitative trait loci and the value of heritability. With the increase of QTL effect and heritability, the accuracy of QTL mapping improved slightly. When the heritability and QTL effect were 0.5 respectively, the power attained maximum (88%).

Key words random effect model; linkage analysis; QTL mapping; maximum likelihood

基因组研究的飞速发展和大量分子标记的涌现,使人们在分子水平上了解数量性状的遗传机理得以实现。当前分子遗传标记的主要用途是构建连锁图谱和进行基因检测。对于数量性状而言,主要是检测与定位影响数量性状的基因座(Quantitative Trait Loci, QTL),QTL 定位即是利用基因组中遗传标记的分离来估计与其连锁的 QTL 存在位置和效应大小,QTL 定位效率的高低主要取决于所采用

的试验设计和统计方法是否恰当。近年来,一些用于 QTL 检测与定位的统计方法相继提出^[1-2],并得到了不断的改进完善,QTL 检测的效率大大提高,定位的精度也在不断上升。

目前用于 QTL 检测与定位的统计方法大多都是基于区间定位,在染色体的一个区间里寻找 QTL 最为可能的位置,较为常用的有最小二乘法、最大似然法等。这些方法在多数情况下,通过 QTL 固定

收稿日期:2005-06-27

基金项目:国家重点基础研究发展规划(G2000016103);安徽省教育厅项目(2002jq126,2004kj151);安徽省自然科学基金项目(050410204)

作者简介:殷宗俊,博士,副教授,E-mail:yinzongjun@yahoo.com.cn;张勤,教授,博士生导师,通讯作者,E-mail:qzhang@cau.edu.cn

效应模型,估计出 QTL 替代效应,将 QTL 看作固定效应来处理。而在家畜群体中,由于双亲的 QTL 基因型一般未知,也不知道群体中 QTL 上等位基因的准确数目,另外,家系间基于标记的 QTL 基因型的频率不同。鉴于这些复杂情况,将 QTL 的效应看作随机效应比较现实,在随机 QTL 效应模型下,不仅可以对 QTL 的效应和位置进行估计,而且还能有效地估计出 QTL 的方差贡献。从现有的研究来看,建立在随机模型基础上的 QTL 分析方法主要有 BLUP 法^[3],Bayes 估计^[4],以及基于混合线性模型的最大似然法等,分析的群体主要是同胞或半同胞家系。但在家畜群体中,广泛存在的是更为一般系谱资源,本研究在 QTL 随机模型的框架内,通过对 QTL 方差组分的估计,进行家畜群体中一般系谱的 QTL-标记连锁分析,旨在探索家畜群体中合理有效的 QTL 连锁分析方法。

1 研究方法

1.1 研究设计

QTL 连锁分析的试验设计为全同胞-半同胞混合家系设计。基础群由 10 头公畜和 100 头母畜构成,每头公畜分别与 10 头母畜交配,每头母畜有 5 个信息后代,后代性别比例为 1:1。一共模拟了 12 个世代的群体发展过程,在每个世代中,随机选择 10 头公母和 100 头母畜作为下一世代的亲本,采用随机交配原则,但避免同胞交配,世代间不重叠。只考虑最后 3 个世代,即 10、11、12 世代的资料进行连锁分析。对上述设计的资源群体重复模拟产生 100 次,分别进行 QTL-标记间的连锁分析。

1.2 数据模拟

影响性状的因素包括非遗传效应、多基因效应以及一个位于常染色体上的具有 2 个等位基因的 QTL, QTL 独立于多基因。6 个多态标记均匀分布于 50cM 的染色体区段上,标记间的图距为 10cM, QTL 位于 6 个标记的中点处(25cM)。基础群个体中,各座位基因型(包括 QTL)根据假定的基因频率独立产生。

采用 Falconer 的模型^[5],设 QTL 的 3 种基因型 QQ 、 Qq 和 qq 的基因型值(v)分别为 a 、0 和 $-a$,因而在基础群中 QTL 方差为 $\sigma_{QTL}^2 = 2pq a^2$, p 和 q 分别为 QTL 2 个等位基因的频率,本研究中 QTL 2 个等位基因的频率分别设定为 0.5。对于给定的 QTL 方差,则有 $a = \sqrt{\sigma_{QTL}^2 / 2pq}$ 。个体表型值由模

