

基因组选择技术在农业动物育种中的应用

谈成^{1,2,3}, 边成¹, 杨达⁴, 李宁¹, 吴珍芳^{2,3}, 胡晓湘¹

1. 中国农业大学, 农业生物技术国家重点实验室, 北京 100193;
2. 华南农业大学动物科学学院, 国家生猪种业工程技术研究中心, 广州 510642;
3. 广东温氏食品集团股份有限公司, 云浮 527400;
4. 明尼苏达大学动物科学系, 圣保罗 MN55108

摘要: 基因组选择(genomic selection, GS)是畜禽经济性状遗传改良的重要方法。随着高密度 SNP 芯片和二代测序价格的下降, GS 技术越来越多被应用于奶牛、猪、鸡等农业动物育种中。然而, 降低全基因组 SNP 分型成本、提高基因组育种值(genomic estimated breeding value, GEBV)估计准确性仍然是 GS 研究的主要难题。本文从全基因组 SNP 分型策略和 GEBV 估计模型两个方面进行了综述, 并对目前 GS 技术在主要畜禽品种中的应用现状进行了介绍, 以期 GS 在农业动物育种中的深入开展提供借鉴和参考。

关键词: 基因组选择; 动物育种; 基因分型; 二代测序

Application of genomic selection in farm animal breeding

Cheng Tan^{1,2,3}, Cheng Bian¹, Da Yang⁴, Ning Li¹, Zhenfang Wu^{2,3}, Xiaoxiang Hu¹

1. State Key Laboratory of Agricultural Biotechnology, China Agricultural University, Beijing 100193, China;
2. National Engineering Research Center for Breeding Swine Industry, College of Animal Science, South China Agricultural University, Guangzhou 510642, China;
3. Guangdong Wens Foodstuff Group Co., Ltd., Yunfu 527400, China;
4. Department of Animal Science, University of Minnesota, Saint Paul MN55108, USA

Abstract: Genomic selection (GS) has become a widely accepted method in animal breeding to genetically improve economic traits. With the declining costs of high-density SNP chips and next-generation sequencing, GS has been applied in dairy cattle, swine, poultry and other animals and gained varying degrees of success. Currently, major challenges in GS studies include further reducing the cost of genome-wide SNP genotyping and improving the predictive accuracy of genomic estimated breeding value (GEBV). In this review, we summarize various methods for genome-wide SNP genotyping and GEBV prediction, and give a brief introduction of GS in livestock and poultry breeding. This review will provide a reference for further implementation of GS in farm animal breeding.

Keywords: genome selection; animal breeding; genotyping; next-generation sequencing

收稿日期: 2017-08-29; 修回日期: 2017-10-20

基金项目: 农业部 948 项目(编号: 2012-G1(4))和国家高技术研究发展计划(863 计划)项目(编号: 2011AA100301)资助 [Supported by the 948 Program of the Ministry of Agriculture of China (No.2012-G1(4)) and the National High-Tech Research and Development Program of China (863 Program) (No.2011AA100301)]

作者简介: 谈成, 博士, 研究方向: 动物分子数量遗传学。E-mail: tancheng200508@163.com

通讯作者: 胡晓湘, 教授, 博士生导师, 研究方向: 生物化学与分子生物学。E-mail: huxx@cau.edu.cn;

吴珍芳, 教授, 博士生导师, 研究方向: 动物遗传育种与繁殖。E-mail: wzfemail@163.com

DOI: 10.16288/j.yczs.17-286

网络出版时间: 2017/11/13 11:28:12

URI: <http://kns.cnki.net/kcms/detail/11.1913.r.20171113.1128.006.html>

基因组选择(genomic selection, GS)是对传统遗传评估技术的一次重大革新,利用覆盖全基因组的高密度遗传标记计算个体的基因组估计育种值(genomic estimated breeding value, GEBV)^[1]。与传统基于系谱信息的估计育种值(estimated breeding value, EBV)相比,GEBV 通常能获得更高的估计准确性^[2-4]。由于 GEBV 计算可以不依赖系谱记录和表型信息,这就为早期选择提供了可能,可以大幅度缩短世代间隔,提高遗传进展,降低农业动物的育种成本^[5,6]。此外,对于传统育种受限的性状,如低遗传力的性状和难以测量的性状,GS 也更加具有优势。

GS 一般包括以下步骤(图 1):首先建立参考群体(reference population),参考群体中每个个体都有已知的表型和基因型,通过合适的统计模型可以估计出每个 SNP 或不同染色体片段的效应值;然后对候选群体(candidate population)每个个体进行基因分型,利用参考群体中估计得到的 SNP 效应值来计算

候选群体中每个个体的 GEBV;最后,根据 GEBV 排名对个体进行选留,待选留个体(selected candidates)完成性能测定后,这些个体又可以被放入参考群体,用于重新估计 SNP 的效应值,如此反复。

随着商业化高密度 SNP 芯片的普及和二代测序价格的下降,GS 越来越多被应用于除奶牛以外其他农业动植物的育种实践中,如猪(*Sus scrofa*)^[7,8]、鸡(*Gallus gallus*)^[9,10]、水产动物^[11,12]、水稻(*Oryza sativa*)^[13]、玉米(*Zea mays*)^[14]等。尽管如此,基因分型成本仍然是 GS 技术推广和应用的重大障碍,大多数育种企业限于长久的基因分型成本投入而不敢大规模应用。多数已开展 GS 研究的科研院所或企业也囿于基因分型方面的投入,不得不通过缩小参考群体规模或降低标记密度来降低成本。较小的参考群或较低的标记密度一定程度上影响了 GEBV 的准确性,从而低估了 GS 技术在动物育种中的潜力,反而阻碍了这一技术的推广和应用。因此,降低个

