

系谱错误对 ssGBLUP 估计育种值可靠性的影响

洪一峰¹, 蔡更元^{1,2}, 吴珍芳^{1,2}

(1 国家生猪种业工程技术研究中心/华南农业大学 动物科学学院, 广东 广州 510642;
2 广东温氏种猪科技有限公司, 广东 新兴 527400)

摘要:本研究旨在探讨系谱错误对 ssGBLUP 估计育种值 (EBV) 可靠性的影响程度, 研究不同遗传力、参考群、系谱错误下 ssGBLUP、GBLUP 和 BLUP 对 EBV 可靠性的影响。使用 QMSim 软件进行数据模拟, BLUP90 软件进行数据分析。结果表明, 遗传力的高低与系谱错误率对 ssGBLUP 预测 EBV 可靠性的趋势无关。当参考群逐渐增大时, ss-GBLUP 对系谱错误的容忍程度逐渐降低; 随着系谱错误率增大, ssGBLUP 估计育种值的可靠性呈线性下降趋势。

关键词:遗传力; 参考群; 系谱错误; 估计育种值

Effects of genealogical errors on the reliability of ssGBLUP to estimate breeding values

HONG Yifeng¹, CAI Gengyuan^{1,2}, WU Zhenfang^{1,2}

(1 National Engineering Research Center for Breeding Swine Industry/College of Animal Science, South China Agriculture University, Guangzhou 510642, China; 2 Guangdong Wens Pig Breeding Technology Co., Ltd., Xinxing 527400, China)

Abstract: This study was aimed to study the influence of genealogical errors on the reliability of estimated breeding values (EBV) obtained by ssGBLUP. Under the condition of different heritabilities, reference groups and genealogical errors, effects of ssGBLUP, GBLUP and BLUP on the reliability of EBV were evaluated. Data were simulated using QMSim software and analyzed by BLUP90 software. The results showed that the heritability degree and genealogical error rate had nothing to do with reliability of EBV. The tolerance of ssGBLUP to genealogical errors decreased gradually with the gradual increase of reference group. The reliability of EBV obtained by ssGBLUP decreased linearly with the increase of genealogical error rate.

Key words: heritability; reference group; genealogical error; estimated breeding value

最佳线性无偏预测 (Best linear unbiased prediction, BLUP) 法自提出到现在仍具有顽强的生命力。对于中等或高等遗传力性状, 当系谱信息较完整、表型信息较准确时, BLUP 法能准确估计育种值, 从而显著提高遗传进展。随着商业化高密度基因芯片的不断普及以及高通量测序价格的下降, Meuwissen 等^[1]于 2001 年首次提出基因组选择 (Genomic se-

lection, GS) 方法。近年来, GS 已逐渐应用于畜禽^[2]、作物^[3]和水产生物^[4]生产中。GS 可以实现早期选种, 缩短世代间隔, 且对低遗传力的性状具有优势。一般而言, 基因组预测育种值 (Genomic estimated breeding value, GEV) 的准确性高于利用系谱信息预测的 EBV^[5-6]。相比于传统 BLUP 法, GBLUP (Genomic BLUP, GBLUP) 法通过构建基因组关

收稿日期: 2019-03-28
作者简介: 洪一峰 (1995—), 男, 硕士研究生, E-mail: 2846588197@qq.com; 通信作者: 吴珍芳 (1970—), 男, 教授, 博士, E-mail: wzfemail@163.com
基金项目: 广东省自然科学基金 (2018B030313011); 广东省现代农业产业技术体系生猪创新团队项目 (2019KJ126)

系矩阵(\boldsymbol{G} 矩阵)来替换基于系谱信息构建的分子血缘关系矩阵(\boldsymbol{A} 矩阵),而一步法 GBLUP(Single-step GBLUP,ssGBLUP)结合系谱与 SNP 标记信息共同构建 \boldsymbol{H} 矩阵来替换 \boldsymbol{A} 矩阵。在生产过程中,ssGBLUP 可充分利用已有表型记录、基因组信息和系谱记录等历史数据,会显著提高 GEBV 的准确性。Fragomei 等^[7]与 Christensen 等^[8]证明 ssGBLUP 法对估计育种值的准确性最高。因此,目前在 GS 应用中估计育种值的主流方法是 ssGBLUP^[9]。