型 $y = u + v + e$ 产生,其中: e 为随机环境误差,服从分布 $N(0, \sigma_e^2)$, u 为多基因值,服从分布 $N(0, \sigma_u^2)$ 。不失一般性,可设定表型方差为 1。QTL 效应则以 QTL 方差与总遗传方差($\sigma_G^2 = \sigma_u^2 + \sigma_{QTL}^2$)的比值表示,本研究中 QTL 效应考虑 3 个水平:0.10, 0.30 和 0.50。性状的遗传力(h^2),即总遗传方差与表型方差的比值,考虑 2 个水平 0.10, 0.50。

后代基因型根据亲本产生的单倍型确定,考虑位点间的重组,重组率采用 Haldane 图距函数计算。后代个体多基因效应为 $u_i = 0.5 u_i^s + 0.5 u_i^d + m_i$,其中 u_i^s 和 u_i^d 分别代表其父亲和母亲的多基因值, m 为孟德尔抽样离差。后代个体表型值的产生同基础群,世代间 σ_e^2 保持不变。

1.3 连锁分析方法

标记-QTL 连锁分析模型如下:

$$y = X\beta + Zu + Qv + e$$

其中: y 为观察值向量; β 为固定效应向量(包括总体均值); u 为育种值随机向量,其均值为 0,方差协方差矩阵为 $A\sigma_u^2$ (A 是分子亲缘相关矩阵^[6]); v 是随机 QTL 等位基因效应向量,期望值为 0,方差协方差矩阵为 $G\sigma_{QTL}^2$ (G 是 QTL 配子相关矩阵^[7]), Q 是 v 的关联矩阵; e 为误差向量,其均值为 0,方差协方差矩阵为 $I\sigma_e^2$; X , Z 分别为 y , u 的关联矩阵。 v 与 u 相互独立,相应的混合模型方程组为

$$\begin{bmatrix} X'X & X'Z & X'Q \\ Z'X & Z'Z + k_1 A^{-1} & Z'Q \\ Q'X & Q'Z & Q'Q + k_2 G^{-1} \end{bmatrix} \begin{bmatrix} \lambda \\ \mu \\ \nu \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \\ Q'y \end{bmatrix}$$

其中: k_1 和 k_2 分别为 σ_e^2 / σ_u^2 和 $\sigma_e^2 / \sigma_{QTL}^2$ 。

1.4 显著性检验与检验功效

利用最大似然法对上述方程进行迭代求解^[8],得到各参数的最大似然估计值,并可获得此时的最大对数似然函数值($\ln L_{max}$)。似然比定义为 $LR = 2(\ln L_{max} - \ln L_0)$,其中: L_0 为零假设下(无 QTL 存在)的似然函数,似然比近似服从自由度为 2 的卡方分布^[2]。通过 LR 值进行 QTL 参数的显著性检验。检验功效(Power)定义为重复模拟过程中 QTL 检验统计量显著的次数占总模拟次数的百分比。

2 结果与分析

2.1 QTL 参数估计

估计的 QTL 参数主要有其效应参数(a)和位置参数(r_1),其中 a 值是通过 QTL 等位基因频率

和方差获得 ($a = \sqrt{\frac{2}{QTL} \pi / 2pq}$)。不同遗传力和 QTL 效应下的参数估计结果见表 1 所示。可以看出,性状的遗传力高低直接影响着 QTL 定位的准确度,当模拟的性状遗传力较低时 ($h^2 = 0.1$), QTL 效应和位置参数的估计值与真值的偏差均较大,随着性状遗传力的提高 ($h^2 = 0.5$), QTL 参数估计的准确性也不断上升。特别是在高遗传力和高的 QTL 方差贡献的情况下,参数估计的结果最接近真值, QTL 位置参数估计的标准差也相对较小。

QTL 的方差贡献对 QTL 定位准确度也有非常大的影响。QTL 方差贡献是 QTL 方差在整个遗传方差中所占的比率,反映了 QTL 座位对性状表型的影响效应大小。因此,它会对连锁分析产生较大的影响,随着 QTL 方差贡献的增大,定位的准确性也会不断提高。

表 1 不同参数组合下的 QTL 效应与位置参数估计效果 (染色体长度为 50cM, QTL 在 25cM 处)