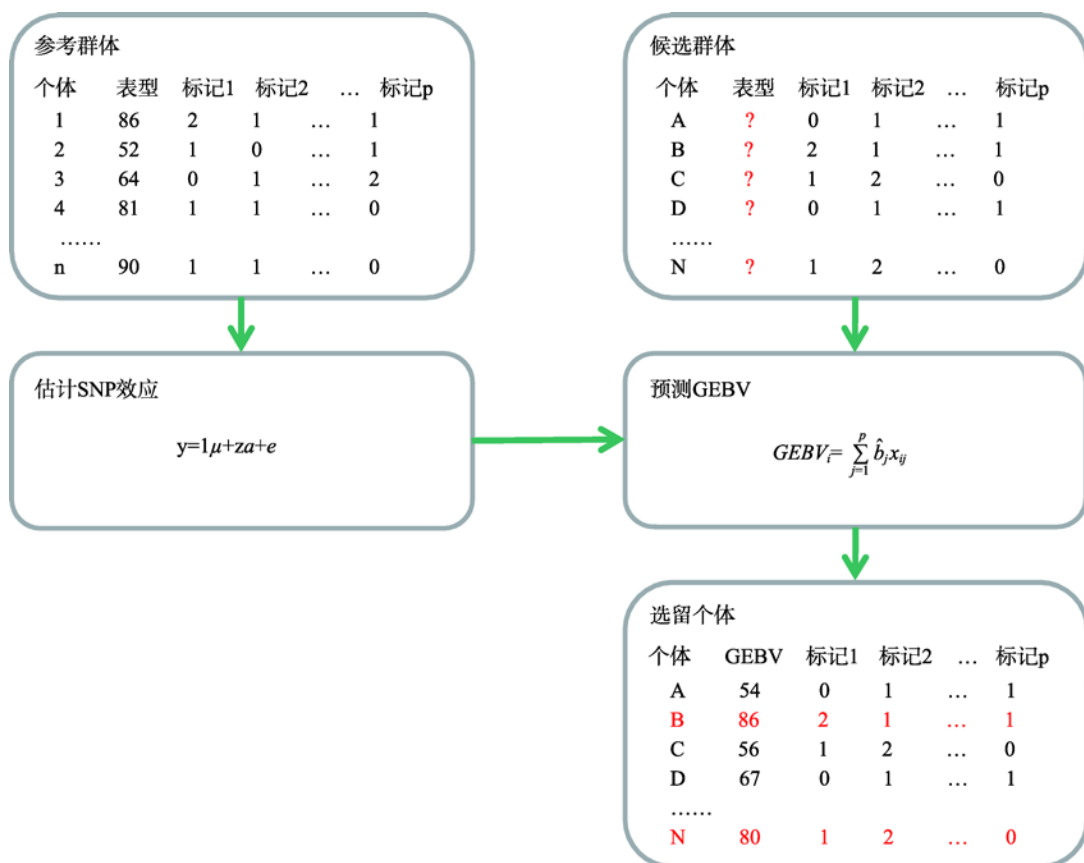


图 1 基因组选择实施示意图

Fig. 1 Schematic overview of genomic selection

体分型费用一直是 GS 的研究热点之一。

GS 技术的另一个重要研究方向是如何针对性状本身的遗传特点, 合理选择最优的模型和算法来提高 GEBV 估计准确性。常用的基于单点 SNP 标记的 GEBV 估计方法主要分两类: 一类是基于估计等位基因效应来计算 GEBV^[15], 如最小二乘法、随机回归-最佳线性无偏预测(ridge regression best linear unbiased prediction, RR-BLUP)、贝叶斯方法等; 第二类是采用 SNP 标记构建基因组关系矩阵(genomic relationship matrix, GRM 或 G 矩阵), 然后采用线性混合模型估计 GEBV^[16], 如基因组最佳线性无偏预测法(genomic best linear unbiased prediction, GBLUP)、一步法 GBLUP。这两类方法均有各自的优缺点, 对于不同性状会有不同的效果。随着基因分型个体数极大累积或标记密度极大提高, 传统基于单点 SNP 信息的统计模型无法利用所有个体表型数据或所有标记位点信息, 此时基于单倍型的 GEBV 预测模型可以解决这一问题。通过将已知功能的基因组学信息, 包括基因结构、甲基化区域、转录因子调控结合位点、选择信号候选区域等信息, 以单倍型信息整合进 GEBV 预测模型, 从而提高 GEBV 的准确性^[17]。

本文从全基因组 SNP 分型策略和 GEBV 估计模型两个方面进行了综述, 并对目前 GS 技术在主要

畜禽品种中的应用现状进行了介绍。

1 全基因组 SNP 分型技术的选择

1.1 高密度 SNP 芯片

SNP 标记作为第三代分子标记, 具有数量多、分布广、遗传稳定等特点, 被广泛应用于全基因组关联分析和基因组选择等领域。SNP 分型技术可分为两个时代: 凝胶电泳时代和高通量时代。凝胶电泳时代的主要方法包括限制性酶切片长度多态性分析、寡核苷酸连接分析、等位基因特异聚合酶链反应分析、单链构象多态性分析、变性梯度凝胶电泳分析等。虽然这些技术与高通量时代的技术原理大致一样, 但是由于它不能进行自动化, 只能进行小规模 SNP 分型, 所以逐渐被淘汰。高通量时代的 SNP 分型技术以高密度 SNP 芯片为代表, Illumina 公司 Infinium 技术和 Affymetrix 公司 Axiom 技术是目前最流行的高密度 SNP 芯片解决方案。

Illumina Infinium 芯片是基于微珠的 BeadArray 生物芯片, 首先在玻璃基片(称为 Wafer)上, 通过光蚀刻的方法蚀刻出许多个排列整齐的小孔, 每个小孔直径约为 3 μm , 这些小孔刚好可以容纳一个微珠(图 2A)。微珠是芯片的核心部分, 每个微珠的表面偶联了几十万条相同序列的 DNA 片段。这些 DNA

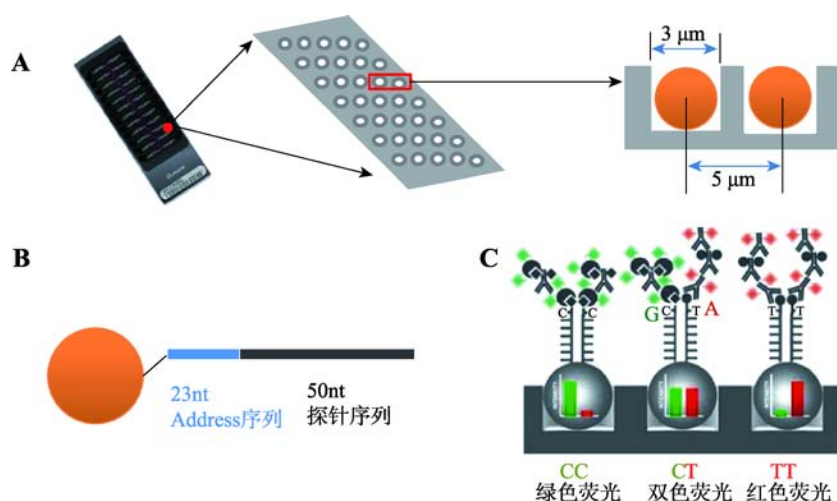


图 2 Illumina BeadArray 芯片结构示意图

Fig. 2 Illustration of Illumina BeadArray structures

A: 在玻璃基片上光蚀刻出直径约为 3 μm 的整齐小孔, 每个小孔刚好可以容纳一个微珠; B: 微珠表面偶联了几十万条相同序列的 DNA 片段, 靠近珠子一侧的 23nt 序列是用于标识微珠的 Address 序列, 而远离珠子端的 50nt 序列是待测 SNP 位点特异性的探针序列; C: 用绿色荧光标记 C 碱基, 用红色荧光标记 T 碱基, 以基因型为 CT 的 SNP 为例, 如果发出的光为绿色, 则是 CC 纯合子, 如果发出的光为红色, 则为 TT 纯合子, 如果绿色和红色荧光都有, 而且光强度差不多, 那么该位点则是 CT 杂合子。

