

基于显性模型的基因组选择中贝叶斯方法研究

王延晖, 朱 波, 李俊雅*

(中国农业科学院北京畜牧兽医研究所, 北京 100193)

摘 要: 本研究旨在探索显性效应对基因组育种值估计准确性的影响。基于贝叶斯 A 模型, 根据加性效应和显性效应相关性, 提出两种子模型: 1) 加性效应和显性效应相互独立的 BayesAD1 模型; 2) 显性系数(Dominant coefficients)与加性效应的绝对值相互独立, 并且显性系数服从正态分布的 BayesAD2 模型。通过模拟数据比较加性效应模型 BayesAD0 和两种显性效应模型下基因组估计育种值(GEBV)的准确性, 并且研究不同数量性状基因座(Quantitative trait locus, QTL)数、全同胞家系内个体数和加性方差与显性方差的比重对 GEBV 准确性的影响。结果表明, 显性模型可以减缓随着世代变化 GEBV 的准确性降低的趋势。另外, 显性方差的比重越大, 对 GEBV 的准确性影响越大。当加性方差与显性方差之比达到 0.25 时, BayesAD2 有较大优势, 分别比 BayesAD1 和 BayesAD0 准确性高 20.3% 和 28.4%。全同胞数越多, GEBV 的准确性越高, QTL 数目增多, GEBV 估计的准确性随之下降。结果显示, 对显性效应占较大比重即低遗传力的性状进行育种值估计时, 考虑显性效应可以提高育种值估计准确性。

关键词: 全基因组选择; 贝叶斯模型; 显性效应; 显性系数

中图分类号: S813.1

文献标志码: A

文章编号: 0366-6964(2017)01-0060-08

Bayesian Models including Dominant Effects for Genomic Selection

WANG Yan-hui, ZHU Bo, LI Jun-ya*

(Institute of Animal Science, Chinese Academy of Agricultural Sciences, Beijing 100193, China)

Abstract: The aim of this study was to investigate the impact of dominant effects on the predictive accuracy of genomic breeding values. Based on the dependency between additive and dominant effects using BayesA model, we proposed two submodels: 1) the additive effect and dominant effect were independent of each other in BayesAD1 model; 2) the dominant coefficient and absolute values of additive effect were independent of each other in BayesAD2 model, in which the dominant coefficient followed normal distribution. Using simulated datasets, we compared the predictive accuracy of genomic estimated breeding value (GEBV) among the additive model (BayesAD0) and dominant models (BayesAD1 and BayesAD2). We further investigated the effect of number of QTLs (quantitative trait loci), size of full-sibs family and ratio of additive variance to dominant variance on predictive accuracy of GEBV. The results showed that the dominant models slowed down the declining of the predictive accuracy of GEBV in subsequent generations. Moreover, the predictive accuracy of GEBV increased as the ratio of dominant variance increase. When the ratio of additive variance to dominant variance reached 0.25, BayesAD2 was 20.3% and 28.4% higher than the accuracy of the BayesAD1 and BayesAD0, respectively. In addition, the size of full-sibs family affected the predictive accuracy of GEBV positively. And an increase in the number of QTL was accompanied by a reduction on the predictive accuracy of GE-

收稿日期: 2016-10-27

基金项目: 现代农业(肉牛)产业技术体系岗位科学家(CARS-38); 中国农业科学院科技创新工程-牛遗传育种(ASTIP-IAS03)

作者简介: 王延晖(1988-), 女, 甘肃景泰人, 硕士, 主要从事动物遗传育种与繁殖(数量遗传)研究, E-mail: wangyanhui@mail@126.com

* 通信作者: 李俊雅, 研究员, E-mail: jll@iascaas.net.cn

BV. These results indicate that a better prediction of genetic values is intended, when the dominant variance are large just as low-heritability traits.

Key words: genome selection; Bayesian model; dominant effects; dominant coefficients

基因组选择就是利用基因组范围的单核苷酸多态性 (Single nucleotide polymorphisms, SNPs) 来估计育种值并进行个体选择。基因组选择利用的是连锁不平衡信息,即假设 QTLs 和标记连锁不平衡,那么,由 QTL 解释的遗传变异就可以由其相邻标记捕获,通过估计标记效应就可以得到个体的估计育种值。T. H. E. Meuwissen 等在提出基因组选择概念的同时也提出了几个用于基因组选择的统计方法, BayesA 和 BayesB 等^[1]的方法都基于一个加性模型,不同之处在于假设标记的加性效应服从不同的先验分布^[2-3]。

