

比值性状 QTL 的一种贝叶斯定位策略

孙丽华¹, 方 铭², 高会江^{3*}, 杜红丽¹

(1. 华南理工大学生物科学与工程学院, 广州 510006; 2. 中国农业大学动物科学技术学院, 北京 100193; 3. 中国农业科学院北京畜牧兽医研究所, 北京 100193)

摘要: 本研究旨在对比值性状 QTL 采用贝叶斯定位方法进行探讨。将两个具有正态分布的普通数量性状之比定义为比值性状。由于此类性状不服从正态分布, 因此在遗传分析上不同于普通性状。利用构成比值性状的两个组分性状的线性结合, 我们提出了定位比值性状数量性状位点(QTL)的最大似然法。在此基础上, 进一步探索比值性状的贝叶斯定位策略。首先采用贝叶斯压缩方法估计两个组分性状的 QTL 参数, 然后利用两个组分性状的群体均值和 QTL 遗传效应, 产生一个关于比值性状 QTL 遗传效应的后验样本, 通过对这个后验样本的统计分析, 来推断比值性状的 QTL。模拟研究表明: 新方法表现出比似然法更高的 QTL 检测效力。

关键词: 比值性状; 数量性状位点(QTL); 贝叶斯压缩估计; Monte-Carlo 模拟

中图分类号: S813.1; Q348

文献标识码: A

文章编号: 0366-6964(2010)07-0785-07

A Bayesian Strategy for Mapping QTL of Ratio Traits

SUN Li-hua¹, FANG Ming², GAO Hui-jiang^{3*}, DU Hong-li¹

(1. School of Biosciences and Bioengineering, South China University of Technology, Guangzhou 510006, China; 2. College of Animal Science and Technology, China Agricultural University, Beijing 100193, China; 3. Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, China)

Abstract: The research was conducted to study on the Bayesian strategy for mapping QTL of ratio traits. The ratio trait was defined as a ratio of two regular quantitative traits with normal distribution, which was distinguished from regular quantitative traits in the genetic analysis because it did not follow the normal distribution. A Bayesian mapping strategy was developed for ratio trait based on the maximum likelihood method using a special linear combination of the two component traits. The two component traits were analyzed by Bayesian shrinkage method, and then a new posterior sample of genetic effects was generated for a ratio trait from ones of population means and genetic effects for the two component traits, finally, QTL for the ratio trait was inferred via post MCMC analysis for the new posterior sample. Simulations analysis demonstrated that the new method had higher detecting power for the QTL than maximum likelihood method.

Key words: ratio trait; QTL; Bayesian shrinkage; MCMC

大多数重要的经济性状都属于比值性状, 如吸收率、饲料报酬、生长速度、乳脂率等, 它们都被定义为两个具有正态分布的普通数量性状的比值。在实际研究中, 两个组分性状的原始记录通常也被保留

着。研究者们早就开始关注此类性状的统计分析问题^[1-4]。然而, 由于比值性状不服从正态分布, 且当两个组分性状相互独立时, 它服从柯西分布^[5-6], 违背了采用最大似然法定位 QTL 所要求的正态性的

收稿日期: 2009-07-07

基金项目: 国家自然科学基金(30700565)

作者简介: 孙丽华(1970-), 女, 黑龙江牡丹江人, 讲师, 博士, 主要从事动物营养与免疫及生物统计应用方面的研究, E-mail: lhsun@scut.edu.cn

* 通讯作者: 高会江, E-mail: gaohj111@sina.com

前提条件。为定位比值性状的基因序列,人们首先想到采用数学转换方法将比值数据进行正态化处理,Box-Cox 转换是数据转换的一般方法,但应用在比值性状上并不是最优^[7];最小二乘也是常用的方法,但一系列模拟研究表明,其对非正态分布性状 QTL 的定位分析很难保证一致性^[8-11],且最小二乘法不能充分利用 QTL 基因型的混合分布信息;以百分比所定义的性状不同于这里的比值性状,它是由两个不连续变量(计数资料)构成,显然,用来定位百分比性状的统计方法,如阈转换法^[12-13]也不适用于比值性状。到目前为止,有许多关于比值性状 QTL 定位的文献报道,但大都是将比值性状视为普通数量性状进行分析^[14-18]。Yang 等^[19]采用两个组分性状的线性结合,提出了一种用来分析比值性状 QTL 的最大似然法,并通过一系列的模拟和实际资料对比分析证实:此方法同直接把比值性状视为普通数量性状的定位分析方法相比,能够充分地提高 QTL 检测的统计效力。