10K、55K、62K、68K 和 650K, 其中 Illumina 公司 PorcineSNP60 v2 芯片(64 232 SNP)和 GeneSeek 公司的 PorcineSNP80 芯片(68 528 SNP)应用最为普遍。然而这些 SNP 芯片主要是根据国外瘦肉型猪种(如长白猪、大白猪、杜洛克猪、皮特兰猪等)的简化基因组测序数据而设计的^[18], 很多并不适用于我国地方品种。因此, 针对特定群体进行 GS 应用时, 自定义芯片或基于二代测序技术的全基因组 SNP 分型方法可能会是一个更好的选择。

1.2 简化基因组测序

与全基因组重测序相比, 简化基因组测序(reduced-representation genome sequencing, RRGs)只对很小一部分的基因组进行测序, 这不仅能降低测序成本, 而且能大幅降低测序数据的复杂性和生物信息分析所需的时间^[19]。已有大量文献报道了不同的简化基因组测序建库方法, 如 CroPS(complexity reduction of polymorphic sequences)^[20]、RRL(reduced representation libraries)^[21]、RAD-seq(restriction site associated DNA sequencing)^[22]、MSG(multiplexed shotgun genotyping)^[23]、GBS(Genotyping-by-Sequencing)^[24~27]、2b-RAD(double digest RAD-seq)^[28]、GGRS(genotyping by genome reducing and sequencing)^[29]等, 其原理均是通过合适的限制性内切酶对基因组进行酶切, 然后连接二代测序平台所需的接头序列, 再选择一定长度范围内的连接产物进行测序。GBS、ddGBS 和 RAD-seq 是目前最普及的 3 种简化基因组测序方法, 它们的主要区别在于是否在接头连接后进行片段选择, 以及 PCR 后是否进行片段选择^[30]。

Elshire 等^[24]首先提出的 GBS 的实验流程为: 利用限制性内切酶分别打断所有个体的基因组, 随后通过连接反应连接接头 (GBS 使用两类接头: 一类是含有一端测序结合位点以及“标签”的接头; 另一类是仅含有另一端测序结合位点的普通接头); 连接反应后会得到 3 种片段, 分别为两端均连接有“标签”接头的片段、两端均连有普通接头的片段以及一端连有“标签”接头而另一端连有普通接头的片段; 将连接后的所有样本片段混合, 只有片段长度小于 1000 bp 且两端分别连有“标签”接头和普通接头的片段才可以进行有效地扩增, 这一步骤实际上会丢

弃许多片段, 但保留下来的片段仍然可以满足大部分研究的需要; 最后对扩增后的片段进行测序。GBS 相较于其他简化基因组测序技术具有相对简单的建库步骤, 省去了包括机械随机打断、片段长度选择、末端修复等步骤, 在很大程度上节约了人力和物力, 可以在一次实验中对大量的样本进行同时测序。GBS 技术适合应用于有高质量参考基因组的物种; 对于没有参考基因组的物种, 也可以通过 *de novo* 拼接, 以拼接后的片段作为参考序列来开发标记。GBS 分型技术所获得的 SNP 位点通常在目标群体中多态性较好, 而不像高密度 SNP 芯片那样只能固定地检测特定的位点。这一特点使得该技术非常适合在中国地方品种或高度纯化的品系中进行分型, 从而获得更大的信息量。

1.3 基因组重测序

基因组重测序(whole-genome resequencing, WGS)是对参考基因组序列已知物种的个体进行全基因组测序, 并在个体或群体水平进行序列差异性分析的测序方法。随着二代测序成本的降低和畜禽参考基因组序列的完善, 基因组重测序已成为最为迅速和全面的群体遗传变异鉴定和检测方法。相比高密度 SNP 芯片或者简化基因组测序, 基因组重测序的方法可以全面地挖掘基因序列差异和结构变异, 在全基因组水平上鉴定并检测与重要经济性状相关的突变位点, 具有重大的科研价值和产业价值。通过基因组重测序可以找到大量的 SNP、拷贝数变异(copy number variation, CNV)、插入缺失(InDel)、结构变异(structural variation, SV)等遗传变异。由于当前测序价格相对较高, 因此降低测序乘数是主要的测序策略, 一般为 0.1~2 \times , 即低覆盖度重测序。Hickey^[31]于 2013 年提出 GS2.0 概念, 用以区别基于高密度 SNP 芯片的 GS1.0, 认为可以通过对百万动物进行低覆盖度重测序, 结合基因型填充技术, 获得每个个体基因组上几乎所有的遗传变异。GS2.0 能够发现大量与性状紧密连锁的 QTL 或 QTN, 能够大幅提高基因组选择的准确性。

基于基因组重测序的 GS 还有另外一个策略(图 4), 即首先对群体中遗传贡献较大的个体进行 10~30 \times 的中高覆盖度的重测序, 结合单倍型估计技术构建群体主要的单倍型库; 对于大多数个体采用

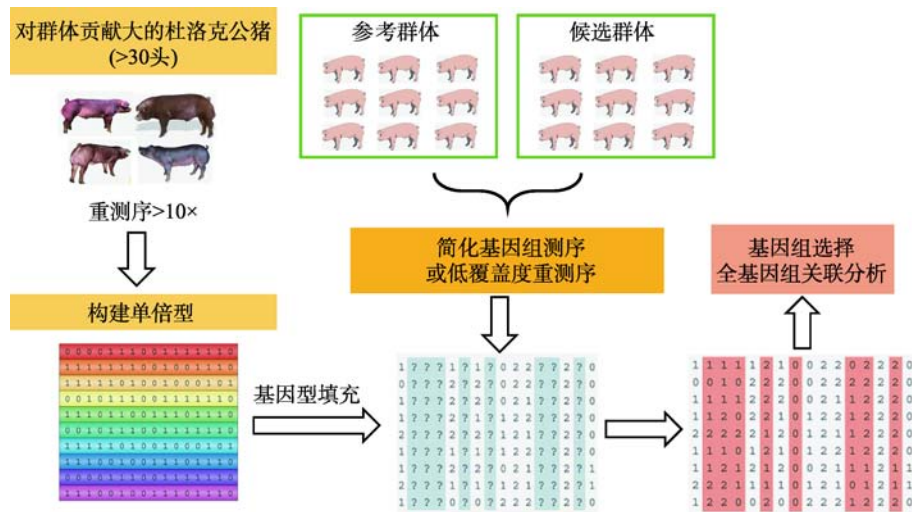


图 4 基于二代测序的基因组选择流程

Fig. 4 A diagram of genomic selection using next-generation sequencing technologies

高密度 SNP 芯片或简化基因组测序技术进行分型，通过基因型填充获得全基因组的遗传变异，从而用于 GEBV 的预测。

尽管当前受制于成本和分析的压力，基于基因组重测序的 GS 方法仍然处于研究阶段。但是，随着二代测序技术的发展，基于全基因组重测序的 GS 方法会逐渐成为主要的技术手段。