在生产过程中由于种种原因会导致系谱错误,例如,精液的记录混乱、耳标的缺失、饲养员寄养问题等等。对于有错误的系谱,我们很难及时对其进行纠正,而系谱错误会严重影响 ssGBLUP 估计育种值的准确性,从而降低遗传进展和生产效益。因此,本研究的目的在于探讨系谱错误对 ssGBLUP 估计育种值的影响,当系谱错误率达到一定的比例时,可能 ssGBLUP 估计育种值的准确性会低于 GBLUP 的。因此,有必要对系谱错误率与 EBV 准确性的关系进行摸索,以期为今后 GS 育种提供一定的科学依据和参考。

1 研究方法

1.1 数据来源

本研究利用 QMSim 软件进行数据模拟。QM-Sim 被设计用来模拟家畜的多种遗传结构和种群结构,可以模拟大规模的基因分型数据和复杂的系谱。模拟数据分 2 步进行,第一,模拟历史世代来创建理想的 LD(Linkage disequilibrium)水平,建立遗传突变、漂移等;第二,生成用于进行数据分析和验证的群体^[10]。

1.1.1 历史群的模拟 为了建立初始的 LD 以及实现突变-漂移平衡,进行历史种群的模拟,在历史种群无选择、无迁移、随机交配的前提下,通过突变产生新的变异,进而完成历史种群的模拟。历史群体模拟参数如表 1 所示。

表 1 历史群体的模拟	
参数	设定值
初始有效群体含量	100
世代数	1 000
公母比例	1 : 1
第 1 000 世代群体数	2 000

1.1.2 基础群和基因组参数的模拟 在历史种群的最后一代中,随机挑选部分个体作为最近种群的创建者,然后以逐渐增大最后保持不变的速度进行扩群,详细参数见表 2。基因组参数的模拟详见表 3。

1.2 数据处理

对最终产生的 20 444 个基础群个体进行处理,用于产生参考群体与估计群体。选取最后 15 000 个个体,使其具有完整的系谱信息与表型信息,分别选取最后 1 500、2 500、3 500、4 500 和 5 500 个个体使其具有基因型信息,并将最后 500 个个体表型信息设置为缺失值,用于估计群体。

表 2 基础群体的模拟	
参数	设定值
初始群体公母数	200 (200)
世代数	16
各世代公母比例	1 : 1
选配方式	随机交配
选种方式	随机选种
群体增长率	0.2 ~ 0.8 (逐渐增长)
模拟重复次数	10

表 3 基因组参数的模拟	
参数	设定值
染色体条数	19
染色体长度/cM	2 603
标记密度/cM	23.05
QTLs 密度/cM	0.26
标记位置	随机分布
QTLs 位置	随机分布
标记等位基因数	2
QTLs 等位基因数	2
标记突变率	2.5×10^{-3}
QTLs 突变率	2.5×10^{-5}
干涉 ¹⁾	25

1)干涉现象:在减数分裂时,非姐妹染色单体的交叉互换会影响到相邻等位基因之间的关联水平和 LD 程度,一个单次交换会影响临近位点发生交换的概率

将 15 000 个个体的系谱信息分别按照 4% , 8% , 12% , 16% , 20% , 24% , 28% , 32% , 36% , 40% 的错误率来将一些个体的系谱随机与附近个体的系谱进行交换,从而得到错误的系谱。

1.3 分析模型

$$\begin{aligned} Y &= \boldsymbol{Xb} + \boldsymbol{Za} + \boldsymbol{e}, \\ E(Y) &= \boldsymbol{Xb}, \\ E(\boldsymbol{a}) &= E(\boldsymbol{e}) = 0, \\ \text{var}(\boldsymbol{a}) &= \boldsymbol{G}, \\ \text{var}(\boldsymbol{e}) &= \boldsymbol{R}, \text{cov}(\boldsymbol{a}, \boldsymbol{e}) = 0. \end{aligned}$$