但最大似然法仍是建立在单个 QTL 模型基础上的,当一个数量性状是由多个 QTLs 控制时,它将不再是一个优化的方法。相比之下,贝叶斯方法由于是能够较容易的处理包括多个 QTLs 的复杂模型,而越来越广泛应用于数量性状的 QTL 定位分析中^[20-28]。在多个 QTLs 的贝叶斯定位分析中,最复杂的问题是确定 QTLs 的个数。为此,人们提出了一系列解决此问题的方案,如可逆跳跃 MCMC 法^[20-25],贝叶斯压缩分析法^[29-30],随机搜寻变量选择法(SSVS)^[31]及贝叶斯模型选择法^[27-31]等。模拟和实际数据分析指出,压缩分析方法在一定程度上优于其他的变量选择法,它不仅能提高马尔可夫链的收敛速度,还能减少 QTL 被遗漏的机会^[29-30]。本研究将利用压缩分析方法的优点,探索一种比值性状 QTL 的贝叶斯定位策略,即通过对两个组分性状的贝叶斯压缩分析,间接地实现对比值性状

$$\begin{aligned}
 p(y | \mu, b, \Sigma, \Sigma_e, x) &= \prod_{i=1}^n p(y_i | \mu, b, \Sigma, \Sigma_e, x) \\
 &\propto |\Sigma_e|^{-n/2} \exp\left[-\sum_{i=1}^n (y_i - \mu - \sum_{j=1}^q x_{ij}\beta_j)^T \Sigma_e^{-1} (y_i - \mu - \sum_{j=1}^q x_{ij}\beta_j)\right] \\
 p(m | x, \lambda) &= \prod_{i=1}^n \frac{p(m, x_{i\cdot} | \lambda)}{p(x_{i\cdot} | \lambda)}
 \end{aligned}$$

这里 $y = \{y_i\}$, $x = \{x_{ij}\}$, $b = \{\beta_j\}$, $\lambda = \{\lambda_j\}$, $i = 1, 2, \dots, n$, $j = 1, 2, \dots, q$ 。

在标记之间没有分离干预的前提下, $p(m, x_{i\cdot} |$

QTL 的定位。

1 理论和方法

首先采用贝叶斯压缩方法估计两个组分性状的 QTL 参数,然后利用两个组分性状的群体均值和遗传效应产生一个新的有关比值性状 QTL 遗传效应的后验样本,通过对这个新的后验样本的统计分析推断比值性状的 QTL。

1.1 两个组分性状 QTL 参数的贝叶斯压缩估计

为方便起见,以一个简单的回交群体为例,观测 n 个回交个体的表型值和具有中等以上密度的分子标记信息,假设两个组分性状是由 q 个数量性状位点控制,不考虑它们之间的交互,个体 i 的表型值 y_i 可以描述为:

$$y_i = \mu + \sum_{j=1}^q x_{ij}\beta_j + e_i \quad (1)$$

这里, μ 是群体均值的二维向量; β_j 是第 j 个 QTL 加性效应的二维向量, $j = 1, L, q$; 变量 x_{ij} 是个体 i 在第 j 个位点的基因型指示变量,两种基因型的指示变量分别定义为 1 和 -1; e_i 是随机误差的二维向量,假设它是服从于均值为 0,协方差矩阵为 Σ_e 的二元正态分布。

贝叶斯估计方法是把所有参数的联合先验和所有参数条件下观测值的似然组合起来,获得关于所有参数的联合后验分布,然后从联合后验分布中分别提取每个参数的边际后验分布,由每个参数的边际后验分布进行 MCMC 抽样。

所有参数条件下观测值的概率分布称为似然。观测值包括表型值数据和标记数据。由于两种观测值相互独立,故,

$$p(y, m | \mu, b, \Sigma, \Sigma_e, x, \lambda) = p(y | \mu, b, \Sigma, \Sigma_e, x) p(m | x, \lambda),$$

