

DOI: 10.1360/yc-007-0668

# 贝叶斯统计在 QTL 作图中的应用研究进展

敖雁, 朱明星, 徐辰武

扬州大学江苏省作物遗传生理重点实验室, 扬州 225009

**摘要:** 在许多复杂情况下, 贝叶斯统计方法比经典数理统计方法能更直接解决问题, 且可有效整合部分先验信息, 但其需要高强度计算的特性曾限制了其广泛应用。近几十年来, 随着高速计算机的发展以及 MCMC 算法的不断提出, 贝叶斯方法已被用于群体遗传学、分子进化、连锁作图和数量遗传学等研究领域, 文章综述了数量遗传学中 QTL 作图的贝叶斯方法从简单到复杂的发展历程。

**关键词:** 贝叶斯统计; QTL 作图; 进展

## Advances in researches on application of Bayesian methods to QTL mapping

AO Yan, ZHU Ming-Xing, XU Chen-Wu

Jiangsu Provincial Key Laboratory of Crop Genetics and Physiology, Yangzhou University, Yangzhou 225009, China

**Abstract:** In many cases, Bayesian methods can solve problems of interest more directly than a classical approach. Its utility lies on the incorporation of prior information. In recent years, with the development of high-speed computer and advances of MCMC algorithm, Bayesian methods have been employed in many genetic areas, such as population genetics, molecular evolution, linkage mapping, and quantitative genetics and so on. In this review, we reviewed the development of Bayesian approaches for quantitative trait locus (QTL) mapping in quantitative genetics.

**Keywords:** Bayesian statistics; QTL mapping; advance

贝叶斯概率是复杂事件的条件概率, 由之发展起来的贝叶斯统计学在现代统计学、决策学、生态学、医疗诊断学、分子生物学和数量遗传学等领域都已有着广泛的应用, 并已成为这些学科领域中一些重要原理和方法的依据之一<sup>[1-4]</sup>。国际上有关贝叶斯统计在数量遗传学上的应用研究早已开展, 多见于人类<sup>[5, 6]</sup>, 植物上相对较少<sup>[7, 8]</sup>。而在国内, 汤在祥等<sup>[9]</sup>探讨过贝叶斯统计在遗传连锁分析中的应用, 包括遗传重组率的贝叶斯估计、遗传连锁的贝叶斯因子检验和基于马尔可夫链蒙特卡罗理论的遗传连锁图谱构建, 但还未见有关贝叶斯统计方法在 QTL

作图中研究进展的报道。本文综述了 QTL 作图从单标记分析到多标记分析; 从单 QTL 模型到多 QTL 模型; 从分析信息完全数据到分析信息不完全数据; 从利用双亲本杂交衍生的分离群体到利用多亲本衍生的分离群体; 从区间作图到多区间作图; 从加性—显性模型到加性—显性—上位性模型的贝叶斯统计方法的发展历程。

### 1 QTL 作图

QTL 作图<sup>[10]</sup>, 就是根据标记基因型和数量性状表现型, 应用一定的统计方法, 在遗传连锁图上标

收稿日期: 2006-10-20; 修回日期: 2006-12-22

基金项目: 国家自然科学基金(编号: 30270724 和 30370758)资助[Supported by National Natural Science Foundation of China(No.30270724, 30370758)]

作者简介: 敖雁(1974-), 女, 达斡尔族, 讲师, 在职博士生, 研究方向: 数量遗传学

通讯作者: 徐辰武(1965-), 男, 教授, 博士生导师, 研究方向: 生物统计学和统计基因组学。Tel: 0514-7979358; E-mail: qtls@yzu.edu.cn

定有关QTL的位置,并估计其效应。其基本统计方法称为标记性状关联分析(marker-trait association analysis)。其基本原理是<sup>[10]</sup>:如果在某标记座位M附近有着一个影响某性状数量大小的QTL,则在该M与该QTL均杂合而发生共分离时,必使M的不同类别( $M_1M_1$ 、 $M_1M_2$ 和 $M_2M_2$ )与该性状表现的大小构成关联;因此测验M的不同类别间数量性状平均数的差异或相关,就可能发现此种关联。其作用可概括为<sup>[11]</sup>:(1)可更好的理解物种间和物种内复杂性状的遗传结构;(2)有利于重要作物的进一步改善(比如抗旱水稻)<sup>[12,13]</sup>,也可提高人类对复杂生理疾病(如高血压)的治疗效果<sup>[14]</sup>;(3)可对自然群体或作物驯化时进行强选择的那些性状进行研究,因此QTL作图不仅是一个发现基因的工具,还提供了数量进化的遗传进程的关键信息。其技术路线为<sup>[15]</sup>:(1)选择具有相对数量性状的纯系A和B,进行杂交和自交,获得分离世代 $F_2$ 、BC或重组群体DH(加倍单倍体)、RIL(重组近交系)等;或进行多系杂交得到相应的分离群体。(2)检测世代群体中每一个体的标记基因型和数量性状值。(3)分析标记基因型和数量性状值之间是否连锁,发现QTL并估计其遗传参数。实行上述路线,要特别注意两个问题:(1)纯系亲本间相对性状的差异应尽可能大,这样,该性状的分离世代就能提供最丰富的标记和性状连锁的信息,提高检测效率。分子标记应具备中性、共显性、准确性和广布于整个基因组(标记间的平均遗传距离 $<15\sim 20$  cM)的基本特征。