如果是估计育种值,那么通常只考虑加性效应,但是显然生物作用模式有异于纯粹的加性模型的假设,为更好地了解复杂性状的遗传结构可能还需要加入显性效应和上位效应。一般对遗传效应的估计,尤其是加性效应,在模型中考虑非加性效应的时候更有利于其准确的估计^[4]。比如, G. Su 等在猪的平均日增重 (30~100 kg) 的全基因组选择研究中,显性方差为表型方差的 5.6%^[5],当排除非加性效应时,非加性效应的方差会被分配到模型中其他的方差组分中。也就是说在育种值估计中如果不考虑显性效应,会过高估计加性效应。J. E. de Almeida Filho 等研究认为,当显性方差与表型方差比率为 0.2 的时候,在基因组预测中包含显性效应的模型会显著优于加性模型^[6]。S. H. Lee 等研究显示,在贝叶斯模型中考虑显性效应会明显提高基因组选择的准确性,对于杂交小鼠群体的毛色性状,考虑显性效应后预测准确性可提高 2%^[7]。因此,基因组选择模型中有必要考虑显性效应,这样不仅可以提高基因组选择的准确性,还能够通过计算配合力对交配组合做出指导^[8-9]。

R. Wellmann 等对加性效应和显性效应的相关性进行了模拟研究^[10],本试验利用该研究结果,采用两种策略:一种假设加性效应和显性效应相互独立;另一种假设显性系数 (Dominant coefficients) 与加性效应的绝对值相互独立,并且显性系数服从正态分布。分别称之为 BayesAD1 和 BayesAD2,构建马尔科夫链通过 MH (Metropolis-Hastings) 抽样和 Gibbs 抽样求解得到标记的加性效应和显性效

应,估计基因组育种值。研究各模型在不同性状遗传结构、不同群体结构下,显性效应对基因组估计育种值的影响。

1 材料与方法

1.1 统计模型

首先,定义 a 和 d 分别为 QTL 的加性效应和显性效应。对于二等位基因的 QTL,等位基因分别编码为 0 和 1,等位基因 1 的基因频率为 p ,则另一等位基因的基因频率为 $1-p$ 。那么对于 3 种不同的基因型 00,01 和 11,它们的基因型值分别为 0, $a+d$ 和 $2a$ ^[11]。

使用如下线性模型^[12]进行基因组选择:

$$y = Xb + Za + Wd + e \quad (1)$$

其中, y 表示 $n \times 1$ 的表型向量 (n 为观察到的表型数), b 为 $k \times 1$ 表示固定效应向量, X 为 b 的关联矩阵,维数为 $n \times k$; a 和 d 分别是加性效应向量和显性效应向量,维数均为 $m \times 1$ (m 为标记数), Z 和 W 分别为它们的关联矩阵,大小为 $n \times m$; e 为残差效应向量。

根据上述模型,个体的基因组估计基因型值 (GEGV)、基因组估计育种值 (GEBV) 和显性离差 (GEDV) 可以表示为^[13-14]:

$$GEGV = \sum_{j=1}^m (\hat{a}_j + (2 - x_j) \hat{d}_j) x_j \quad (2)$$

$$GEBV = \sum_{j=1}^m (\hat{a}_j + (q_j - p_j) \hat{d}_j) (x_j - 2p_j) \quad (3)$$

$$GEDV = \sum_{j=1}^m -\hat{d}_j x_j (x_j - 1 - 2p_j) - 2p_j^2 \hat{d}_j \quad (4)$$

其中, x_j 是第 j 个标记中等位基因 1 的拷贝数, \hat{a}_j 是第 j 个标记的加性效应, \hat{d}_j 是第 j 个标记的显性效应, p_j 是第 j 个标记中等位基因 1 的基因频率, q_j 是第 j 个标记中等位基因 0 的基因频率。

对于标记的加性效应和显性效应的独立性,一种假设是加性效应与显性效应相互独立;另一种是显性系数 δ (Dominant coefficients) 与加性效应的绝对值 $|a|$ 相互独立,并且显性系数服从正态分布^[15]。这里,显性系数为 $\delta_j = \frac{d_j}{|a_j|}$,如果 $\delta_j > 1$

或 $\delta_j < -1$ 为超显性或超隐性;如果 $\delta_j = 1$ 或 $\delta_j = -1$ 则为完全显性或隐性;如果 $-1 < \delta_j < 0$ 或 $0 < \delta_j < 1$ 则为不完全显性或隐性。

1.2 计算方法

根据 BayesAD 模型中各变量的层级依赖关系^[16],也可将模型描述如下:

$$| a_j | | \sigma_{aj}^2 \sim (iid) | N(0, \sigma_{aj}^2) | \quad j = 1, \dots, m \quad (5)$$

$$\sigma_{aj}^2 \sim \chi^{-2}(v_a, s_a^2) \quad (6)$$

$$d_j | | a_j |, \sigma_{aj}^2 \sim N(\mu_d(| a_j |), \sigma_d^2(| a |, \sigma_{aj}^2)) \quad (7)$$

$$b, \sigma_e^2 \sim uniform \quad (8)$$

其中,均值 $\mu_d(| a_j |) = E(d_j | | a_j |)$, 方差 $\sigma_d^2(| a |, \sigma_{aj}^2) = Var(d_j | | a_j |, \sigma_{aj}^2)$ 。