式中, Y 是个体的表型值; \boldsymbol{b} 是固定效应向量,包括性别效应; \boldsymbol{a} 是动物所有随机效应(育种值)构成的向量; \boldsymbol{e} 是残差效应所构成的向量; \boldsymbol{X} 、 \boldsymbol{Z} 分别为 \boldsymbol{b} 、 \boldsymbol{a} 的结构矩阵, \boldsymbol{G} 和 \boldsymbol{R} 是已知的正定矩阵。

1.4 统计分析

将处理过的数据进行整理,按照 BLUPf90 软件

的要求将其整理为数据文件和系谱文件。

BLUP90 是一个全面的集合程序,主要用于动物育种,该程序可以使用多种方法进行方差组分和育种值估计^[11],本研究利用 AIREMLF90 程序来估计育种值。

Henderson 在 1984 年提出用理论可靠性(Theoretical Reliability, REL)来评价估计育种值的可靠性,其计算方法为:

$$REL = 1 - SE^2/Va,$$

式中,SE 为个体 EBV 的标准误差;Va(加性遗传方差)由约束性最大似然(REML)方法进行估计。

2 结果与分析

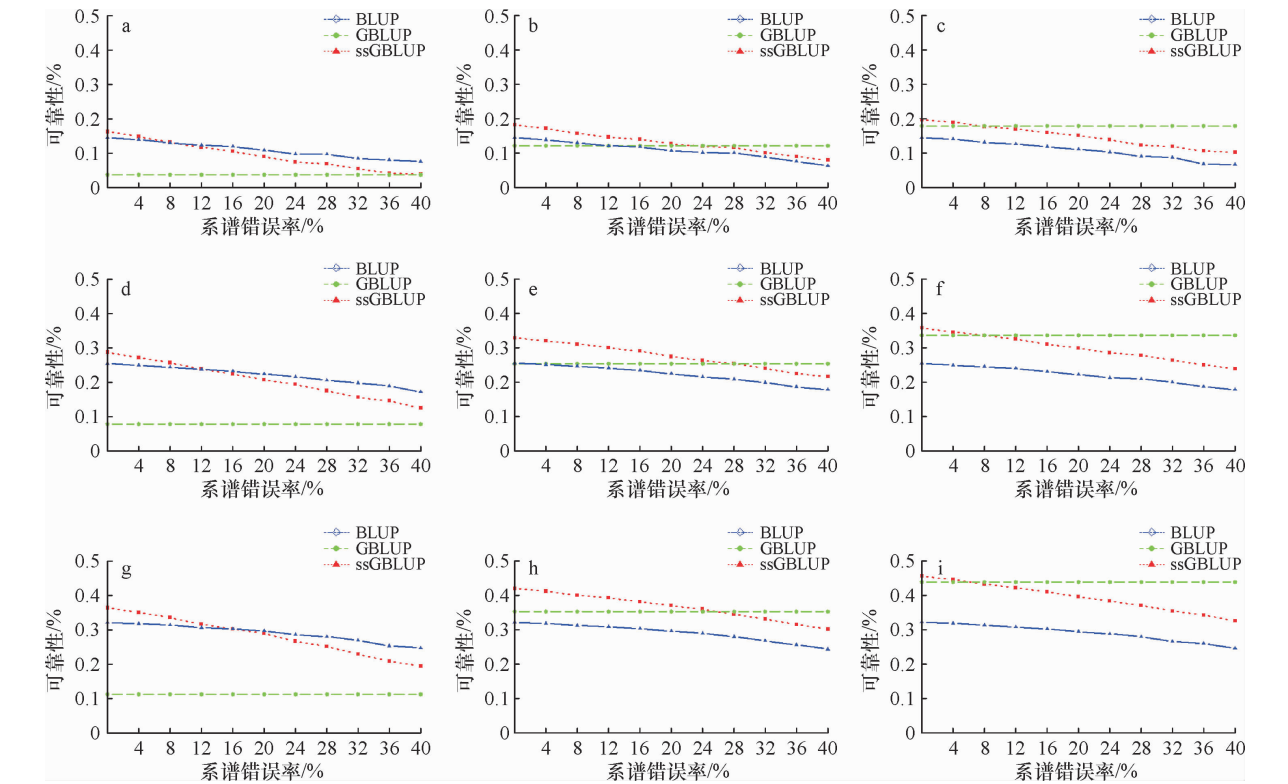
2.1 在不同参考群下,系谱错误对 ssGBLUP 估计育种值可靠性的影响

由图 1a、图 d 和图 1g 可知,当参考群个体数量为 1 000 时,GBLUP 估计育种值的可靠性比较低,甚至低于 BLUP 预测的可靠性,ssGBLUP 与 BLUP 估计育种值的可靠性随着系谱错误率的提高逐渐下

降,且 ssGBLUP 下降的趋势更加明显;当参考群个体数量为 3 000 时(图 1b、图 1e、图 1h),GBLUP 估计育种值的可靠性明显得到显著提高,BLUP 估计育种值的可靠性相对于参考群为 1 000 时基本不变,且当系谱错误率达到 24% 左右时,ssGBLUP 估计育种值的可靠性会低于 GBLUP 的;当参考群个体数量为 5 000 时(图 1c、图 1f、图 1i),GBLUP 估计育种值的可靠性基本接近与 ssGBLUP 的可靠性,且当系谱错误率达到 8% 左右时,ssGBLUP 估计育种值的可靠性会低于 GBLUP 的可靠性。