其中

λ 和 $p(x_{i\cdot} | \lambda)$ 可以通过马尔可夫模型推断。

在对普通性状 QTL 贝叶斯定位分析中,一般假设一个标记区间内最多只存在一个 QTL,那么

QTL 的最大数等于标记的区间数。而在普通性状 QTL 参数的贝叶斯压缩估计中,在中等标记密度下,研究者们经验地约定 QTL 所在的区间长度不能超过 40 cM^[29]。因此, QTL 个数的先验值 $q \in [h_1, h_2]$, 这里 h_1 等于整个基因组的长度除以 40 cM, h_2 是整个基因组标记区间的数目。对于群体均值 μ 来说,几乎没有任何先验信息,因此可将其先验分布假设为 $p(\mu) \propto \text{常数}$ 。根据贝叶斯压缩估计的观点^[29],可以认为每个 QTL 加性效应 β_j 的估计值

$$\begin{aligned} p(\mu, b, \Sigma, \Sigma_e, x, \lambda | y, m) &\propto p(y, m | \mu, b, \Sigma, \Sigma_e, x, \lambda) p(\mu, b, \Sigma, \Sigma_e, x, \lambda) \\ &= p(y | \mu, b, \Sigma, \Sigma_e, x) p(m | x, \lambda) p(\mu, b, \Sigma, \Sigma_e, x, \lambda) \\ &= p(y | \mu, b, \Sigma, \Sigma_e, x) p(m | x, \lambda) p(x | \lambda, m) p(\lambda) p(\mu) \\ &\quad \times p(b | \Sigma) p(\Sigma | v_\beta, S_\beta) p(\Sigma_e | v_e, S_e) \end{aligned} \quad (2)$$

这里, m 是已知的标记信息, $\Sigma = \{\Sigma_j\}$ 为 QTL 效应的先验方差, $j=1, 2, \dots, q$ 。

由(2)式,通过固定其他参数,获得任意目标未知参数的边际后验分布。在此基础上,概括两个组分性状群体均值和 QTL 参数的 MCMC 抽样步骤如下:

1.1.1 每个未知参数的初始化 直接赋予每个未知参数一个合理的初始值,或将相应先验分布抽样作为此参数的初始值。

1.1.2 更新群体均值 μ 从均值为 $\frac{1}{n} \sum_{i=1}^n (y_i - \sum_{j=1}^q x_{ij} \beta_j)$, 协方差矩阵为 $\frac{1}{n} \Sigma_e$ 的多元正态分布中抽样。

1.1.3 更新每个 QTL 效应 $\beta_j (j=1, 2, \dots, q)$ 其抽样分布是均值为 $[I \sum_{i=1}^n x_{ij}^2 + \Sigma_e(\Sigma_j)^{-1}]^{-1} \sum_{i=1}^n x_{ij} (y_i - \mu - \sum_{k \neq j} x_{ik} \beta_k)$, 协方差矩阵为 $[I \sum_{i=1}^n x_{ij}^2 + \Sigma_e(\Sigma_j)^{-1}]^{-1} \Sigma_e$ 的多元正态分布,这里的 I 是两维的单位矩阵。

1.1.4 更新每个 QTL 效应的协方差矩阵 $\Sigma_j (j=1, 2, \dots, q)$ 从参数为 $v_\beta + 1$ 和 $(v_\beta + 1) S_\beta + \beta_j^T \beta_j$ 所决定的逆-Wishart 分布中抽样。

1.1.5 更新剩余协方差矩阵 Σ_e 从参数 $v_e + n$ 和 $(v_e + n) S_e + \sum_{i=1}^n (y_i - \mu - \sum_{j=1}^q x_{ij} \beta_j)^T (y_i - \mu - \sum_{j=1}^q x_{ij} \beta_j)$ 所决定的逆-Wishart 分布中抽样。

1.1.6 更新 QTL 位置 $\lambda_j (j=1, 2, \dots, q)$ 和基因型 由于 QTL 位置的后验分布没有显性的表达式,需要利用 Metropolis-Hastings^[32-33] 算法来对 QTL 位置 λ 进行抽样。而 QTL 基因型(x)依赖于 QTL 位置(λ),因此一起抽取 QTL 位置和相应的基