BayesAD0 表示加性模型,即模型中仅有加性效应,因此 $\mu_d(| a_j |) = 0$, $\sigma_d^2(| a |, \sigma_{aj}^2) = 0$, $var(a_j) = E(\sigma_{aj}^2) = s_{aj}^2 \frac{v}{v-2}$ 。

BayesAD1 模型假设加性效应和显性效应相互独立, $\mu_d(| a_j |) = \mu_d$, $\sigma_d^2(| a |, \sigma_{aj}^2) = \sigma_d^2$ 。

BayesAD2 模型假设显性系数和加性效应的绝对值相互独立,则 $\delta_j = \frac{d_j}{| a_j |} \sim N(\mu_\delta, \sigma_\delta^2)$, 因此, $\mu_d(| a_j |) = | a_j | \mu_\delta$, $\sigma_d^2(| a |, \sigma_{aj}^2) = a_j^2 \sigma_\delta^2$ 。

BayesAD 通过构建马尔科夫链 (Markov chain),从模型中各变量的完全条件后验分布中使用 Gibbs 抽样和 MH 抽样进行求解,经过循环迭代达到收敛。

1.3 数据模拟

使用全基因组遗传标记模拟程序 GPOPSIM^[17] 模拟一个群体的标记基因型和系谱。首先是系谱的模拟,给定有效群体数 $\delta_j = \frac{d_j}{| a_j |}$ 为 100,也就是初始群体。经过 1 000 个历史世代的传递后得到一个基础群体 generation0, generation0 中的雄性和雌性个体随机交配,每对亲本的基因组经过交换与重组过程产生相同后代数得到群体 generation1,规模扩大为 1 000 个个体。与产生 generation1 的过程相同,从中随机选择雄性 \times 雌性分别为 500×500 、 100×100 、 50×50 、 20×20 和 10×10 产生全同胞家系,作为 generation2。其中每个全同胞家系内个体数分别为 4、20、40、100 和 200,从每个全同胞家系随机抽取一半个体构成参考群体,剩余一半为验证群体。然后是标记基因型的模拟,设定基因组由 5

条染色体构成,每条染色体长度 1M(Morgan),每条染色体 3 000 个标记,设定突变率为 2.5×10^{-3} ^[18]。

上述方法生成全基因组遗传标记数据和系谱后,还需要参考群体的表型和验证群体的真实育种值(TBV)、显性离差(TDV)。根据广义遗传力将随机残差、真实育种值和显性离差加和得到表型值。

有研究表明,显性系数服从均值大于零的正态分布^[19-20]。本研究中设定显性系数 $\delta \sim N(0.2, 0.3^2)$, $a | \delta \sim N(0, \exp(3 * \delta))$, 抽样获得 QTL 的加性效应 a_j 和显性效应 d_j 后计算真实育种值(TBV)、真实显性离差(TDV)和真实基因型值(TGV)^[18,21-23]:

$$TBV = \sum_{j=1}^m (a_j + (q_j - p_j)d_j)(x_j - 2p_j) \quad (9)$$

$$TDV = \sum_{j=1}^m -d_j x_j (x_j - 1 - 2p_j) - 2p_j^2 d_j \quad (10)$$

$$TGV = \sum_{j=1}^m (a_j + (2 - x_j)d_j)x_j \quad (11)$$

加性方差为 $VA = \sum 2p_j q_j a^2$, 显性方差为 $VD = \sum (2p_j q_j d)^2$, α 为替代效应 $a_j + (q_j - p_j)d_j$, p_j 和 q_j 为 generation1 中第 j 个 QTL 的等位基因频率^[24]。第 i 个个体的表型值为:

$$P_i = TBV_i + DBV_i + e_i \quad (12)$$

e_j 服从正态分布 $N(0, (VA + VD)(1 - H^2)/H^2)$, 其中 H^2 是广义遗传力。

1.4 评价指标

准确性:即基因组估计育种值与真实值的皮尔逊相关系数,取值范围 $[-1, 1]$ 。

$$r(\hat{a}, a) = \frac{\text{cov}(GEBV, TBV)}{\sqrt{\text{Var}(GEBV)\text{Var}(TBV)}} \quad (13)$$

其中, $\text{cov}(GEBV, TBV)$ 为基因组估计育种值(GEBV)与真实育种值(TBV)之间的协方差, $\text{Var}(GEBV)$ 为基因组估计育种值的方差, $\text{Var}(TBV)$ 为真实育种值的方差。 $r(\hat{d}, d)$ 和 $r(\hat{g}, g)$ 通过相应的基因组估计显性离差和真实显性离差、基因组估计基因型值和真实基因型值的相关系数。