2.2 在不同遗传力下,系谱错误对 ssGBLUP 估计育种值可靠性的影响

由图 1 可知,通过纵向对比,在 3 种遗传力(H^2)作用下,随着系谱错误的不断增多,ssGBLUP 估计育种值的可靠性(红色线条)与 GBLUP 估计育种值的可靠性(绿色线条)的交叉点基本一致。结果表明当参考群数量相等,遗传力的不同与系谱错误率对 ssGBLUP 估计育种值的趋势影响较小。



a ~ c $H^2 = 0.1$, d ~ f $H^2 = 0.3$, g ~ i $H^2 = 0.5$; a、d、g 参考群个体数量为 1 000, b、e、h 参考群个体数量为 3 000, c、f、i 参考群个体数量为 5 000

图 1 系谱错误对 ssGBLUP 估计育种值可靠性的影响

3 讨论与结论

3.1 参考群的大小与系谱错误率对 ssGBLUP 估计育种值的影响

ssGBLUP 估计育种值的可靠性随着系谱错误的

增大逐渐降低,两者呈线性关系,GBLUP 估计育种值的可靠性保持不变。由于 GBLUP 没有运用系谱信息,因此其不会因系谱错误而影响可靠性。当参考群较小时,GBLUP 估计育种值效果极差^[12],而 ss-GBLUP 因为可以充分利用系谱,其预测的可靠性会

略高于 BLUP,但随着参考群数量的逐渐增多,GBLUP 与 ssGBLUP 估计育种值的可靠性均有明显增加,但 GBLUP 增加的幅度更大,因此,当参考群逐渐增多时,ssGBLUP 对系谱错误的容忍程度逐渐降低。由此得到的启发是,在生产过程中,当参考群足够大时,GBLUP 估计育种值的可靠性基本接近于 ssGBLUP 的可靠性,所以我们可以考虑使用 GBLUP 来估计育种值以及进行多方面考虑分析。

3.2 遗传力的高低与系谱错误率对 ssGBLUP 估计育种值的影响

随着遗传力的增高,BLUP、GBLUP 和 ssGBLUP 估计育种值的可靠性均明显升高,但由于 ssGBLUP 与 GBLUP 增加的幅度基本一致,所以遗传力的高低并不会影响 ssGBLUP 对系谱错误的容忍程度。根据可靠性公式,遗传力会影响加性遗传方差(V_a),对 BLUP、GBLUP 和 ssGBLUP 均有影响,影响程度可能不一致,故遗传力的高低与系谱错误对 ssGBLUP 估计育种值的可靠性的趋势基本没有影响。