1.4 GS 应用中分型技术的选择

高密度 SNP 芯片是利用已知的 SNP 位点侧翼的序列设计探针，将探针固定合成在芯片上后，将待测定样本的 DNA 与芯片杂交并扫描荧光信号，从而鉴定这些 SNP 位点的基因型。简化基因组测序使用限制性内切酶酶切基因组 DNA，然后富集部分酶切产物进行下一步的测序，通过对测序数据的分析鉴定 SNP 等遗传变异信息；因为同一个物种的酶切位点比较稳定，所以不同个体最终的测序区域基本保持一致。这两种技术都是高性价比的全基因组 SNP 分型方法，主要区别在于高密度 SNP 芯片仅对已知的 SNP 位点进行检测，确定样本在这个位点的基因型，因此需要预先知道该物种的基因组 SNP 多态性信息（一般来源大规模重测序），然后筛选 SNP 来设计芯片。也就是说，高密度 SNP 芯片只能对已有的 SNP 位点进行分型，不能发现新的 SNP^[32]。而简化基因组测序本质上还是测序，即使这个物种没有任何已知的 SNP 信息，也能使用简化基因组测序进行

全基因组 SNP 分型。在价格上，高密度 SNP 芯片的单价从几百到上千人民币不等，主要取决于定制量、位点密度等。简化基因组测序的单价通常比 SNP 芯片低，主要取决于所研究物种的基因组大小和所选择的内切酶。Gorjanc 等^[33]认为在作物的双亲群体 (biparental segregating populations) 中，低至 1× 的 GBS 测序可以获得与 SNP 芯片一样的准确性，而其成本却只有 SNP 芯片的一半。

高密度 SNP 芯片和简化基因组测序具有各自的优势。随着测序价格的继续下降，以及生物信息学分析的流程化，基于二代测序技术的全基因组重测序进行 SNP 的鉴定和分型会越来越普及，特别是对于一些没有标准化芯片的非模式物种。当然，也可以选择 20~30 个代表性的个体进行重测序，获得这个群体具有一定代表性的 SNP 数据集，利用这些数据设计芯片，从而进行“廉价”的大规模基因分型。但是，遇到无标准化芯片的物种，测序无疑是最佳选择。即使对于一些有标准化芯片的模式生物，全基因组重测序也常常是更好的选择。在商业化 SNP 芯片密度不够的情况下，简化基因组测序是很好的替代方法，因为即使只对基因组的很小一部分进行测序，就可以轻松获得几十万的 SNP 标记。在地方猪种中，高密度 SNP 芯片也有其局限性。以藏猪为例，PorcineSNP60 芯片效果不理想，因为芯片上的 SNP 位点都是从常见的品种中筛选出来的，这些位

点在藏猪这样的特殊亚种中多态性比较差。因此, 简化基因组测序或全基因组重测序的效果会优于芯片。总之, 在具体研究中, 应该根据具体情况来选择全基因组 SNP 分型方案。随着测序价格的不断降低, 测序的确会不断侵蚀芯片的市场空间, 成为一种趋势。

2 统计分析模型的选择

2.1 贝叶斯方法

Meuwissen 等^[1]首次提出 GS 理论时, 提供了两种贝叶斯(Bayesian, Bayes)方法用以解决 SNP 标记数目通常远远多于表型记录的问题, 即 BayesA 和 BayesB。BayesA 假设所有 SNP 位点都有效应, 且所有 SNP 效应的方差服从尺度逆卡方分布的正态分布, 其中的两个参数——自由度和尺度参数与遗传结构直接相关, 能够确定遗传结构, 而这两个参数都是事先给定的。BayesA 使用 MCMC(马尔科夫链蒙特卡洛方法)方法构建 Gibbs 抽样链, 在模型中对标记效应进行求解。BayesB 与 BayesA 的区别在于对 SNP 效应的先验假设不同。BayesA 假设所有 SNP 都有效应, 而 BayesB 假设只有一小部分标记位点有效应, 其他大部分染色体片段效应为 0 (无效应位点的比例为 π); 这一小部分有效应的位点, 其效应方差服从的分布与 BayesA 一样。BayesB 中有效应位点的比例($1-\pi$)在模型中是预先设定的, 通常为 0.05 左右。与 BayesA 不同, BayesB 使用混合分布作为标记效应方差的先验, 所以难以构建标记效应和方差各自的完全条件后验分布, 因此 BayesB 使用 MH (Metropolis-Hasting) 抽样对标记效应和方差进行联合抽样。

在 Meuwissen 提出的两种贝叶斯模型基础之上, 研究人员又提出了多种 GEBV 估计的 Bayes 模型。BayesC 模型^[34]使用混合分布作为标记效应的先验分布, 但在 BayesC 模型中, π 是未知的, 需要在模型中求解得到, 其他与 BayesB 相同。Habier 等^[34]还提出了 BayesC π 、BayesD π 方法: BayesC π 与 BayesC 的主要区别在于 BayesC π 假设有效应的 SNP 的效应方差相同, 而 BayesC 假设其效应方差是不同的。BayesC 和 BayesC π 相比于 BayesA 和 BayesB

的最大改进之处在于对模型中无效应 SNP 位点的比例 π 进行估计^[16,35]。BayesD π ^[34]在模型中假设 π 值服从 $U[0, 1]$ 的均匀分布, 同时也要对尺度参数进行求解, 其假设尺度参数的先验分布为 Gamma(1, 1) 分布, 其后验分布也是 Gamma 分布, 可以直接从后验分布中抽样。Bayes LASSO (Least Absolute Shrinkage and Selection Operator)^[36~40]假定 SNP 标记效应服从 Laplace 分布, 而 Lplace 分布等价于方差服从指数分布的正态分布。Laplace 分布与 BayesA 假定的 SNP 标记效应 Student's t 分布具有相同的特点: 具有比正态分布更重的两尾概率。在高维数据线性回归中常用到的方法除了 LASSO, 还有 EN (Elastic net)^[41], 这两种方法都是在惩罚函数中增加约束项, 同时实现变量选择与参数估计, 可以将很多变量效应压缩至(接近)0。Bayes LASSO 在 GS 算法应用中的一种拓展策略是使用广义正态分布作为标记效应先验。Bayes LASSO 的思想和 BayesA 一样, 不同之处在于它假设标记效应服从另一种分布—拉普拉斯分布, 所以标记效应的后验分布也随之改变。

其他的贝叶斯方法, 如 Bayes SSVS^[42]、fBayesB^[43]、wBSR^[44]、BAL/IAL^[45]、emBayesR^[46]、EBL^[47]、BayesRS^[48]和 Bayes TA^[49]等, 都是在前人研究基础之上对模型中的先验假设进行变换和模型中的参数进行优化, 以期寻找最适合群体的假设模型和参数。目前广泛应用的贝叶斯算法仍是 BayesA、BayesB、BayesC π 和 Bayes LASSO, 这是由于它们计算的结果稳定, 具有较高的 GEBV 估计准确性。总之, 在经典贝叶斯方法基础上, 贝叶斯方法的改进算法及其参数优化策略围绕着以提高 GEBV 估计的准确性为目的, 通过生物遗传算法与实际的群体情况相结合, 寻找最适的模型假设和参数优化方法, 使得 GEBV 更具有育种价值。