2 结果

2.1 标记效应的模拟

模拟数据中设定显性系数服从正态分布,给定显性系数后加性效应的条件分布为拉普拉斯分布,

加性效应绝对值 a_j 和显性系数 δ 的联合分布如图 1 所示。A. Caballero 等研究表明^[25],对于具有较大的加性效应的等位基因,更倾向于存在不完全显性或隐性,而且杂合子的基因型值通常在两种纯合子基因型平均值之上;而对于微效基因座,其显性系数具有高度变异性。本研究模拟数据中的标记效应的设定正是基于这种结果。

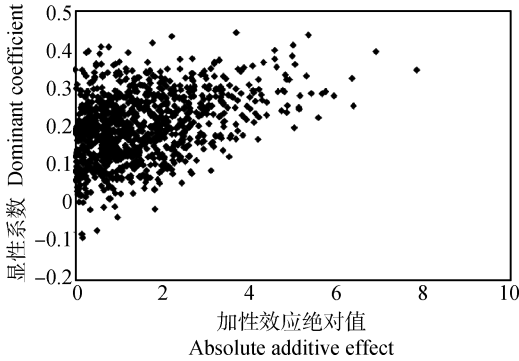
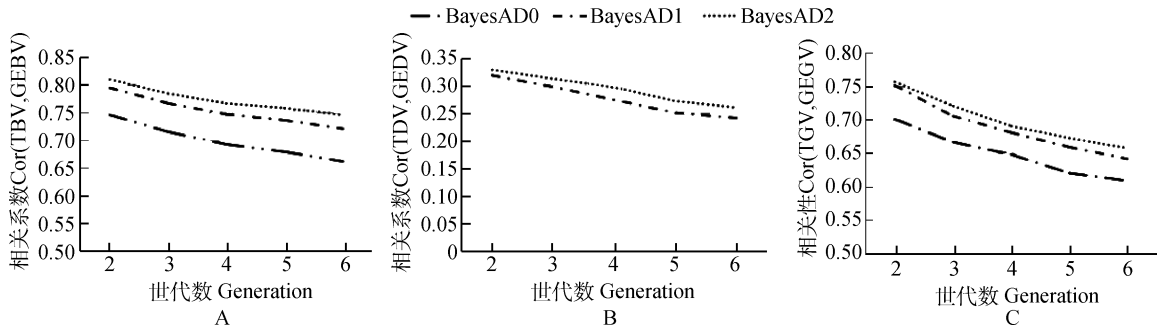


图 1 加性效应绝对值与显性系数的联合分布散点图
Fig. 1 Scatter plots of the considered joint distributions of $|a_j|$ and δ_j

2.2 世代间估计育种值的准确性

不同模型在第 2 到第 6 世代的基因组估计育种



A. 基因组估计育种值; B. 基因组估计显性离差; C. 基因组估计基因型值
A. Accuracy of GEBV; B. Accuracy of GEDV; C. Accuracy of GEGV

图 2 世代间基因组育种值、基因组显性离差及基因组基因型值估计准确性
Fig. 2 Accuracy of GEBV, GEDV and GEGV from generation 2 to 6 with different models

2.3 显性方差比重对育种值估计准确性的影响

全同胞数为 100 个, QTL 数为 10 个时,不同模型在不同遗传方差比重下的育种值估计准确性见表 1,因为没有考虑上位效应,因此本研究中遗传方差只包括加性方差和显性方差,模拟中显性方差占遗传方差的比重分别为 20%, 50% 和 80%。随着加性方差的比重升高,不同模型育种值估计的准确性均提高,在加性方差较小时, BayesAD2 有较大优势,分别比 BayesAD1 和 BayesAD0 准确性高 20.3% 和

值(GEBV),基因组显性离差(GEDV)和基因组基因型(GEGV)准确性的变化趋势列于图 2。模拟数据集参数:每个家系内全同胞个数为 50 个, QTL 数为 10 个,加性与显性方差比值为 3/1,狭义遗传力 0.3。

由图 2A 中可以看出,在各世代中, BayesAD2 估计准确性最高,其次是 BayesAD1,准确性最低的是 BayesAD0。也就是说,考虑显性效应后,育种值估计的准确性相对提高,在第 2 世代中, BayesAD1 较 BayesAD0 准确性高 6.6%, BayesAD2 较 BayesAD0 高 8.6%。从第 2 到第 6 世代, BayesAD0、BayesAD1 和 BayesAD2 估计基因组育种值准确性分别下降 0.089、0.075 和 0.064,准确性最高的 BayesAD2 随着世代数增加估计准确性降低的最少。

图 2B 显示了在各世代中基因组显性离差估计准确性的变化,与育种值估计的结果一致,从第 2 到第 6 世代, BayesAD1, BayesAD2 估计准确性均逐渐降低,估计显性离差准确性分别下降 0.077 和 0.068。