是很多研究者对它的评价,且 $\beta_j \sim N(0, \Sigma_j)$, 其中 Σ_j 服从逆-Wishart 分布,即, $\Sigma_j \sim IW [v_\beta, (v_\beta S_\beta)^{-1}]$, $j=1, 2, \dots, q$ 。 Σ_e 的先验分布服从由超参数 v_e 和 S_e 所决定的逆-Wishart 分布,即 $\Sigma_e \sim IW (v_e, (v_e S_e)^{-1})$ 。 QTL 位置的先验概率 $p(\lambda_j) = \frac{1}{d_j}$, 这里, d_j 是第 j 个 QTL 所在区间的长度。

由上述各参数的先验分布和观测值的条件似然,构造所有未知参数的联合后验分布:

因型 $\{\lambda_j, x_j\}$ 。采用可变区间法^[29-34] 对 QTL 位点进行抽样,区间的边界是相邻 QTL 位置。那么 λ_j 的先验概率是:

$$p(\lambda_j) = U(\lambda_j; \lambda_{j-1}, \lambda_{j+1}) = 1/(\lambda_{j+1} - \lambda_{j-1})$$

这里 λ_{j-1} 和 λ_{j+1} 是第 j 个 QTL 左右两侧的 QTL 位置。令 $\lambda_j^{(t)}$ 为第 j 个 QTL 当前的抽样位置, $x_j^{(t)} = [x_{1j} \dots x_{nj}]^T$ 是所有个体在此位置的基因型向量。首先在 QTL 位置的先验分布中抽取一个新位置 $\lambda_j^* = \lambda_j + \delta$, 这里的 δ 是从均匀分布 $U(-s, s)$ 中抽样, s 是一个小的正数(调节参数),一般取 1-2 cM。在新位置上抽取新 QTL 基因型 x_j^* , 然后通过 M-H 规则^[29,34] 来判定是否接受新的位置 λ_j^* 。若 λ_j^* 被接受,则更新 QTL 位置和基因型,即 $\lambda_j^{(t+1)} = \lambda_j^*$, $x_j^{(t+1)} = x_j^*$, 否则, λ_j 和 x_j 保持不变,即 $\lambda_j^{(t+1)} = \lambda_j^{(t)}$, $x_j^{(t+1)} = x_j^{(t)}$ 。

1.1.7 抽取缺失标记基因型 标记基因型的后验分布为 $p(m | x, \lambda)$ 。如果标记缺失,可根据两侧最近的非缺失标记推断其可能基因型存在的概率,然后根据缺失标记的各种基因型存在的概率,在每轮迭代中随机抽样产生新的基因型。然而,在此研究中,通过相邻标记推断的概率是先验概率,如果将表型值嵌入标记(或 QTL)效应中,相应的概率就变成了后验概率,就可用后验概率来抽取缺失标记的基因型。

1.1.8 MCMC 抽样 重复 1.1.2-1.1.7 步骤,直到 MCMC 链达到事先给定的长度为止。

1.2 比值性状 QTL 的推断

比值性状基因定位的目标,是检测导致构成它的两个组分性状群体均值之比发生变化的 QTL。通过上述 MCMC 抽样,获得有关两个组分性状的

群体均值和 QTL 效应实现值。于是我们计算比值性状的两个 QTL 基因型效应为 $\frac{\mu_1 + \beta_1}{\mu_2 + \beta_2}$ 和 $\frac{\mu_1 - \beta_1}{\mu_2 - \beta_2}$ 。根据 QTL 基因型效应和遗传效应的关系,间接地生成比值性状 QTL 的加性效应 β_R :

$$\beta_R = \frac{1}{2} \left(\frac{\mu_1 + \beta_1}{\mu_2 + \beta_2} - \frac{\mu_1 - \beta_1}{\mu_2 - \beta_2} \right)$$