Table 1 Mean estimate of QTL parameters under different parameter case

遗传力	QTL 效应	QTL 基因型真值 (a)	a 估计均值 (标准差)	QTL 位置估计均值 (标准误)
0.10	0.10	0.141 4	0.187 1(0.087 5)	18.56(7.25)
0.10	0.30	0.244 9	0.205 2(0.122 0)	20.35(7.48)
0.10	0.50	0.316 2	0.253 0(0.135 4)	20.73(6.52)
0.50	0.10	0.316 2	0.344 5(0.122 4)	19.50(6.76)
0.50	0.30	0.547 7	0.452 6(0.146 0)	21.62(6.44)
0.50	0.50	0.707 1	0.628 3(0.142 8)	23.85(6.21)

2.2 检验功效

检验功效定义为检验显著次数占模拟总数(同一参数组合)的百分率,功效的高低反映着分析方法

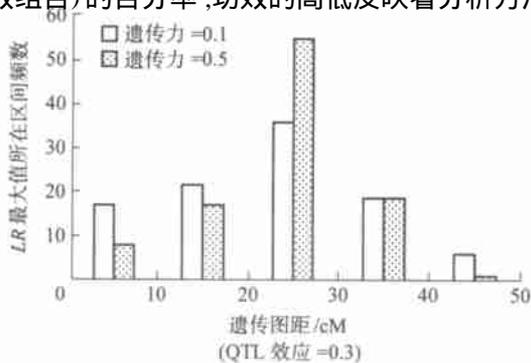


图 1 不同参数组合下 QTL 定位的区间分布频数(100 次模拟)

Fig. 1 Frequency when QTL was located in different marker bracket under some parameter cases (100 replicates)

的优劣。不同参数组合下的 QTL 检验功效和获得的最大 LR 值(似然检验统计量)列于表 2。

表 2 不同参数组合下的最大似然统计量 (LR_{max}) 和 QTL 定位的功效 ($\alpha = 0.05, 100$ replicates)

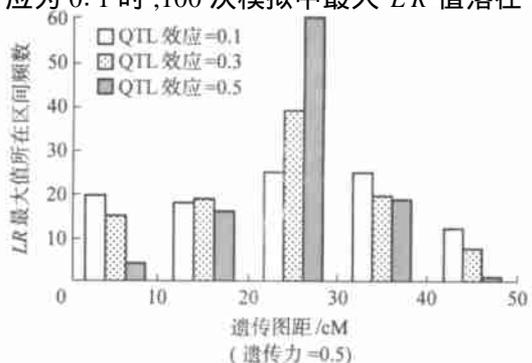
Table 2 The maximum LR statistic and power of QTL mapping under different parameter case

遗传力	QTL 效应	LR_{max}	检验功效
0.10	0.10	6.82	0.17
	0.30	9.53	0.34
	0.50	13.26	0.49
0.50	0.10	12.13	0.42
	0.30	16.78	0.68
	0.50	20.56	0.88

模拟研究的结果表明,不同参数组合下获得的 LR 差异较大,一般来说似然检验统计量愈高,检验显著的概率相应就愈大。因此,检验统计量的高低直接影响着检验功效大小。从表 2 的结果中可以看出,检验统计量的高低受到性状遗传力和 QTL 方差贡献的影响,随着性状遗传力和 QTL 方差贡献的提高,所获得的检验统计量也随之增加,相应的检验功效也不断增加。

2.3 QTL 定位区间分布

每一参数组合下,做了 100 次重复模拟研究,每次重复可获得一个最大 LR 值,统计最大 LR 值在每个标记区间内的出现频率(图 1)。模拟研究中, QTL 实际存在于标记 3 与标记 4 之间,即 25cM 处。由图 1 的结果可以看出,随着性状遗传力和 QTL 效应的增加,最大 LR 值落在 QTL 真实区间内的频数也随之增大, QTL 定位的准确性提高。相对而言, QTL 效应对定位准确性的影响最大,如当 QTL 效应为 0.1 时,100 次模拟中最大 LR 值落在 QTL 真