图 2C 显示了从第 2 到第 6 世代基因组基因型值估计的准确性, BayesAD0 估计的准确性最低, BayesAD1 和 BayesAD2 在第 2 世代准确性估计的差异很小。

28.4%。加性方差与显性方差相等时, BayesAD2 的优势减弱,分别比 BayesAD1 和 BayesAD0 准确性高 5.9% 和 8.6%。加性方差和显性方差比重为 0.4/0.1 时 3 种模型估计准确性相当。

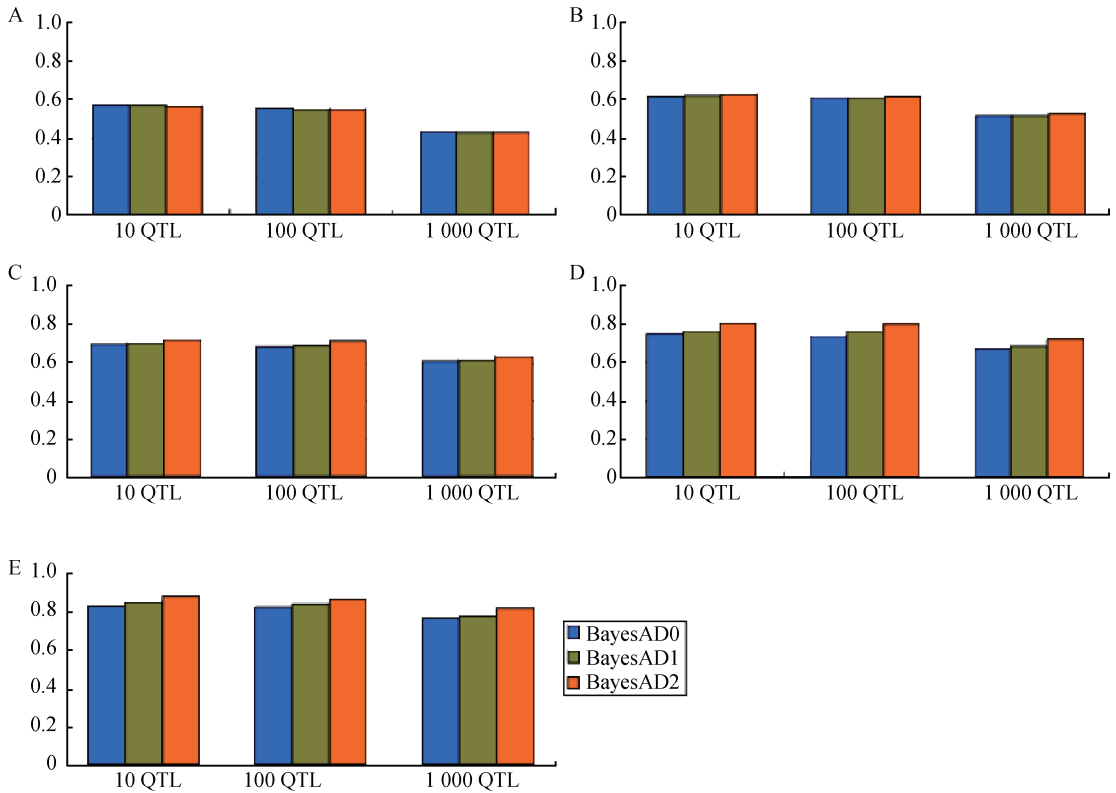
2.4 QTL 数和全同胞数对育种值估计准确性的影响

图 3 列出的是 3 种模型在全同胞数和 QTL 数不同时估计基因组育种值的准确性。全同胞数相同时, QTL 数目的增多使得育种值估计的准确性下降。不论全同胞数是多少,都可以观察到这个趋势。

表 1 加性方差与显性方差比重不同时育种值估计的准确性

Table 1 The predictive accuracy of estimated TBV($r(\hat{a}, a)$) with different dominant degree

模型 Model	VA/VD		
	VA/VD=0.1/0.4	VA/VD=0.25/0.25	VA/VD=0.4/0.1
BayesAD0	0.462(0.042)	0.746(0.019)	0.911(0.007)
BayesAD1	0.493(0.039)	0.765(0.021)	0.912(0.003)
BayesAD2	0.593(0.045)	0.810(0.019)	0.913(0.003)



A. 全同胞个体数为 4; B. 全同胞个体数为 10; C. 全同胞个体数为 40; D. 全同胞个体数为 100; E. 全同胞个体数为 200

A. 4 individuals in full-sibs family; B. 10 individuals in full-sibs family; C. 40 individuals in full-sibs family; D. 100 individuals in full-sibs family; E. 200 individuals in full-sibs family

图 3 全同胞家系内个体数与 QTL 数目不同时基因组育种值估计的准确性

Fig. 3 The predictive accuracy of GEBV for different size of full-sibs family and number of QTLs