2.2 GBLUP 方法

另一种估计 GEBV 的方法是通过构建基因组关系矩阵(G 矩阵)替换基于系谱信息构建的分子血缘关系矩阵(numerator relationship matrix, NRM 或 A 矩阵), 进而使用最佳线性无偏预测(best linear unbiased prediction, BLUP)方法直接估计 GEBV, 即 GBLUP 法。VanRaden^[16]提出了 G 矩阵的构建方法, 将 SNP

基因型进行编码, 纯合基因型分别编码为 1 和 -1, 杂合基因型为 0, 假设 p_i 为位点 i 的第二个等位基因的频率, 则每个基因型编码后的值减去相应的处于哈代-温伯格平衡 (Hardy-Weinberg equilibrium, HWE) 时的均值 $2 \times (p_i - 0.5)$, 得到 Z 矩阵, 按照公式 (1-1) 即可计算得到 G 矩阵:

$$G = ZZ' / (2 \sum p_i(1 - p_i)) \quad (1-1)$$

根据混合模型方程组 (mixed model equation, MME) 对育种值进行求解:

$$\begin{bmatrix} \hat{\beta} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} XX & XZ \\ Z'X & Z'Z + \lambda G^{-1} \end{bmatrix}^{-1} \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \quad (1-2)$$

其中, G^{-1} 为基因组关系矩阵 G 的逆矩阵,

$$\lambda = \frac{\sigma_e^2}{\sigma_a^2}.$$

与贝叶斯方法相比, GBLUP 不需要先利用参考群体估计 SNP 标记效应, 再计算 GEBV; 而是可以直接将有表型及无表型个体可以放在同一个模型中, 同时估计出有表型和无表型个体的 GEBV 及其准确性^[50]。从计算速度来讲, GBLUP 比 Bayes 方法快很多, 因此更加适用于现场应用时快速获得 GEBV^[51]。Azevedo 等^[52]比较了 GBLUP 与几种改进的 Bayes LASSO 之间的准确性, 通过对多种不同遗传力、不同遗传基础性状的 GEBV 估计, 认为 GBLUP 总体上表现最好, 其次是 BayesA 或 BayesB。

2.3 一步法 GBLUP

一步法 GBLUP (single-step GBLUP, ssGBLUP) 模型是传统基于系谱信息的 BLUP 法和基于 SNP 标记信息的 GBLUP 法的合并, 它的模型形式上与 BLUP 及 GBLUP 法并无区别。如公式 (1-2) 中, ssGBLUP 是用 H 矩阵替代 GBLUP 中的 G 矩阵, 从而将没有基因型的个体与有基因型的个体放在同一个模型中进行 EBV 或 GEBV 的估计, H 矩阵结构如公式 (1-3):

$$H = \begin{bmatrix} A_{11} - A_{12}A_{22}^{-1}A_{21} + A_{12}A_{22}^{-1}GA_{22}^{-1}A_{21} & A_{12}A_{22}^{-1}G \\ GA_{22}^{-1}A_{21} & G \end{bmatrix} \quad (1-3)$$

其中, A_{11} 和 A_{22} 分别为无基因型个体与有基因型个体的基于系谱信息的个体间分子血缘关系矩阵, G 为有基因型个体间的基因组关系矩阵。通过 H 矩

阵, 有基因型个体的 G 矩阵与基于系谱信息的 A 矩阵便合并到了同一个模型中, GS 便可与传统遗传评估方式实现有机整合。而对于 ssGBLUP 的求解, 同样可以采用 MME 方程组求解, 只需将 G^{-1} 矩阵替换为 H^{-1} 矩阵即可^[53-55]:

$$H^{-1} = A^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & G_w^{-1} - A_{22}^{-1} \end{bmatrix} \quad (1-4)$$

$$G_w = (1 - w)G + wA_{22}$$

其中, w 为加权因子, 表示 SNP 标记信息未能捕获的加性遗传效应的比例; w 的取值主要取决于 SNP 标记所能解释的遗传方差比例, 可用方差组分估计的方式确定, 或尝试不同的值确定最优参数^[56-58]。

ssGBLUP 有效地解决了畜禽 GS 实施中如何充分利用已有表型记录的历史数据等问题, 广泛被用于育种实践中^[58,59]。因为大多数育种群体具有丰富翔实的系谱和表型记录, 但限于经费、人力、时间等因素, 难以对全部个体进行基因型测定, 特别是一些年代久远的个体因为没有 DNA 组织样品而无法再进行基因型测定。ssGBLUP 能将无基因型个体的系谱及表型数据和有基因型的个体的基因组信息结合起来^[60,61], 会大大提高基因组选择的准确性。Christensen 等^[58]对 2668 头杜洛克猪群体的日增重和饲料转化率两个性状进行基因组选择准确性分析, 结果表明多性状模型的 ssGBLUP 准确性最高, 比常规 BLUP 方法准确性提高了 16%。Tsuruta 等^[62]用有系谱和表型记录的 7 860 183 头美国荷斯坦奶牛和 17 293 头有基因型数据的个体, 对荷斯坦奶牛的 18 个性状进行分析, 结果也证明 ssGBLUP 的准确性明显高于常规 BLUP 方法。

3 GS 在畜禽育种中的应用

3.1 牛

GS 在奶牛育种中的应用领先于其他畜禽。由于奶牛个体育种价值大, 群体组织好, 数据记录完善, 而且传统的奶牛育种完全依赖于后裔测定对种公牛进行遗传评估, 使得世代间隔较长 (达 53~70 个月), 采用 GS 可准确地预测青年公牛种用价值, 使世代间隔缩短为 21 个月 (图 5), 因此 GS 在奶牛育种中具有巨大的应用价值。从 2008 年开始, 主要几个奶业发达国家, 包括新西兰 (2008 年)、美国 (2009 年)、