这里 μ_1 和 μ_2 是两个信息性状的群体均值实现值, β_1 和 β_2 是相应的加性效应实现值。

由所派生出来的比值性状 QTL 加性效应的后验样本,可通过构造 Z 统计量来推断整个基因组上任意位置对比值性状作用大小的显著性, $Z(\lambda) = \frac{\beta_R(\lambda)}{\sqrt{V_R(\lambda)}}$, 这里的 $\beta_R(\lambda)$ 是位点 λ 对比值性状的加性效应, $V_R(\lambda)$ 是位点 λ 相应加性效应的样本方差。在大样本情况下, $Z(\lambda)$ 服从标准正态分布,在 0.05 显著标准下,用来判断位置 λ 对比值性状影响显著性的 $Z(\lambda)$ 统计量的临界值为 1.96。为清晰显示 QTL 位置(整个基因组上对比值性状有显著影响的位置),在作 Z 统计量轮廓图时,将位置 λ 的 Z 统计量计算公式重新定义如下:

$$Z(\lambda) = \begin{cases} \frac{\beta_R(\lambda)}{\sqrt{V_R(\lambda)}} & Z(\lambda) \geq 1.96 \\ 0 & Z(\lambda) \leq 1.96 \end{cases}$$

2 结果分析

为考察此方法的可行性,模拟了一个包括 150 个个体的回交群体。在一条长度为 600 cM 的染色体片段上均匀地排列 61 个共显性标记。假设有 9 个 QTLs 控制着构成比值性状的两个组分性状,这些 QTLs 的效应及其在染色体片段上的位置见表 1,它们对于每个组分性状都解释了 80% 的表型变异。这些模拟的 QTL 对每个性状影响不同:有的只影响一个性状,如 QTL2, QTL3, QTL4, QTL6 和 QTL7; 有的同时影响两个性状,表现为一因多效。群体均值 $\mu = (12.8 \quad 6.4)$, 残差协方差矩阵为:

$$\Sigma_e = \begin{bmatrix} 1 & 0.25 \\ 0.25 & 1 \end{bmatrix}$$

在进行 MCMC 抽样之前,对所有参数进行初始化。其中将 QTL 个数初步确定为 20 个,即模型中的 $q = 20$ (可变区间平均长度为 $600/20 = 30$ cM)。超参数 v_β 和 v_e 被赋予 0, 而 S_β 和 S_e 被赋予单位矩阵。其它参数的初始值从其先验分布中抽取。设定马尔

可夫链总长为 35 000, 即进行 35 000 轮 MCMC 抽样, 删除处于非平稳期的前 5 000 轮抽样, 剩余 30 000 轮中每隔 30 轮记录一次数据。可见, 用来进行后验分析的后验样本将包括 $30\,000/30 = 1\,000$ 个独立数据。模拟试验重复 5 次。鉴于 5 次试验均表现出较小的差异, 本文仅统计其中一次抽样试验的结果。

为分析比值性状 QTL, 首先从两个组分性状的总体均值和加性效应中产生一个关于比值性状的 QTL 效应的后验样本, 然后计算基因组上每个位点的 Z 统计量, 以便作 Z 统计量轮廓图。由图 1 可见, 有 7 个陡峭峰超过了临界值所在水平线。清楚地检测到了影响比值性状的 9 个 QTL 中的 7 个。相应的参数估计结果列于表 2, 模拟结果表明, 本文提出的新方法能够较灵敏的检测到影响比值性状的 QTL。没有被检测到的 QTL 要么其效应很小, 要么是事先设定的对两个组分性状有影响但对比值性状没有影响, 如 QTL8, 由于它对两个组分性状的加性效应与两个组分性状的群体均值成比例, 因此认为其对比值性状的加性效应为 0。本文也用最近提出的最大似然法^[19] 分析了相同的模拟资料。图 2 是似然比检验统计量轮廓图。通过 500 次重复排列模拟数据来确定 5% 显著性水平临界值为 313.1 (看图 2 中水平线)。显然, 似然法所检测出的 QTL 数目少于本研究提出的贝叶斯方法, 且贝叶斯方法能够精确地估计模拟的 QTL 效应, 而最大似然法一般倾向于过高估计 QTL 效应(表 2)。