在相同 QTL 时,全同胞个体数的增加会提高育种值估计的准确性,比如,当 QTL 数为 10,全同胞数为 4 时,BayesAD2 的 $r(\hat{a}, a)$ 为 0.567;当全同胞数为 200 时,相应的准确性则为 0.884, $r(\hat{a}, a)$ 提高了 0.317。

不同模型的表现是,全同胞数为 4 时,考虑显性效应后,对育种值估计并无提高。全同胞数为 20 时, BayesAD1 和 BayesAD0 准确性相当, BayesAD2 有轻微提高。全同胞数高于 40 时, BayesAD1 和 BayesAD2 开始表现出对育种估计准确性的提高,并且全同胞数的增多使得这种提

高更加明显, QTL 数为 10,全同胞数为 200 时, BayesAD1 比 BayesAD0 准确性提高 0.02, BayesAD2 提高了 0.053。全同胞个数较少时, BayesAD1 和 BayesAD2 之间也没有明显的差异,当全同胞数增多到 40 时, BayesAD2 的准确性开始高于 BayesAD1。

3 讨论

3.1 标记效应的先验分布

目前全基因组选择中常用的 Bayes 算法主要有 BayesA、BayesB 及 BayesCpi,这些算法的不同主要

在于标记效应的先验分布不同, BayesA 模型假设所有遗传标记或染色体片段都有效应, 且标记效应服从方差服从逆卡方分布的正态分布, 也就是大部分标记具有非常微小的效应值, 而只有少数标记有大效应值。 BayesB 模型假设只有少部分遗传标记或染色体片段具有效应, 大部分遗传标记或染色体片段没有效应, 也就是标记效应服从混合分布, 控制有效应标记的比例 π 在模型中预先设定, 该部分标记效应服从方差服从逆卡方分布的正态分布。 BayesCpi 模型与 BayesB 模型假设相似, 只是有效应标记的比例 π 在模型中求解, 且标记效应方差相同。 R. Wellmann 等的研究详细描述了考虑显性效应的 BayesB 模型^[10], 研究显示, 考虑显性效应能提高基因组育种值估计准确性。 本试验是在 BayesA 模型中考虑显性效应, 并研究不同遗传结构与种群结构下该模型估计基因组育种值的表现。

除了考虑标记效应的先验分布, 通过对加性效应和显性效应独立性的不同假设, 采用不同的标记效应模拟策略, 体现在显性相关系数 δ 与和加性效应绝对值 $|a|$ 的先验分布上。 那么在模型中就有一些参数需要提前设定, 也就是显性系数 δ 的均值 μ_δ 和方差 σ_δ 。 在模拟数据中我们可以通过模拟数据得到真实的显性系数分布中的参数, 那么将这个值作为显性相关系数 δ 先验分布中的参数, 将有利于得到更准确的后验分布抽样值, 从而获得更准确的估计育种值和显性离差, 但是如果是将 BayesAD1 和 BayesAD2 模型用于真实数据中, 就需要先估计 δ 的相关参数, 一个可选的方法就是先对标记的效应和表型进行相关性检验^[23], 得到的相关系数在很大程度上反映了加性效应和显性效应的关系, 利用这个信息估计模型中的两个超参, 并作为 MC-MC (Markov Chain Monte Carlo) 抽样中的初值会得到更可靠的后验分布抽样值。

另一个 BayesAD1 和 BayesAD2 方法中加性效应和显性效应的假设先验分布中的问题是, 我们对加性效应的假设的是加性效应绝对值 $|a|$ 的先验分布, 最终我们要得到的是加性效应值 a , 那么还需要假设效应的方向, BayesAD1 和 BayesAD2 方法中的假设是 $\text{sign}(a_j) = -\text{sign}(q_j - p_j) \text{sign}(d_j)$, 这个假设的合理性在于: 当 δ 与 $|a|$ 相互独立时, 加性效应的方向则与等位基因频率有关。 如果选择的是隐性等位基因, 那么选择效率通常较低, 因为隐性等位基因通常频率较低, 等位基因 1 如果是隐性的,

那么, 如果 $a > 0$ 则 $d < 0$, 如果 $a < 0$ 则 $d > 0$, 而且 $p < 0.5$, 那么可以得到 $\text{sign}(a_j) = -\text{sign}(q_j - p_j) \text{sign}(d_j)$; 如果选择的是显性等位基因, 等位基因 1 如果是显性的, 那么另一个等位基因频率就较低, 那么 $p > 0.5$, 如果 $a > 0$ 则 $d > 0$, 如果 $a < 0$ 则 $d < 0$, 同样得到 $\text{sign}(a_j) = -\text{sign}(q_j - p_j) \text{sign}(d_j)$ ^[10]。