## 2 贝叶斯统计

近几十年来,贝叶斯统计已发展为与经典统计并列的两大流派之一。假定统计模型中的参数以 $\theta$ 表示,试验前对 $\theta$ 往往有某些先验知识,即先验分布。先验知识可由过去的某些数据资料分析而得,也可以凭过去对 $\theta$ 的某些经验知识获得。贝叶斯统计推断的思想是:利用先验分布和样本观测值 $\Psi$ 提供的信息,组成较完整的后验信息。再通过分析后验分布的性质(通常为平均数和众数等),得到参数的点估计;通过分析后验分布的置信区间(包含参数的 $1-\alpha$ 的区间, $\alpha$ 是显著水平,一般取0.05或0.01),获得参数的区间估计。而经典统计仅利用样本观测值 $\Psi$ 进行统计推断。因此,贝叶斯统计实质上比经典统计有着更广泛的应用范围,其理论与方法已渗透到数据分析的各个领域中。

在大部分情况下,贝叶斯统计的后验分布没有一个明确的形式,因此较难估计,一般运用Markov chain Monte Carlo (MCMC)算法或其它计算技术从后验分布中产生后验样本。MCMC算法<sup>[16,17]</sup>并非新技术,但直到Gibbs取样方法<sup>[18]</sup>的引入,其在统计学中才被广泛应用,并逐步应用于基因作图等方面的研究<sup>[19,20]</sup>。MCMC描述了一种模拟特定类型随机过程的方法,称为马尔可夫链,以研究不易被其他方法研究的复杂概率分布的性质。其基本思想是构建具有稳定分布的马尔可夫链,该链产生的一系列将来状态的随机变量,完全由该链任一点目前的状态所决定。经过足够多的循环后,马尔可夫链将会有个稳定的不再依靠循环次数和变量初值的概率分布,然后从中抽样进行统计推断。在贝叶斯分析里,概率分布通常是一个或多个参数的联合后验分布,如果条件后验分布有标准的形式(如服从正态分布),则可从该标准分布中直接抽取每一轮循环的参数估计值,该方法称为Gibbs抽样方法<sup>[18,21]</sup>;但如果后验分布没有一个明确的形式,则可采用Metropolis-Hastings(MH)算法<sup>[16,17,22]</sup>,也称为接受-拒绝方法。MH算法类似于MC模拟程序,即从稳定的马尔可夫链中抽样以模拟服从一定概率分布的观察值。但这些观察值不是独立的,而是从链中模拟连续的值直至收敛,再以一定间隔从连续的值中抽样以模拟具有稳定分布的独立样本。其具体算法可见参考文献<sup>[16,17,22]</sup>。MH算法的优点是:由于模拟的变量均是以高概率稳定在链上的值,因此当状态空间很大时,MH算法能提高模拟的效率。其缺点是:在大部分实践应用中,还没有严格的方法决定链何时收敛,或决定抽取最多信息而又能保持观察值间具有独立性的最佳区间。

## 3 研究进展

### 3.1 单 QTL 作图

利用贝叶斯方法进行连锁分析的报道早已有之<sup>[23-25]</sup>。Tai<sup>[26]</sup>, Thomas和Cortessis<sup>[23]</sup>, Smith和Roberts<sup>[27]</sup>, Hoeschele和Vanranden<sup>[24]</sup>, Stephens和Smith<sup>[28]</sup>等考虑曾利用贝叶斯推断进行基因作图。如:Thomas和Cortessis<sup>[23]</sup>在一个大的系谱里使用Gibbs取样技术实现了用贝叶斯方法估计连锁的LOD分值,但未讨论多个座位的位置及其效应的置信区间,即未估计影响感兴趣性状座位的数目。Hoeschele和Vanranden<sup>[24]</sup>在女儿和孙女设计的动物模型里计算了一个QTL和一个标记连锁的后验概