3.3 系谱错误的处理

杜绝系谱错误无疑是最好的,在现有条件下,可通过加强管理,严格做到先打耳标再调栏,使用电子耳标与普通耳标相结合来减少系谱错误的发生等。未来,随着猪脸识别技术的不断发展^[13],发生系谱错误的概率将逐渐降低,甚至没有。当发生系谱错误的时候,尽管很难进行纠正,但依然可以找到一些办法,例如,在现在批次化生产过程中,每一批猪只的父本及母本是确定的,根据基因组信息可以进行亲子鉴定,从而纠正其部分系谱。

3.4 结论

本研究分析显示,遗传力的高低与系谱错误对 ssGBLUP 估计育种值的趋势基本没有影响,当参考群逐渐增大时,ssGBLUP 对系谱错误的容忍程度逐渐降低。当参考群个体为 3 000 时,系谱错误率达到 24% 左右时,ssGBLUP 估计育种值的可靠性会低于 GBLUP 的;当参考群个体为 5 000 时,系谱错误率达到 8% 左右时,ssGBLUP 估计育种值的可靠性会低于 GBLUP 的。

参考文献:

[1] MEUWISSEN T H, HAYES B J, GODDARD M E. Pre-

diction of total genetic value using genome-wide dense marker maps[J]. *Genetics*,2001,157(4):1819-1829.

- [2] ROTHSCCHILD M F, PLASTOW G S. Applications of genomics to improve livestock in the developing world[J]. *Livest Sci*, 2014, 166:76-83.
- [3] HEFFNER E L, SORRELLS M E, JANNINK J L. Genomic selection for crop improvement [J]. *Crop Sci*, 2017, 49(1):1-12.
- [4] TORO M A, SAURA J M, FERNANDEZ B, et al. Accuracy of genomic within - family selection in aquaculture breeding programmes[J]. *J Anim Breed Genet*,2017,134(3):256-263.
- [5] OLSON K M, VANRADEN P M, TOOKER M E, et al. Differences among methods to validate genomic evaluations for dairy cattle [J]. *J Dairy Sci*, 2011, 94(5):2613-2620.
- [6] PSZCZOLA M, CALUS M P L. Updating the reference population to achieve constant genomic prediction reliability across generations[J]. *Animal*, 2015, 10(6):7.
- [7] FRAGOMENI B O, LOURENCO D A L, TSURUTA S, et al. Use of genomic recursions and algorithm for proven and young animals for single-step genomic BLUP analyses; a simulation study[J]. *J Anim Breed Genet*, 2015, 132(5):340-345.
- [8] CHRISTENSEN O F, MADSEN P, NIELSEN B, et al. Single-step methods for genomic evaluation in pigs[J]. *Animal*, 2012,10(6):1565-1571.
- [9] LEGARRA A, AGUILAR I, MISZTAL I. A relationship matrix including full pedigree and genomic information [J]. *J Dairy Sci*, 2009, 92(9):4656-4663.
- [10] SARGOLZAEI M, SCHENKEL F S. QMSim: A large-scale genome simulator for livestock[J]. *Bioinformatics*, 2009, 25(5):680-681.
- [11] MISZTAL I, TSURUTA S, STRABEL T, et al. BLUPF90 and related programs (BGF90)[C]//Proceedings of the 7th world congress on genetics applied to livestock production. 2002, 33:743-744.
- [12] YAMAGUCHI N, MACDONALD D W. The burden of Co-Occupancy: Intraspecific resource competition and spacing patterns in American mink, *Mustela vison*[J]. *J Mammal*, 2003, 84(4):1341-1355.
- [13] 赵静. 猪脸识别技术帮助金融机构获得牲畜“身份证”[J]. *农业知识*, 2018(2):37.