3.2 影响育种值和显性离差估计准确性的因素

BayesAD0 在第 2 至第 6 世代育种值估计的准确性下降了 0.089, 而 BayesAD1 和 BayesAD2 分别下降了 0.075 和 0.064, 也就是说考虑显性效应可以减缓随着世代增加育种值估计准确性的下降程度, 说明加性效应和显性效应的确一起影响着育种值的估计。 另外, 相较育种值显性离差的估计准确性在世代间的降低要更缓慢一些, 因为我们模拟的群体是全同胞家系, 如果是在亲缘关系更远的群体中, 重组可能会使显性离差估计准确性下降的更快, 那么就需要和 QTL 强连锁不平衡的标记。 如果是使用真实数据, 那么就需要增加标记密度保证足够的标记与 QTL 连锁。

本试验研究了性状的加性方差与显性方差比值、QTL 数和全同胞家系大小对 BayesAD0、BayesAD1 和 BayesAD2 方法中估计育种值的影响, 结果显示, 显性方差的比重确实会影响育种值估计的准确性, 不考虑上位效应, 在相同的广义遗传力下, 显性方差比重越高, 在模型中考虑显性效应后对育种值估计的准确性影响越高。 S. H. Lee 等对杂交小鼠群体的 3 个遗传力大小不同的性状估计育种值, 结果也是狭义遗传力较小的性状, 考虑显性效应后对育种值估计准确性有较大的提高, 而对于狭义遗传力很高的性状, 考虑显性效应几乎不能影响育种值估计的准确性^[7]。

T. H. E. Meuwissen 等模拟了 20% 的遗传方差作为显性方差, 不考虑上位效应, 在 QTL 数为 3 时, 在 BayesB 模型中显性效应的加入使得育种值估计的准确性下降 0.132, QTL 数为 30 个时, 对育种值估计的准确性不再有影响, 他认为, 一个加性效应很大的 QTL, 如果解释的显性效应也很大, 那么显性方差会显著的引起基因组选择准确性下降, 但是如果显性方差分配给很多 QTLs, 则很难降低准确性^[26]。 在本研究中, 当全同胞数为 4 和 20, 模拟 10 个 QTLs 时考虑显性效应后育种值估计准确性并没有下降, 当全同胞数更多的时候, BayesAD1 和

BayesAD2 比 BayesAD0 育种值估计的准确性反而提高,那么可能是全同胞数对育种值估计准确性的影响弥补了 QTL 数较少时大的显性方差对准确性的降低。

随着 QTL 数目的增多, BayesAD0、BayesAD1 和 BayesAD2 育种值估计的准确性都有下降,这可能是与模拟的加性效应值有关,因为在模拟数据中假设加性效应 $a \mid \delta \sim N(0, \exp(3 * \delta))$, 这是一个比正态分布具有更重的两尾的分布,也就是说 $|a|$ 为极大值的概率更大,而 T. H. E. Meuwissen 等的结果显示, BayesA 和 BayesB 模型能够估计一些大效应 QTL(这里指方差占遗传方差超过 10% 的 QTL) 的位置,对于效应更小的位点则很难识别,但对于 BayesA 模型,更小的 QTL 在所有可能的位置上的后验概率很低,但不是 0,所以小效应的标记仍然有助于估计总的遗传方差^[26]。因此, BayesA 对 QTL 数目较少时也就是单个 QTL 解释的方差较大的遗传结构中育种值估计的准确性更高,随着单个 QTL 方差减小,育种值估计准确性随之下降。

即使加性方差和显性方差相同,对育种值估计的准确性高于显性离差估计的准确性,说明对显性离差的估计要比育种值估计更难^[13,27],如果要达到和育种值估计相当的准确性可能需要更多的标记, R. Wellmann 等模拟了不同标记密度对显性离差和基因型值估计的影响,增大标记密度来提高 QTL 和标记的连锁不平衡,当 QTL 和标记间的最大的 r^2 从 0.49 提高的 0.80 时,估计育种值的准确性只有轻微提高,而显性离差和基因型值估计准确性表现出了显著提高^[10]。

参考文献 (References):