率。Thaller和Hoeschele<sup>[29,30]</sup>发展MCMC算法实现了Hoeschele和Vanranden<sup>[24]</sup>的一个QTL和一个标记之间连锁的贝叶斯分析。此后,Knott和haley<sup>[31]</sup>认为与利用单个标记进行QTL作图相比,多个连锁标记的使用可以增加QTL检测的功效和提高参数估计的准确度,并且可以降低QTL位置估计的偏度。由于图谱上越来越多的标记被发现,故可利用一个连锁群上的所有标记检测染色体上QTL的存在并估计其位置。有鉴于此,Uimari<sup>[5]</sup>提出了利用多个连锁标记的信息进行QTL作图的贝叶斯方法。

### 3.2 多 QTL 作图

传统的 QTL 作图方法主要涉及多元回归模型和极大似然估计,这些方法对具有大效应 QTL 的检测是有效的。而对起修饰作用或与大效应交互作用的多个小的遗传效应的检测至今仍是一个挑战。而这些小的效应能够潜在的促进作物育种和进一步提高人类对复杂疾病遗传效应的理解,量化这些效应可填补遗传效应分布有关知识的空白。

随着高速计算机的发展以及MCMC算法的不断提出,贝叶斯方法借助其广泛的适用性,已被用于解决多QTL作图等复杂问题<sup>[32]</sup>。该方法的优势<sup>[33]</sup>在于:第一,贝叶斯的灵活性和通用性,使之可以处理具有任意复杂性的模型。第二,贝叶斯方法充分考虑了所有未知因素的不确定性。由于目标参数的估计值均由潜在模型的平均值进行推断,因此并非像非贝叶斯方法那样选择单一模型。因此,与传统的非贝叶斯方法相比,贝叶斯推断更为稳健。第三,贝叶斯方法的全部分析均是直接基于后验分布,不需要分别采用参数估计、假设测验和多重比较等统计方法。此外,贝叶斯方法为用户在分析中合并使用先验信息提供了正规程序。许多研究者利用复杂的可逆跳跃的MCMC算法<sup>[34]</sup>实现的贝叶斯QTL作图方法估计QTL的数目和性状的效应<sup>[8,35,36]</sup>。

#### 3.2.1 简单效应的 QTL 作图

##### 3.2.1.1 模型选择估计方法

Piepho等认为多QTL的检测是一个变量选择的统计问题<sup>[37-39]</sup>,为了解决该类问题,Satagopan等发展了经由MCMC算法实现的多QTL作图的贝叶斯方法<sup>[7,8,35,36,40-42]</sup>,Satagopan<sup>[8]</sup>的方法假定了被研究的染色体上QTL的数目,使用贝叶斯方法的多座位模型,利用MCMC技术同时鉴别了多个QTL及其效应

的大小,而不是一次只配合一个座位。与此同时,Stephens和Fisch<sup>[43]</sup>也同时考虑了所有染色体,把染色体数目作为一个随机变量。但是Satagopan和Yandell的例子中仅包括一条染色体,而Stephens和Fisch没有考虑缺失数据。此外,Uimari<sup>[6]</sup>将贝叶斯统计的QTL作图拓展到分析两个连锁的QTL。

1998年,Sillanpaa<sup>[35]</sup>提出了处理不完全近交系组合数据的贝叶斯作图方法,将QTL的数目作为一个未知的随机变量且使用了与复合区间作图相似的思想,考虑了其他染色体上的QTL效应。1999年<sup>[44]</sup>其进一步提出了对远交物种的F<sub>2</sub>或BC家系的完全和不完全数据进行作图的方法。Vogl等<sup>[45]</sup>利用贝叶斯方法处理标记不完全信息的系谱资料,利用MCMC算法更新标记座位基因型时,采用了descent-graph算法,模拟试验证明该方法可处理复杂系谱并且可在一个连锁群上检测2个QTL。Yi<sup>[46]</sup>提出了相似的方法,不同之处在于采用可逆跳跃的MCMC算法,沿着基因组可同时搜索多个QTL,甚至推断QTL数目的后验分布。随后,Xu<sup>[7]</sup>将贝叶斯回归分析方法拓展到同时估计全基因组上所有标记的遗传效应,该方法能够处理的效应数目可大于观察值数目,成功的关键在于每一个标记效应有其变异参数,变异参数又有其先验分布,从而可从数据估计该变异。在分等级的模型下该方法能够处理大量的具有很小效应的标记,因此可估计标记效应的分布。该方法的难点为:(1)如何在一个模型中处理大量的标记?(2)如何处理效应接近0的标记?Xu等在随机回归系数模型下使用贝叶斯方法解决上述问题,其作法是:每一个基因效应被假定遵循平均数为0和特定方差的正态先验,这个特定效应的先验方差进一步被假定遵循一个模糊先验,从而可从数据估计该方差,这个方法与Meuwissen等<sup>[47]</sup>的贝叶斯方法相类似。