实区间内的次数只有 23 次,而当 QTL 效应增加到 0.5 时,相应的频数增加到 60 次,QTL 定位的准确性提高。

3 讨论

在家畜群体中,由于双亲的 QTL 基因型一般未知,也不知道群体中 QTL 上等位基因的确切数目,另外,家系间基于标记的 QTL 基因型的频率不同。鉴于这些复杂情况,将 QTL 的效应看作随机效应比较现实^[2]。同时,如果标记座位和 QTL 连锁不平衡,标记将提供 QTL 效应的一阶原点矩,此时可采用回归分析,利用表型和标记的信息进行 QTL 效应的估计。如果标记座位和 QTL 处于连锁平衡状态,则标记将不能提供 QTL 的基因型信息,此时难以采用常规的固定效应模型直接估计 QTL 的效应值。但即使在连锁平衡状态下,标记信息也会提供 QTL 效应间的方差协方差,因此可采用 QTL 随机效应模型估计出 QTL 的方差组分^[9]。

利用随机模型分析 QTL 的方差组分,并实现对 QTL 的定位,这时不再将 QTL 效应看作固定效应处理,而是作为一个随机效应来进行方差组分的估计。QTL 随机效应模型的另一优点是可以充分利用资源群体间的亲缘相关信息,提高分析的准确性。亲属间的遗传相似性(遗传协方差或遗传相关)是 QTL 方差组分估计的基础,在标记与 QTL 连锁的情况下,亲属 QTL 效应间的相关表现在 2 个方面,一是系谱信息上的分子亲缘相关(A 阵),再者,还可以借助标记信息推测标记与 QTL 等位基因在上下代的传递情况,以构建个体间的配子相关信息阵(同源概率矩阵,IBD 阵)。亲属间遗传相关信息推测的准确性直接影响着 QTL 连锁分析的效力,本研究中个体间的矩阵采用 Gamal 等^[7]提出的方法构建,并考虑了 2 个侧翼标记的情况^[8],从模拟研究的结果可以看出,采用 QTL 随机效应模型可以有效的进行标记-QTL 间的连锁分析并可获得较高的统计功效,QTL 位置与效应参数的估计均在合理的范围内。模拟研究的结果还显示,性状遗传力和 QTL 方差贡献对 QTL 定位准确度和功效都有较大的影响,随着性状遗传力和 QTL 方差贡献的增大,参数估计的准确性和检验功效有所提高。

进一步提高 QTL 检测与定位的精度,实现精

细定位和高效定位,是今后 QTL 定位研究的重点,除了要不断改进试验设计外,合理优化的统计分析方法也是必不可少的,统计方法的好坏直接影响着 QTL 检测定位的准确性和可靠性。连锁分析与连锁不平衡分析相结合应是今后家畜离散性状 QTL 精细定位的主要手段,虽然在 QTL 精细定位方面连锁不平衡分析优势明显,但其定位效力受到标记与 QTL 间连锁不平衡程度的限制^[10],因此有必要结合二者的优势进行 QTL 的高效定位。

参 考 文 献

- [1] Christoph L. Mapping quantitative trait loci using generalized estimating equations [J]. *Genetics*, 2001, 159: 1325-1337
- [2] Weller, Joel Ira. *Quantitative Trait Loci Analysis in Animals* [M]. Trowbridge: Printed and Bound in the UK by Cromwell Press, 2001
- [3] Van Arendonk J A M, Tier B, Kinghorn B P. Use of multiple genetic markers in prediction of breeding values [J]. *Genetics*, 1994, 137: 319-329
- [4] Hoeschele I, Vanranden P. Bayesian analysis of linkage between genetic markers and quantitative trait loci. I. prior knowledge[J]. *Theor Appl Genet*, 1993, 85: 953-960
- [5] Falconer D S, Mackay T F C. *Introduction to Quantitative Genetics* [M]. 4th ed. Essex: Longman, 1996.
- [6] Henderson C R. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values[J]. *Biometrics*, 1976, 32: 69-83
- [7] Gamal A A, Freeman A E. A rapid method for computing the inverse of the gametic covariance matrix between relatives for a marked quantitative trait locus[J]. *Genet Sel Evol*, 2001, 33: 153-173
- [8] Liu H Y, Zhang Q, Zhang Y. Relative efficiency of marker assisted selection when marker and QTL are incompletely linked[J]. *Chinese Sci Bull*, 2001, 46 (24): 2058-2063
- [9] Xu S, Atchley W R. A random model approach to interval mapping of quantitative trait loci[J]. *Genetics*, 1995, 141: 1189-1197
- [10] Spielman R S, Ewens W J. A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test [J]. *American Journal of Human Genetics*, 1998, 62: 450-458