3 讨论

正如所预期的那样, 本研究所提出的比值性状 QTL 贝叶斯定位策略表现出比最大似然法更高的 QTL 检测效力。尽管最大似然法用两个组分性状的一个特殊线性结合来间接定位比值性状 QTL, 而不是直接分析两个组分性状, 但它和新的贝叶斯定位方法仍是可比的。因为两种方法都是将比值性状的加性效应定义为 $\beta_R = \frac{1}{2} \left(\frac{\mu_1 + \beta_1}{\mu_2 + \beta_2} - \frac{\mu_1 - \beta_1}{\mu_2 - \beta_2} \right)$ 。最大似然法之所以没有直接分析两个组分性状, 是因为非线性的无效假设 $\beta_R = 0$, 导致不能获得关于 QTL 参数的显性表达式, 而不便于 EM 法迭代求解。

实际上, 比值性状 QTL 的贝叶斯定位策略能够较容易地扩展到分析由两个组分性状合成的任意组合性状。在回交群体中, 假设任意组合性状的群体均值 (μ_c) 是两个组分性状群体均值 (μ_1 和 μ_2) 的

表1 模拟试验中 QTL 的参数

Table 1 QTL parameters used in simulations

QTL 编号 QTL No.	部位 Location	性状 1 Trait 1		性状 2 Trait 2		比值 Ratio
		加性效应 Additive effect	遗传力 Heritability	加性效应 Additive effect	遗传力 Heritability	加性效应 Additive effect
1	23	1.20	0.282	0.51	0.052	0.026
2	56	0.00	0.000	1.07	0.229	-0.286
3	148	0.72	0.102	0.00	0.000	0.113
4	193	0.00	0.00	1.24	0.307	-0.325
5	267	0.22	0.009	0.55	0.060	-0.127
6	332	0.67	0.088	0.00	0.000	0.105
7	390	0.00	0.000	0.58	0.067	-0.166
8	476	1.28	0.321	0.64	0.082	0.000
9	574	0.10	0.002	0.10	0.002	-0.015
总和 Sum			0.800		0.800	

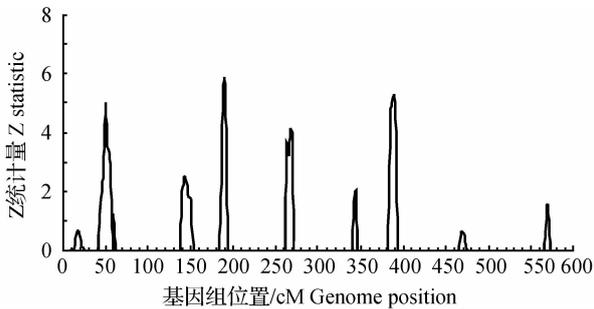


图1 Z 统计量轮廓图

Fig. 1 The profile of Z statistic

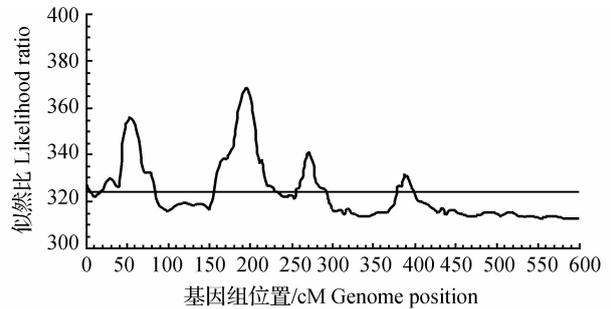


图2 似然比统计量轮廓图

Fig. 2 The profile of likelihood ratio statistic

表2 QTL 位置和加性效应的估计值

Table 2 QTL location and the evaluated values of additive effect

QTL 编号 QTL No.	贝叶斯法 Bayesian method		最大似然法 Maximum likelihood	
	位置 Location	效应 Effect	位置 Location	效应 Effect
1	—	—	—	—
2	52	-0.335(0.149)	60	-0.504(0.201)
3	146	0.168(0.105)	—	—
4	192	-0.416(0.171)	198	-0.550(0.212)
5	270	-0.139(0.068)	270	-0.268(0.170)
6	346	0.068(0.048)	—	—
7	390	-0.244(0.106)	388	-0.154(0.121)
8	—	—	—	—
9	572	-0.017(0.011)	—	—

函数: $\mu_c = f(\mu_1, \mu_2)$, 可用如下公式计算控制组合性状的 QTL 效应:

$$\beta_c = \frac{1}{2} [f(\mu_1 + \beta_1, \mu_2 + \beta_2) - f(\mu_1 - \beta_1, \mu_2 - \beta_2)]$$