截至2003年,已经发展了多QTL作图的许多统计方法,包括多区间作图法<sup>[48,49]</sup>,变量选择法<sup>[38,40,50]</sup>,利用可逆跳跃的MCMC算法实现的贝叶斯方法<sup>[8,32,35,36,51-54]</sup>。这些方法均涉及模型决定和变量选择等问题<sup>[55]</sup>。由于QTL的数目定义了模型的维数,因此研究QTL的数目是个复杂的过程。最早的研究工作是Jansen<sup>[56]</sup>和Stam<sup>[57]</sup>,Kao<sup>[48]</sup>采用了逐步回归方法逐步增加和删除QTL直到模型稳定。Ball<sup>[40]</sup>,

Piepho和Gauch<sup>[37]</sup>, Broman和Speed等<sup>[50]</sup>对贝叶斯信息标准(BIC)进行了研究。贝叶斯因子<sup>[8, 58]</sup>或可逆跳跃MCMC均影响QTL数目的估计<sup>[34, 35]</sup>。但估计QTL数目时, 参数空间的不确定维数常引起马尔可夫链的混合。为了解决该问题, YI<sup>[42]</sup>基于Stochastic Search Variable Selection(SSVS)方法发展了鉴别多QTL的贝叶斯方法。早在1993年, George和McMulloch<sup>[59]</sup>就对SSVS进行过研究, 他们试着避免改变模型的维数但仍然运用了变量选择的思想。在YI的SSVS分析里, 每一个回归系数(效应)被假定遵循两个正态分布的混合, 一个分布平均数为0方差很小, 另一个分布平均数为0方差较大, 如果属于大方差分布的模型效应的后验概率是高的, 该效应即被选择, 否则被剔除。参数空间的维数以这种方式被固定, 然后再以较大效应标记的后验概率表示QTL的显著性。该不完全选择方法已避免了由全模型选择方法引起的许多问题。但小标准差的分布模拟小的效应的做法, 导致在QTL效应较大时, 相应的效应也具有较小的后验概率, 因此降低了QTL作图程序的效率。鉴于此, Zhang<sup>[60]</sup>发展了一种分类方法, 该方法建立在整合了重要先验信息的基础上, 即大部分标记不与一个QTL共分离或者它们的QTL效应是可忽略的。这个方法特别适用于具有大量标记但相对小的样本容量的数据。众所周知, 贝叶斯分析的一个缺点是需要高强度的计算, 缺失值较少时计算上不存在问题。尽管Zhang的贝叶斯方法可以容易地解决缺失数据的问题, 但如有大量的缺失基因型数据, 计算速度较慢。一个可供考虑的解决方法是在每一对相邻的标记之间至多有一个QTL并采用YI<sup>[61]</sup>的复合空间表示法, 此外先验的特化也能影响贝叶斯算法的性能。

### 3.2.1.2 压缩估计方法

在多QTL作图方法中, 另有一些不需要变量选择的方法, 这些方法也可处理大量的模型效应。如最初的脊回归方法, 模型中包含所有有潜力的效应, 但被估计的效应被压缩接近0, 这类估计被称为压缩估计或非模型选择估计。Xu<sup>[7]</sup>认为脊回归方法中模型效应的数目不宜太大。在脊回归分析中加在系数矩阵对角线上的数字等于剩余方差与QTL效应的方差参数的比率, Xu修改了QTL效应先验分布的方差参数并令方差参数随着QTL变动, 该方法与Sauerbrei<sup>[62]</sup>的方法类似。

### 3.2.2 复杂上位性效应的 QTL 作图

在统计基因组学领域, 将贝叶斯方法应用于主效应QTL作图已有近10年时间。如起初发展用于QTL数目固定的多QTL作图的贝叶斯方法<sup>[5, 8, 28]</sup>以及发展用于联合推断QTL数目、基因组位置和效应大小的可逆跳跃MCMC(RJ-MCMC)算法<sup>[34]</sup>。此类方法已被成功应用于一些实例分析<sup>[8, 63, 64]</sup>。数量性状主效应的加性-显性模型只涉及等位基因加性效应与座位内等位基因间的互作(显性)的效应, 未涉及不同座位间的基因互作。然而, 这是不可忽略的。这是因为: (1)上位性可能是引起杂种优势的原因<sup>[65]</sup>; (2)上位性在解释数量性状遗传变异中占有相当的比重<sup>[66]</sup>; (3)基因网络研究需要对上位性<sup>[67]</sup>的理解。