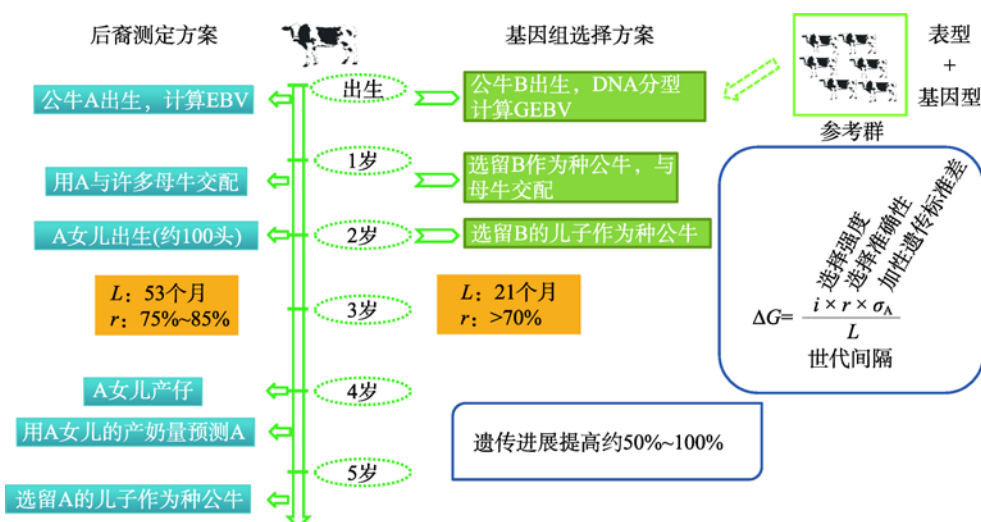


图 5 奶牛基因组选择降低世代间隔

Fig. 5 Genomic selection reduces generation intervals of dairy cattle

加拿大(2009年)、德国(2009年)、澳大利亚(2011年)、意大利和瑞士(2011年),开始将GS应用到遗传评估中。除此之外,这些国家间还不断寻求合作以建立区域化的奶牛GS育种体系,如2008年美国、加拿大、意大利组成的北美奶牛育种体系以及2010年比利时、丹麦、芬兰、法国、德国、荷兰和瑞典联合形成的欧洲奶牛育种体系。区域化的奶牛GS育种体系使得参考群体规模迅速扩大,GS选择准确性逐渐提高,大大促进了GS技术在奶牛育种中的普及。除了常规的产奶量和乳脂率等性状,奶牛GS研究也开始关注一些常被忽视但非常有应用潜力的性状,如肢蹄健康、饲料转化率和甲烷排放量等^[63]。我国于2008年开始启动奶牛GS研究,2012年正式将GS技术应用于荷斯坦奶牛的遗传评估中,评估的性状共14个,包括产奶量、体细胞计数、体型评分等,其中产奶性状基因组预测准确性为0.59~0.76,比传统BLUP方法提高了0.13~0.30。

目前,奶牛GS的参考群体一般由经后裔测定和基因型测定的验证公牛组成。除了通过区域合作扩大参考群体外,对母牛进行基因型测定,也是扩大参考群体的一条重要途径。随着高密度SNP芯片和二代测序成本的下降,将来所有母牛可能进行基因型测定。Buch等^[64]使用验证公牛与经基因型测定的女儿构建参考群体,结果表明GS的准确性比仅使用验证公牛时要高。对于荷斯坦以外的奶牛品种,

如娟姗牛、瑞士褐牛、西门塔尔、挪威红牛等,由于其群体相对小,并且地理分布分散,参考群相对较小,影响了GS技术的应用。

3.2 猪

与奶牛不同,在猪的育种体系中,GS应用的前提是提高GEBV估计的准确性^[58],这是因为在传统育种体系中,猪的世代间隔已经控制在较短的时间,很难再对其进行大幅缩减。因此,猪的GS应用主要是靠提高GEBV的准确性来获得额外的遗传进展,特别是对传统育种中选择准确性低的性状,如繁殖性状、屠宰性状、肉质性状、抗病性状等,GS具有更大的优势。目前,国际上主要的猪育种公司都已经开始GS的研究和应用,如英国PIC公司、丹麦DanBred公司、荷兰TOPIGS公司和Hypor公司。PIC公司自2013年开始,其核心群综合遗传指数的进展幅度增加了35%,这主要得益于GS的应用。在我国,猪GS的研究和应用方兴未艾,温氏集团自2011年开始构建父系杜洛克猪的参考群体,主要对生长性状和饲料利用率等性状进行GS选择。

GS对猪的生长或胴体性状具有很好的预测准确性。Ostersen等^[7]利用Illumina PorcineSNP60芯片对1911头杜洛克猪的日增重和饲料转化率两个性状的GS准确性进行分析,结果表明当使用DEBV作为反应变量时,GEBV的可靠性分别达0.34和0.20。Christensen等^[58]对2668头丹系杜洛克猪30~100 kg

日增重和饲料转化率的 GS 准确性进行研究 (其中仅 1474 头猪具有饲料转化率的表型记录), 结果显示改进后的 ssGBLUP 获得最高的准确性, 其 GEBV 与校正表型的相关系数分别达 0.353 和 0.352。Andersen-Ranberg 等^[65]在第 10 届遗传学应用于畜牧生产世界大会(10th WCGALP)上表示, 通过对 9745 头挪威长白猪的 GS 分析表明, 对于生长和屠宰性状 (达 40 kg 日龄、40~120 kg 日龄、40~120 kg 饲料采食量、瘦肉率和胴体比例) ,GS 的应用相比传统 BLUP 育种能获得额外的 8%~34%的遗传进展。Miar 等^[66]采用 PorcineSNP60 芯片对 2384 头三元杂杜长大猪及其父母进行基因分型, 分析了肉质性状的 GS 准确性, 结果显示在纯种中肉质性状的 GEBV 准确性约为 0.21, 而杂种群体中 GEBV 的准确性提高至 0.25。

然而, GS 在母猪的繁殖性状中应用受限于参考群体的大小。由于繁殖性状的遗传力低, 为获得足够的准确性, 需要大量的表型和基因型数据。但母猪的个体价值相对低, 高密度 SNP 芯片对其来说成本依然高, 限制了 GS 在母系猪育种中的应用。但该问题可通过新的基因型分型策略来解决, 例如使用低密度芯片与高密度芯片结合测定基因型, 采用基因型填充技术, 把低密度标记填充为高密度标记, 或者开发新的基于测序的全基因组 SNP 分型方法, 如 GBS 技术或极低覆盖度重测序技术等。Simianer 等^[67]研究指出, 对于猪的产仔性状, GS 可提高年遗传进展 37%。Cleveland 等^[68]在参考群中用传统 BLUP 育种值代替母猪繁殖性状表型 (如总产仔数、死胎数) 作为反应变量, 可使 GEBV 准确性达 0.63~0.83, 近似于奶牛较大参考群体时的准确性。总之, GS 为新的育种设计及新性状的评估提供了可能。

3.3 鸡

从个体角度, GS 在鸡育种中很难普及, 因为其个体价值远小于基因分型成本。但是, 如果从整个育种体系的角度来看, 为了培育特定目标群体进行配套系生产的话, GS 在鸡的应用还是有一定的潜力的。Long 等^[69]利用 5000 个与后代死亡率相关的 SNP 标记对肉鸡 0-14 天龄的死亡率进行预测, 结果显示当利用 SNP 信息的进行选择时, 其准确性与不使用 SNP 信息相比提高了 50%~90%。随后, 陆续开展了关于饲料转化效率、肉鸡 6 周龄体重、胸肌面积、