再通过分析 β_c 的后验样本来推断组合性状的 QTL。

在动植物育种中, 有很多性状是由其他正态性状以各种函数的形式组合而成, 如标准产奶量和多性状综合指数等, 我们不妨称其为组合性状。在没有发展适合这类性状的基因定位分析的专用方法之前, 人们还习惯和普通正态性状一样, 直接分析组合

性状。根据比值性状的贝叶斯定位策略, 一旦我们通过多性状的贝叶斯定位方法首先获得各个组分性状 QTL 参数的后验样本, 利用这些后验样本就可以较高的检测效率简便地实现对任意组合性状基因位点的估计和评价。

本研究仅以简单的回交群体为例来描述比值性状 QTL 定位的贝叶斯方法, 但将该方法扩展到 DH, RIL, F_2 群体和全同胞家系并不困难。在贝叶斯 QTL 定位框架中, 随机搜寻变量选择法 (SS-VS)^[30] 和贝叶斯模型选择法^[27, 31] 也都是常见的统

计方法。Wang 等^[29]比较了 SSVS 法和压缩分析法后,发现两种方法表现出基本一致的定位效果。相比之下,贝叶斯模型选择方法则由于在每轮抽样中仅估计显著的 QTL 而大大地节省了计算时间。因此,在多个 QTL 模型分析中,如果同时考虑两个 QTL 之间的交互作用,与贝叶斯压缩估计和 SSVS 相比,贝叶斯模型选择应该是一个比较好的方法。

4 结 论

比值性状在动植物育种中比较广泛地存在。为了高效率地检测控制这类性状的基因位点,可以首先采用贝叶斯压缩方法抽取两个组分性状的 QTL 参数的后验样本,然后通过对这些后验样本的统计分析,来间接推断比值性状的 QTL。该法很容易推广为由多个正态性状组成的任意组合性状的基因定位分析。

参考文献:

- [1] SUTHERLAND T M. The correlation between feed efficiency and rate of gain, a ratio and its denominator [J]. *Biometrics*, 1965, 21:739-749.
- [2] GUNSETT F C, ANDRIANO K N, RUTLEDGE J J. Estimating the precision of estimates of genetic parameters realized from multiple-trait selection experiments[J]. *Biometrics*, 1982, 38(4):981-989.
- [3] GUNSETT F C. Linear index selection to improve traits defined as ratios[J]. *Journal of Animal Science*, 1984, 59:1185-1193.
- [4] GUNSETT F C. Merit of utilizing the heritability of a ratio to predict genetic change of a ratio[J]. *Journal of Animal Science*, 1987, 65:936-942.
- [5] HINKLEY D V. On the ratio of two correlated normal random variables[J]. *Biometrika*, 1969, 56:635-639.
- [6] KENDALL M, STUART A. The advanced theory of statistics[M]. 4th edition. Griffin:London,1979: 1-355.
- [7] YANG R, YI N, XU S. Box-Cox transformation for QTL mapping[J]. *Genetica*, 2006, 28:133-143.
- [8] JANSEN R C. A general mixture model for mapping quantitative trait loci by using molecular markers[J]. *Theor Appl Genet*, 1992, 85: 252-260.
- [9] REBAÏ A. Comparison of methods for regression interval mapping in QTL analysis with non-normal traits[J]. *Gene Res*, 1997, 69: 69-74.
- [10] HACKETT C A. Model diagnostics for fitting QTL models to trait and marker data by interval mapping [J]. *Heredity*, 1997, 79: 319-328.
- [11] COPPIETERS W, KVASZ A, ARRANZ J J, et al. A rank-based non parametric method to map QTL in outbred half-sib pedigrees: application to milk production in a grand-daughter design [J]. *Genetics*, 1998, 149:1547-1555.
- [12] CHIB S, GREENBERG E. Analysis of multivariate probit models[J]. *Biometrika*, 1998, 85 (2): 347-361.
- [13] MAO Y, XU S. Mapping QTLs for traits measured as percentages[J]. *Genet Res*, 2004, 83: 159-168.
- [14] TSAROUHAS V, GULLBERG U, LAGERCRANTZ U. An AFLP and RFLP linkage map and quantitative trait locus (QTL) analysis of growth traits in *Salix*[J]. *Theor Appl Genet*, 2002, 105(2-3): 277-288.
- [15] VAN KAAM J B, BINK M C, BOVENHUIS H, et al. Scaling to account for heterogeneous variances in a Bayesian analysis of broiler quantitative trait loci[J]. *J Animal Sci*, 2002, 80: 45-56.
- [16] VIITALA S M, SCHULMAN N F, DE KONING D J, et al. Quantitative trait loci affecting milk production traits in finnish ayrshire dairy cattle[J]. *J Dairy Sci*, 2003, 86: 1828-1836.
- [17] WELLER J I, GOLIK M, SEROUSSI E, et al. Population-wide analysis of a QTL affecting milk-fat production in the Israeli Holstein population[J]. *J Dairy Sci*, 2003, 86: 2219-2227.
- [18] CASAS E, SHACKELFORD S D, KEELE J W, et al. Detection of quantitative trait loci for growth and carcass composition in cattle[J]. *Journal of Animal Science*, 2003, 81: 2976-2983.
- [19] YANG R, LI J H, XU S. Mapping quantitative trait loci for traits defined as ratios[J]. *Genetica*, 2008, 132:323-329.
- [20] SATAGOPAN J M, YANDELL B S, NEWTON M A, et al. A Bayesian approach to detect quantitative trait loci using Markov chain Monte Carlo[J]. *Genetics*, 1996, 144: 805-816.
- [21] STEPHENS D A, FISCH R D. Bayesian analysis of quantitative trait locus data using reversible jump Markov chain Monte Carlo [J]. *Biometrics*, 1998, 54: 1334-1347.
- [22] SILLANPÄÄM J, ARJAS E. Bayesian mapping of multiple quantitative trait loci from incomplete out-