近年来, 部分研究者尝试将贝叶斯方法用于复杂上位性基因作图研究。如, Sen和Churchil<sup>[41]</sup>提出了一套上位性QTL作图的贝叶斯方法, 但是其基本假设仍然采用双QTL模型。Yi和Xu<sup>[68]</sup>提出利用RJ-MCMC算法联合推断QTL数目、基因组位置、主效应和上位性效应。他们在模型中考虑了所有可能的主效应和上位性效应, 并利用RJ-MCMC算法向模型添加新QTL或删除已有QTL。但是, 对于大量数目的假定QTL而言, 包含所有可能的上位性效应必然导致数目巨大的潜在参数, 从而造成遗传模型选择的问题。另一方面, Yi和Xu的方法并不适用于没有主效应、但有互作效应的两个QTL的检测。最近, Yi等<sup>[69-71]</sup>发展了一套专用于BSB小鼠模型复杂上位性QTL作图的贝叶斯新方法。他们提出利用指示向量代表模型中实际效应的数目和大小, 可以方便地从模型中删除不显著的主效应和上位性效应。

若假定每一标记区间都存在一个QTL且QTL间存在互作, 则Bayesian压缩估计方法<sup>[72, 73]</sup>能解决其参数估计问题。Wu和Li<sup>[74]</sup>认为Bayesian压缩估计方法<sup>[71, 72, 75]</sup>将QTL定位策略推进到利用所有标记的上位性QTL的全基因组检测。该法的不足就是运算时间太长。Zhang等人<sup>[76]</sup>提出了估计QTL上位性效应的惩罚极大似然方法, 该方法能够处理过饱和模型, 即效应的数目比观察值的数目大许多倍的模型。这个方法将被假定的效应压缩为接近0, 而具有大效应的QTL不被压缩, 因此不再需要模型选择, 与模型选择有关的所有问题都不必考虑, 例如模型参数空间的充分探究和高强度的计算时间等问题。该方法与贝叶斯压缩分析相比较计算速度大大提高。此

外, 可变区间 Bayesian 压缩估计方法<sup>[71, 75]</sup>也可解决 Bayesian 压缩估计方法运算时间太长的问題。

#### 4 小结与展望

起初 QTL 作图的 Bayesian 方法发展较慢。但近些年来, 已成为数量遗传学研究领域的一个热点。截至目前可大致将其分为两大类方法: 可逆跳跃 MCMC 的 Bayesian 方法和 Bayesian 压缩估计方法。前者主要是将 QTL 数目作为变量并假定它具有一定的先验分布, 推导出其后验分布后, 再利用 MCMC 技术在后验分布中不断抽样, 样本收敛后可得 QTL 数目、位置和效应的估计值。但该方法收敛较慢, QTL 数目变化具有可逆性。后者假定每一标记区间都存在一个 QTL, 若 QTL 效应估计值接近 0, 说明该位置无 QTL, 否则有 QTL 存在。但该方法在变量过多时导致参数估计时间过长。最近, 章元明等将 Bayesian 压缩估计方法似然化, 大大节约了运算时间。其方法是选择恰当的参数先验分布, 以参数后验密度函数为目标函数, 用最大似然法估计其参数。即惩罚最大似然方法<sup>[76]</sup>。

本文旨在综述 QTL 作图的贝叶斯方法从简单到复杂的发展历程, 明确贝叶斯方法在 QTL 作图中的优势及适用场合, 并非将之取代经典的 QTL 作图方法。随着 QTL 作图研究的深入和进一步复杂化, 模型中参数将急剧增多, 比如: 在基因网络和 eQTL 作图等方面的研究上, 贝叶斯方法有望发挥一定的效用。