腿评分等性状^[70,71]及蛋鸡产蛋量、蛋重、沙门氏菌抗性等性状^[72,73]的 GS 研究。美国海兰公司对蛋鸡进行 GS 研究, 研究了包括产蛋量、蛋重、蛋壳质量和性成熟等重要经济性状, 目标是将世代间隔缩短一半, 即从 12 个月减为 6 个月, 从而提高年遗传进展; 其方案是使用高密度 SNP 芯片测定个体基因型, 并对 GS 群体的大小及结构进行优化, 使选择准确性最大化, 且年近交速率不增加。美国爱维茵公司在 GS 方面的研究重点是通过 GEBV 提高选择准确性, 但并不改变世代间隔。然而, GS 在鸡等家禽育种上的应用仍旧存在一些障碍, 其中最为显著的就是成本问题。相比奶牛等可以通过冻精和人工受精扩大影响范围的动物, 种鸡对群体改良的影响能力要明显较小, 因此在鸡中进行 GS 应用带来的经济效益也就较低。但是, 随着二代测序技术的发展, 基因分型成本可能会降低到育种公司可以接受的范围, 此时, GS 在家禽育种上可能会逐渐普及。

4 展 望

GS 作为一种新的畜禽遗传评估方法, 比传统 BLUP 方法有明显的优势。随着基因分型成本的下降, GS 技术有望在国内大型育种公司中应用普及。由于已有的 SNP 芯片在特定群体中多态性差, 越来越多的研究和应用开始使用自定制高密度 SNP 芯片, 如猪的 PorcineSNP55K 芯片, 在肉鸡、肉牛和蛋鸡等物种中也有相应报道。随着二代测序价格的下降, 基于二代测序技术的 GS 优势日趋明显, 有望在 5~10 年内替代高密度 SNP 芯片, 成为 GS 应用的主要分型手段。基于二代测序的 GS 技术主要面临的是分型准确性和分析时效性的问题, 分型准确性可以通过优化测序方案和基因型填充策略来提高; 而数据分析的时效性需要借助于自动化/智能化分析流程的建立以及基于“硬件加速”技术的软件开发。随着基因分型个体数极大累积或标记密度极大提高, 当前常用的基于单点 SNP 的 GEBV 估计模型将具有局限性, 基于单倍型信息的统计模型可能会成为未来的研究方向。除此之外, 在统计模型中增加显性效应、上位效应和印记效应的应用会越来越多^[74]。最后, 借助目前智能化农牧设备开发的热潮, 应该尽快将 GS 应用从“育种场→实验室→育种场”的所需样品收集、表型测定、数据分析、个体选留等流

程整合进入常规育种生产中, 加快 GS 在动物育种中的应用。

参考文献(References):

- [1] Meuwissen THE, Hayes BJ, Goddard ME. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*, 2001, 157(4): 1819–1829. [DOI]
- [2] Goddard ME, Hayes BJ. Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat Rev Genet*, 2009, 10(6): 381–391. [DOI]
- [3] Olson KM, VanRaden PM, Tooker ME, Cooper TA. Differences among methods to validate genomic evaluations for dairy cattle. *J Dairy Sci*, 2011, 94(5): 2613–2620. [DOI]
- [4] Pszczola M, Calus MPL. Updating the reference population to achieve constant genomic prediction reliability across generations. *Animal*, 2016, 10(6): 1018–1024. [DOI]
- [5] VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS. *Invited review*: Reliability of genomic predictions for North American Holstein bulls. *J Dairy Sci*, 2009, 92(1): 16–24. [DOI]
- [6] Spelman RJ, Hayes BJ, Berry DP. Use of molecular technologies for the advancement of animal breeding: genomic selection in dairy cattle populations in Australia, Ireland and New Zealand. *Anim Prod Sci*, 2013, 53(9): 869–875. [DOI]
- [7] Ostensen T, Christensen OF, Henryon M, Nielsen B, Su GS, Madsen P. Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. *Genet Sel Evol*, 2011, 43: 38. [DOI]
- [8] Tribout T, Larzul C, Phocas F. Efficiency of genomic selection in a purebred pig male line. *J Anim Sci*, 2012, 90(12): 4164–4176. [DOI]
- [9] Liu TF, Qu H, Luo CL, Li XW, Shu DM, Lund MS, Su GS. Genomic selection for the improvement of antibody response to newcastle disease and avian influenza virus in chickens. *PLoS One*, 2014, 9(11): e112685. [DOI]
- [10] Wolc A, Zhao HH, Arango J, Settar P, Fulton JE, O'Sullivan NP, Preisinger R, Stricker C, Habier D, Fernando RL, Garrick DJ, Lamont SJ, Dekkers JCM. Response and inbreeding from a genomic selection experiment in layer chickens. *Genet Sel Evol*, 2015, 47: 59. [DOI]
- [11] Palaiokostas C, Ferraresso S, Franch R, Houston RD, Bargelloni L. Genomic prediction of resistance to pasteurellosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. *G3 (Bethesda)*, 2016, 6(11): 3693–3700. [DOI]
- [12] Dou JZ, Li X, Fu Q, Jiao WQ, Li YP, Li TQ, Wang YF, Hu XL, Wang S, Bao ZM. Evaluation of the 2b-RAD method for genomic selection in scallop breeding. *Sci Rep*, 2016, 6: 19244. [DOI]
- [13] Beyene Y, Semagn K, Mugo S, Tarekegne A, Babu R, Meisel B, Sehabiague P, Makumbi D, Magorokosho C, Oikeh S, Gakunga J, Vargas M, Olsen M, Prasanna BM, Banziger M, Crossa J. Genetic gains in grain yield through genomic selection in eight Bi-parental maize populations under drought stress. *Crop Sci*, 2015, 55(1): 154–163. [DOI]
- [14] Zhao YS, Gowda M, Liu WX, Würschum T, Maurer HP, Longin FH, Ranc N, Reif J. Accuracy of genomic selection in European maize elite breeding populations. *Theor Appl Genet*, 2012, 124(4): 769–776. [DOI]
- [15] Meuwissen TH. Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. *Genet Sel Evol*, 2009, 41(1): 35. [DOI]
- [16] VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci*, 2008, 91(11): 4414–4423. [DOI]
- [17] Da Y. Multi-allelic haplotype model based on genetic partition for genomic prediction and variance component estimation using SNP markers. *BMC Genet*, 2015, 16: 144. [DOI]
- [18] Ramos AM, Crooijmans RPMA, Affara NA, Amaral AJ, Archibald AL, Beever JE, Bendixen C, Churcher C, Clark R, Dehais P, Hansen MS, Hedegaard J, Hu ZL, Kerstens HH, Law AS, Megens HJ, Milan D, Nonneman DJ, Rohrer GA, Rothschild MF, Smith TPL, Schnabel RD, Van Tassell CP, Taylor JF, Wiedmann RT, Schook LB, Groenen MAM. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PLoS One*, 2009, 4(8): e6524. [DOI]
- [19] Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet*, 2011, 12(7): 499–510. [DOI]
- [20] Van Orsouw NJ, Hogers RCJ, Janssen A, Yalcin F, Snoeijers S, Verstege E, Schneiders H, Van Der Poel H, Van Oeveren J, Versteegen H, Van Eijk MJT. Complexity reduction of polymorphic sequences (CROPS): A novel approach for large-scale polymorphism discovery in complex genomes. *PLoS One*, 2007, 2(11): e1172. [DOI]
- [21] Van Tassell CP, Smith TPL, Matukumalli LK, Taylor JF, Schnabel RD, Lawley CT, Haudenschild CD, Moore SS, Warren WC, Sonstegard TS. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat Methods*, 2008, 5(3): 247–252. [DOI]
- [22] Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS One*, 2008, 3(10): e3376. [DOI]
- [23] Andolfatto P, Davison D, Erezylmaz D, Hu TT, Mast J, Sunayama-Morita T, Stern DL. Multiplexed shotgun genotyping for rapid and efficient genetic mapping. *Genome Res*, 2011, 21(4): 610–617. [DOI]
- [24] Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K,