- bred offspring data[J]. *Genetics*, 1999, 151: 1605-1619.
- [23] YI N, XU S. Bayesian mapping of quantitative trait loci under the identity-by-descent-based variance component model[J]. *Genetics*, 2000, 156: 411-422.
- [24] UIMARI P, SAILLANPÄÄM J. Bayesian oligogenic analysis of quantitative and qualitative traits in general pedigrees [J]. *Genet Epidemiol*, 2001, 21 (3): 224-242.
- [25] CORANDER J, SILLANPÄÄM J. A unified approach to joint modeling of multiple quantitative and qualitative traits in gene mapping[J]. *J Theor Bio*, 2002, 218: 435-446.
- [26] YI N, XU S. Mapping quantitative trait loci with epistatic effects[J]. *Genet Res*, 2002, 79: 185-198.
- [27] YI N, YANDELL B S, CHURCHILL G A, et al. Bayesian model selection for genome-wide epistatic quantitative trait loci analysis [J]. *Genetics*, 2005, 170: 1333-1344.
- [28] XU S. Estimating polygenic effects using markers of the entire genome[J]. *Genetics*, 2003, 163: 789-801.
- [29] WANG H, ZHANG Y M, LI X, et al. Bayesian shrinkage estimation of quantitative trait loci parameters[J]. *Genetics*, 2005, 170: 465-480.
- [30] YI N, GEORGE V, ALLISON D B. Stochastic search variable selection for identifying multiple quantitative trait loci [J]. *Genetics*, 2003, 164: 1129-1138.
- [31] YI N, ALLISON D B, XU S. Bayesian model choice and search strategies for mapping multiple epistatic quantitative trait loci[J]. *Genetics*, 2003, 165: 867-883.
- [32] METROPOLIS N, ROSENBLUTH A W, ROSENBLUTH M N, et al. Equations of state calculations by fast computing machines[J]. *J Chem Phys*, 1953, 21: 1087-1091.
- [33] HASTINGS W K. Monte Carlo sampling methods using Markov chains and their applications [J]. *Biometrika*, 1970, 57: 97-109.
- [34] ZHANG Y M. Advanced statistical methods for detecting multiple quantitative trait loci[J]. *Recent Res Devel Genet Breeding*, 2005, 2: 1-23.

(编辑 郭云雁)