#### 参考文献(References):

- [1] Berry DA, Stangl DK. In: Bayesian Biostatistics, 2nd ed. New York: Marcel Dekker Inc, 2000, 1-696.
- [2] Dey DK, Ghosh SK, Mallick BK. In: Generalized Linear Models: A Bayesian Perspective. New York: Marcel Dekker Inc, 2000, 1-440.
- [3] Stangl DK, Berry DA. In: Meta-analysis in Medicine and Health Policy. New York: Marcel Dekker Inc, 2000, 1-414.
- [4] Weir BS. In: Genetic Data Analysis II. Massachusetts: Sinauer Associates Inc, Publisher, 1996, 1-445.
- [5] Uimari P, Thaller G, Hoeschele I. The use of multiple markers in a Bayesian method for mapping quantitative trait loci. *Genetics*, 1996, 143(4): 1831-1842.
- [6] Uimari P, Hoeschele I. Mapping-linked quantitative trait loci using Bayesian analysis and Markov Chain Monte Carlo Algorithms. *Genetics*, 1997, 146(2): 735-743.
- [7] Xu S Z. Estimating polygenic effects using markers of the entire genome. *Genetics*, 2003, 163(2): 789-801.
- [8] Satagopan JM, Yandell BS, Newton MA, Osborn TC. A bayesian approach to detect quantitative trait loci using markov chain monte carlo. *Genetics*, 1996, 144(2): 805-816.
- [9] TANG Zai-Xiang, WANG Xue-Feng, WU Wen-Wen, XU Chen-Wu. Bayesian statistics-based method for genetic linkage analysis. *Hereditas(Beijing)*, 2006, 28 (9): 1117-1122. 汤在祥, 王学枫, 吴雯雯, 徐辰武. 基于贝叶斯统计的遗传连锁分析方法. *遗传*, 2006 28 (9): 1117-1122.
- [10] MO Hui-Dong. Look back and reflect on genetic researches of variation for quantitative traits-A challenge for quantitative genetics in post-genome era. *Journal of Yangzhou University (Agricultural and Life Sciences Edition)*, 2003, 24(2): 24-31. 莫惠栋. 数量性状遗传基础研究的回顾与思考-后基因组时代数量遗传领域的挑战. *扬州大学学报(农业与生命科学版)*, 2003, 24(2): 24-31.
- [11] Zhang M, Montooth KL, Wells MT, Clark AG, Zhang D. Mapping multiple quantitative trait loci by Bayesian classification. *Genetics*, 2005, 169(4): 2305-2318.
- [12] Price AH, Courtois B. Mapping QTLs associated with drought resistance in rice: progress, problems and prospects. *Plant Growth Reg*, 1999, 29(1-2): 123-133.
- [13] Price AH, Cairns JE, Horton P, Jones HG, Griffiths H. Linking drought-resistance mechanisms to drought avoidance in upland rice using a QTL approach: progress and new opportunities to integrate stomatal and mesophyll responses. *J Exp Bot*, 2002, 53(371): 989-1004.
- [14] Sugiyama F, Churchill GA, Higgins DC, Johns C, Makaritsis KP, Gavras H, Paigen B. Concordance of murine quantitative trait loci for salt-induced hypertension with rat and human loci. *Genomics*, 2001, 71(1): 70-77.
- [15] Helentjaris T, Burr B. Development and application of molecular markers to problems in plant genetics. *Cold Spring Harbor Press*, 1989, 1-165.
- [16] Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. Equation of state calculations by fast computing machines. *J Chem Phys*, 1953, 21(6): 1087-1092.
- [17] Hastings WK. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 1970, 57(1): 97-109.
- [18] Geman S, Geman D. Stochastic relaxation, Gibbs distribution, and the Bayesian restoration of images. *IEEE Trans Pattn Anal Mach Intell*, 1984, PAMI(6): 721-741.
- [19] Thompson EA. Monte Carlo likelihood in genetic mapping. *Statist Sci*, 1994, 9(3): 355-366.
- [20] Thomas DC, Gauderman WJ. Gibbs sampling methods in genetics, in Markov Chain Monte Carlo in Practice, edited by Gilks WR, Richardson S and Spiegelhalter DJ. London: Chapman & Hall, 1995, 419-440.
- [21] Casella G, George EI. Explaining the Gibbs sampler. *Am Statist*, 1992, 46(2): 167-174.
- [22] Chib S, Greenberg E. Understanding the Metropolis-Hastings algorithm. *Am Statist*, 1995, 49(4): 327-335.
- [23] Thomas DC, Cortessis V. A Gibbs sampling approach to