- [1] MEUWISSEN T H E, HAYES B J, GODDARD M E. Prediction of total genetic value using genome-wide dense marker maps [J]. *Genetics*, 2001, 157(4): 1819-1829.
- [2] GIANOLA D, DE LOS CAMPOS G, HILL W G, et al. Additive genetic variability and the Bayesian Alphabet [J]. *Genetics*, 2009, 183(1): 347-363.
- [3] 王重龙, 丁向东, 刘剑锋, 等. 基因组育种值估计的贝叶斯方法[J]. *遗传*, 2014, 36(2): 111-118. WANG C L, DING X D, LIU J F, et al. Bayesian methods for genomic breeding value estimation[J]. *Hereditas (Beijing)*, 2014, 36(2): 111-118. (in Chinese)
- [4] BOLORMAA S, PRYCE J E, ZHANG Y, et al. Non-additive genetic variation in growth, carcass and fertility traits of beef cattle [J]. *Genet Sel Evol*, 2015, 47: 26.
- [5] SU G, CHRISTENSEN O, OSTERSEN T, et al. Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers [J]. *PLoS One*, 2012, 7(9): e45293.
- [6] DE ALMEIDA FILHO J E, GUIMARAES J F R, E SILVA F F, et al. The contribution of dominance to phenotype prediction in a pine breeding and simulated population [J]. *Heredity*, 2016, 117(1): 33-41.
- [7] LEE S H, VAN DER WERF J, HAYES B J, et al. Predicting unobserved phenotypes for complex traits from whole-genome SNP data [J]. *PLoS Genet*, 2008, 4(10): e1000231.
- [8] GUO X, CHRISTENSEN O F, OSTERSEN T, et al. Genomic prediction using models with dominance and imprinting effects for backfat thickness and average daily gain in Danish Duroc pigs [J]. *Genet Sel Evol*, 2016, 48(1): 67.
- [9] DOS SANTOS J P R, VASCONCELLOS R C D C, PIRES L P M, et al. Inclusion of dominance effects in the multivariate GBLUP model [J]. *PLoS One*, 2016, 11(4): e0152045.
- [10] WELLMANN R, BENNEWITZ J. Bayesian models with dominance effects for genomic evaluation of quantitative traits [J]. *Genet Res (Camb)*, 2012, 94(1): 21-37.
- [11] ZENG Z B, WANG T, ZOU W. Modeling quantitative trait loci and interpretation of models [J]. *Genetics*, 2005, 169(3): 1711-1725.
- [12] HENDERSON C. Best linear unbiased prediction of nonadditive genetic merits [J]. *J Anim Sci*, 1985, 60: 111-117.
- [13] DENIS M, BOUVET J M. Efficiency of genomic selection with models including dominance effect in the context of Eucalyptus breeding [J]. *Tree Genet Genomes*, 2013, 9(1): 37-51.
- [14] VITEZICA Z G, VARONA L, LEGARRA A. On the additive and dominant variance and covariance of individuals within the genomic selection scope [J]. *Genetics*, 2013, 195(4): 1223-1230.
- [15] WELLMANN R, BENNEWITZ J. The contribution of dominance to the understanding of quantitative genetic variation [J]. *Genet Res (Camb)*, 2011, 93(2):

- 139-154.
- [16] FINLEY A O, BANERJEE S, WALDMANN P, et al. Hierarchical spatial modeling of additive and dominance genetic variance for large spatial trial datasets [J]. *Biometrics*, 2009, 65(2):441-451.
- [17] ZHANG Z, LIU J, DING X, et al. Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix [J]. *PLoS One*, 2010, 5(9):e12648.
- [18] WITTENBURG D, MELZER N, REINSCH N. Including non-additive genetic effects in Bayesian methods for the prediction of genetic values based on genome-wide markers [J]. *BMC Genet*, 2011, 12:74.
- [19] BENNEWITZ J, MEUWISSEN T H E. The distribution of QTL additive and dominance effects in porcine F2 crosses [J]. *J Anim Breed Genet*, 2010, 127(3):171-179.
- [20] HAYES B J, GODDARD M. The distribution of the effects of genes affecting quantitative traits in livestock [J]. *Genet Sel Evol*, 2001, 33(3):209-229.
- [21] NISHIO M, SATOH M. Including dominance effects in the genomic BLUP method for genomic evaluation [J]. *PLoS One*, 2014, 9(1):e85792.
- [22] LEE S H, VAN DER WERF J H. Using dominance relationship coefficients based on linkage disequilibrium and linkage with a general complex pedigree to increase mapping resolution [J]. *Genetics*, 2006, 174(2):1009-1016.
- [23] ZENG J, TOOSI A, FERNANDO R L, et al. Genomic selection of purebred animals for crossbred performance in the presence of dominant gene action [J]. *Genet Sel Evol*, 2013, 45:11.
- [24] ERTL J, LEGARRA A, VITEZICA Z G, et al. Genomic analysis of dominance effects on milk production and conformation traits in Fleckvieh cattle [J]. *Genet Sel Evol*, 2014, 46:40.
- [25] CABALLERO A, KEIGHTLEY P D. A pleiotropic nonadditive model of variation in quantitative traits [J]. *Genetics*, 1994, 138(3):883-900.
- [26] MEUWISSEN T H E, GODDARD M. Accurate prediction of genetic values for complex traits by whole-genome resequencing [J]. *Genetics*, 2010, 185(2):623-631.
- [27] DUANGJINDA M, BERTRAND J, MISZTAL I, et al. Estimation of additive and nonadditive genetic variances in Hereford, Gelbvieh, and Charolais by method R [J]. *J Anim Sci*, 2001, 79(12):2997-3001.

(编辑 郭云雁)