- linkage analysis. *Hum Hered*, 1992, 42(1): 63–76.
- [24] Hoeschele I, Vanraden PM. Bayesian analysis of linkage between genetic markers and quantitative trait loci. I. Prior knowledge. *Theor Appl Genet*, 1993, 85(8): 953–960.
- [25] Hoeschele I, vanraden PM. Bayesian analysis of linkage between genetic markers and quantitative trait loci. II. Combining prior knowledge with experimental evidence. *Theor Appl Genet*, 1993, 85(8): 946–952.
- [26] Tai JJ. Application of Bayesian decision procedure to the inference of genetic linkage. *J Am Statist Assoc*, 1989, 84(407): 669–673.
- [27] Smith AFM, Roberts GO. Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods. *J Roy Statist Soc Ser B*, 1993, 55(1): 3–23.
- [28] Stephens DA, Smith AF. Bayesian inference in multipoint gene mapping. *Ann Human Genet*, 1993, 57(1): 65–82.
- [29] Thaller G, Hoeschele I. A Monte Carlo method for Bayesian analysis of linkage between single markers and quantitative trait loci. I. Methodology. *Theor Appl Genet*, 1996, 93(7): 1161–1166.
- [30] Thaller G, Hoeschele I. A Monte Carlo method for Bayesian analysis of linkage between single markers and quantitative trait loci. II. A simulation study. *Theor Appl Genet*, 1996, 93(7): 1167–1174.
- [31] Knott SA, Haley CS. Aspects of maximum likelihood methods for the mapping of quantitative trait loci in line crosses. *Genet Res Camb*, 1992, 60(2): 139–151.
- [32] Hoeschele I. Mapping quantitative trait loci in outbred pedigrees. In: Handbook of Statistical Genetics. Balding D, Bishop M, Cannings C (eds). Wiley, 2001, 599–644.
- [33] Carlin BP, Louis TA. Bayes and Empirical Bayes methods for data analysis. 2nd edition. London, Chapman & Hall, 2000.
- [34] Green PJ. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, 1995, 82(4): 711–732.
- [35] Sillanpää MA, Arjas E. Bayesian mapping of multiple quantitative trait loci from incomplete inbred line cross data. *Genetics*, 1998, 148(3): 1373–1388.
- [36] Stephens DA, Fisch RD. Bayesian analysis of quantitative trait locus data using reversible jump Markov chain Monte Carlo. *Biometrics*, 1998, 54(4): 1334–1347.
- [37] Piepho HP, Gauch HG Jr. Marker pair selection for mapping quantitative trait loci. *Genetics*, 2001, 157(1): 433–444.
- [38] Broman KW, Speed TP. A model selection approach for the identification of quantitative trait loci in experimental crosses. *J R Stat Soc B*, 2002, 64(4): 641–656.
- [39] Sillanpää MJ, Corander J. Model choice in gene mapping: what and why. *Trends Genet*, 2002, 18(6): 301–307.
- [40] Ball RD. Bayesian methods for quantitative trait loci mapping based on model selection: approximate analysis using the Bayesian information criterion. *Genetics*, 2001, 159(3): 1351–1364.
- [41] Sen S, Churchill G. A statistical framework for quantitative trait mapping. *Genetics*, 2001, 159(1): 371–387.
- [42] Yi N, George V, Allison DB. Stochastic search variable selection for identifying multiple quantitative trait loci. *Genetics*, 2003, 164(3): 1129–1138.
- [43] Stephens DA, Fisch RD. Bayesian analysis of quantitative trait locus data using reversible jump Markov chain Monte Carlo. Technical report. (available at <http://www.ma.ic.ac.uk/statistics/techrep.html>) 1996.
- [44] Sillanpää MA, Arjas E. Bayesian mapping of multiple quantitative trait loci from incomplete outbred offspring data. *Genetics*, 1999, 151(4): 1605–1619.
- [45] Vogl C, Xu S. QTL analysis in arbitrary pedigrees with incomplete marker information. *Heredity*, 2002, 89(5): 339–345.
- [46] Yi N, Xu S. Linkage analysis of quantitative trait loci in multiple line crosses. *Genetica*, 2002, 114(3): 217–230.
- [47] Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide marker maps. *Genetics*, 2001, 157(4): 1819–1829.
- [48] Kao CH, Zeng ZB, Teasdale RD. Multiple interval mapping for quantitative trait loci. *Genetics*, 1999, 152(3): 1203–1216.
- [49] Zeng ZB, Kao CH, Basten CJ. Estimating the genetic architecture of quantitative traits. *Genet Res*, 2000, 74(3): 279–289.
- [50] Broman KW, Speed TP. A model selection approach for the identification of quantitative trait loci in experimental crosses. *J R Stat Soc B*, 2002, 64(4): 641–656.
- [51] Heath SC. Markov Chain Monte Carlo segregation and linkage analysis for oligogenic models. *Am J Hum Genet*, 1997, 61(3): 748–760.
- [52] Xu S, Yi N. Mixed model analysis of quantitative trait loci. *Proc Natl Acad Sci USA*, 2000, 97(26): 14542–14547.
- [53] Yi N, Xu S. Bayesian mapping of quantitative trait loci for complex binary traits. *Genetics*, 2000, 155(3): 1391–1403.
- [54] Yi N, Xu S. Bayesian mapping of quantitative trait loci under complicated mating designs. *Genetics*, 2001, 157(4): 1759–771.
- [55] Sillanpää MJ, Corander J. Model choice in gene mapping: what and why. *Trends Genet*, 2002, 18(6): 301–307.
- [56] Jansen RC. Interval mapping of multiple quantitative trait loci. *Genetics*, 1993, 135(1): 205–211.
- [57] Jansen RC, Stam P. High resolution of quantitative traits into multiple loci via interval mapping. *Genetics*, 1994, 136(4): 1447–1455.
- [58] Kass R, Raftery A. Bayes factors. *J Am Stat Assoc*, 1995, 90(430): 773–795.
- [59] George EI, McCulloch RE. Variable selection via Gibbs sampling. *J Am Stat Assoc*, 1993, 88(423): 881–889.
- [60] Zhang M, Montooth KL, Wells MT, Clark AG, Zhang D. Mapping multiple quantitative trait loci by Bayesian classification. *Genetics*, 2005, 169(4): 2305–2318.
- [61] Yi N. A unified Markov Chain Monte Carlo framework for mapping multiple quantitative trait loci. *Genetics*, 2004,

- 167(2): 967-975.
- [62] Sauerbrei W. The use of resampling methods to simplify regression models in medical statistics. *Appl Stat*, 1999, 48(3): 313-329.
- [63] [Bink M, Uimari P, Sillanpää M, Janss L, Jansen R. Multiple QTL mapping in related plant populations via a pedigree-analysis approach. *Theor Appl Genet*, 2002, 104(5): 751-762.
- [64] Hurme P, Sillanpää M J, Arjas E, Repo T, Savolainen O. Genetic basis of climatic adaptation in Scots pine by Bayesian quantitative trait locus analysis. *Genetics*, 2000, 156(3): 1309-1322.
- [65] Hua JP, Xing YZ, Wu WR, Xu CG, Sun XL, Yu SB, Zhang QF. Single-locus heterotic effects and dominance by dominance interactions can adequately explain the genetic basis of heterosis in a elite rice hybrid. *Proc Natl Acad Sci USA*, 2003, 100(5): 2574-2579.
- [66] Carlborg O, Haley CS. Epistasis: too often neglected in complex trait studies? *Nat Rev Genet*, 2004, 5(8): 618-625.
- [67] Segre D, Deluna A, Church GM, Kishony R. Modular epistasis in yeast metabolism. *Nat Genet*, 2005, 37(1): 77-83.
- [68] Yi N, Xu S. Mapping quantitative trait loci with epistatic effects. *Genet Res*, 2002, 79(2): 185-198.
- [69] Yi N, Diament A, Chiu S, Fisler J, Warden C. Characterization of epistasis influencing complex spontaneous obesity in the BSB model. *Genetics*, 2004, 167(1): 399-409.
- [70] Yi N, Diament A, Chiu S, Fisler J, Warden C. Epistatic interaction between chromosomes 7 and 3 influences hepatic lipase activity in BSB mice. *J Lipid Res*, 2004, 45(11): 2063-2070.
- [71] Yi N, Yandell BS, Churchill GA, Allison DB, Eisen EJ, Pomp D. Bayesian model selection for genome-wide epistatic QTL analysis. *Genetics*, 2005, 170(3): 1333-1344.
- [72] Wang H, Zhang YM, Li X, Masinde GL, Mohan S, Baylink DJ, Xu S. Bayesian shrinkage estimation of quantitative trait loci parameters. *Genetics*, 2005, 170(1): 465-480.
- [73] Zhang YM, Xu S. Mapping quantitative trait loci in F2 incorporating phenotypes of F3 progeny. *Genetics*, 2004, 166(4): 1981-1993.
- [74] Wu RL, Li M. Functional mapping -how to map and study the genetic architecture of dynamic complex traits. *Nat Rev Genet*, 2006, 7(3): 229-237.
- [75] Zhang YM, Xu S. Advanced statistical methods for detecting multiple quantitative trait loci. *Recent Res Dev Genet Breed*, 2005, 2(1): 1-23.
- [76] Zhang YM, Xu S. A penalized maximum likelihood method for estimating epistatic effects of QTL. *Heredity*, 2005, 95(1): 96-104.

## 科学出版社生命科学编辑部新书推介 2007-03

### 发育神经生物学 (第二版)

蔡文琴 主编

978-7-03-016833-7/Q·1691 定价: 85.00 2007年3月出版

发育神经生物学是神经科学的一个重要分支学科。本书在参阅国际出版相关专著的基础上,结合近年发育神经生物学的进展。着重介绍了神经系统从发育至老化中的有关问题及其分子调控与研究方法。在第一版的内容上增加了新的内容。全书共19章,由多位专家教授编算而成。

本书适合从事神经科学的教研人员研究生,也可供生命科学相关专业的学者及医学院校师生参考。

### 现代生物工程 (应用生物技术大系)

焦炳华 孙树汉 主编

978-7-03-017544-1/Q·1756 定价: 78.00 2007年3月出版

本书是为落实国家生命科学与技术人才培养基地建设委员会关于加强生物类专业教学的要求而组织编写的。书中全面介绍了现代生物技术的概念、原理、研究方法发展方向及其应用领域。全书共分现代生物工程导论及基因工程、细胞工程、蛋白质工程、酶工程、抗体工程、组织工程和生物芯片7篇,共33章。本书可作为高等院校生物专业学生专业教学的教材,也可供给综合性大学、师范、农林、医药院校有关专业本科生、研究生及教师作为参考用书。

欢迎各界人士邮购科学出版社各类图书(免邮费)。

邮购地址:100717 北京东黄城根北街16号科学出版社 科学分社,联系人:阮芯 联系电话:010-64034622(带传真)

更多精彩图书请登陆网站<http://www.lifescience.com.cn>, 欢迎致电索要书目