- Buckler ES, Mitchell SE. A robust, Simple Genotyping-by-Sequencing (GBS) approach for high diversity species. *PLoS One*, 2011, 6(5): e19379. [DOI]
- [25] Rife TW, Wu SY, Bowden RL, Poland JA. Spiked GBS: a unified, open platform for single marker genotyping and whole-genome profiling. *BMC Genomics*, 2015, 16: 248. [DOI]
- [26] Poland JA, Brown PJ, Sorrells ME, Jannink JL. Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One*, 2012, 7(2): e32253. [DOI]
- [27] Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. Double digest RADseq: An inexpensive method for de novo snp discovery and genotyping in model and non-model species. *PLoS One*, 2012, 7(5): e37135. [DOI]
- [28] Wang S, Meyer E, McKay JK, Matz MV. 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nat Methods*, 2012, 9(8): 808-810. [DOI]
- [29] Chen Q, Ma YF, Yang YM, Chen ZL, Liao RR, Xie XX, Wang Z, He PF, Tu YY, Zhang XZ, Yang CS, Yang HJ, Yu FQ, Zheng YM, Zhang ZW, Wang QS, Pan YC. Genotyping by genome reducing and sequencing for outbred animals. *PLoS One*, 2013, 8(7): e67500. [DOI]
- [30] Jiang ZH, Wang HY, Michal JJ, Zhou X, Liu B, Woods LCS, Fuchs RA. Genome wide sampling sequencing for SNP genotyping: Methods, challenges and future development. *Int J Biol Sci*, 2016, 12(1): 100-108. [DOI]
- [31] Hickey JM. Sequencing millions of animals for genomic selection 2.0. *J Anim Breed Genet*, 2013, 130(5): 331-332. [DOI]
- [32] DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, Del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernysky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*, 2011, 43(5): 491-498. [DOI]
- [33] Gorjanc G, Battagin M, Dumasy JF, Antolin R, Gaynor RC, Hickey JM. Prospects for cost-effective genomic selection via accurate within-family imputation. *Crop Sci*, 2017, 57(1): 216-228. [DOI]
- [34] Habier D, Fernando RL, Kizilkaya K, Garrick DJ. Extension of the bayesian alphabet for genomic selection. *BMC Bioinformatics*, 2011, 12: 186. [DOI]
- [35] Gianola D, De Los Campos G, Hill WG, Manfredi E, Fernando R. Additive genetic variability and the bayesian alphabet. *Genetics*, 2009, 183(1): 347-363. [DOI]
- [36] Colombani C, Legarra A, Fritz S, Guillaume F, Croiseau P, Ducrocq V, Robert-Granié C. Application of Bayesian least absolute shrinkage and selection operator (LASSO) and BayesC π methods for genomic selection in French Holstein and Montbéliarde breeds. *J Dairy Sci*, 2013, 96(1): 575-591. [DOI]
- [37] Hastie T, Tibshirani R. Bayesian backfitting (with comments and a rejoinder by the authors). *Stat Sci*, 2000, 15(3): 196-223. [DOI]
- [38] Legarra A, Robert-Granié C, Croiseau P, Guillaume F, Fritz S. Improved Lasso for genomic selection. *Genet Res*, 2011, 93(1): 77-87. [DOI]
- [39] Xu S. An expectation-maximization algorithm for the Lasso estimation of quantitative trait locus effects. *Heredity*, 2010, 105(5): 483-494. [DOI]
- [40] De Los Campos G, Naya H, Gianola D, Crossa J, Legarra A, Manfredi E, Weigel K, Cotes JM. Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics*, 2009, 182(1): 375-385. [DOI]
- [41] Croiseau P, Legarra A, Guillaume F, Fritz S, Baur A, Colombani C, Robert-Granié C, Boichard D, Ducrocq V. Fine tuning genomic evaluations in dairy cattle through SNP pre-selection with the Elastic-Net algorithm. *Genet Res*, 2011, 93(6): 409-417. [DOI]
- [42] Yi NJ, Xu SZ. Bayesian LASSO for quantitative trait loci mapping. *Genetics*, 2008, 179(2): 1045-1055. [DOI]
- [43] Meuwissen TH, Solberg TR, Shepherd R, Woolliams JA. A fast algorithm for BayesB type of prediction of genome-wide estimates of genetic value. *Genet Sel Evol*, 2009, 41: 2. [DOI]
- [44] Hayashi T, Iwata H. EM algorithm for Bayesian estimation of genomic breeding values. *BMC Genet*, 2010, 11: 3. [DOI]
- [45] Sun W, Ibrahim JG, Zou F. Genomewide multiple-loci mapping in experimental crosses by iterative adaptive penalized regression. *Genetics*, 2010, 185(1): 349-529. [DOI]
- [46] Shepherd RK, Meuwissen TH, Woolliams JA. Genomic selection and complex trait prediction using a fast EM algorithm applied to genome-wide markers. *BMC Bioinformatics*, 2010, 11: 529. [DOI]
- [47] Mutshinda CM, Sillanpää MJ. Extended bayesian LASSO for multiple quantitative trait loci mapping and unobserved phenotype prediction. *Genetics*, 2010, 186(3): 1067-1075. [DOI]
- [48] Brøndum RF, Su GS, Lund MS, Bowman PJ, Goddard ME, Hayes BJ. Genome position specific priors for genomic prediction. *BMC Genomics*, 2012, 13: 543. [DOI]
- [49] Wang CL, Ding XD, Wang JY, Liu JF, Fu WX, Zhang Z, Yin ZJ, Zhang Q. Bayesian methods for estimating GEBVs of threshold traits. *Heredity*, 2013, 110(3): 213-219. [DOI]
- [50] Ricard A, Danvy S, Legarra A. Computation of deregressed proofs for genomic selection when own phenotypes exist with an application in French show-jumping horses. *J Anim Sci*, 2013, 91(3): 1076-1085. [DOI]
- [51] Da Y, Wang CK, Wang SW, Hu G. Mixed model methods for genomic prediction and variance component estimation

- of additive and dominance effects using SNP markers. *PLoS One*, 2014, 9(1): e87666. [DOI]
- [52] Azevedo CF, De Resende MDV, Silva FF, Viana JMS, Valente MSF, Resende MFR Jr, Muñoz P. Ridge, Lasso and Bayesian additive-dominance genomic models. *BMC Genet*, 2015, 16: 105. [DOI]
- [53] VanRaden PM, Null DJ, Sargolzaei M, Wiggans GR, Tooker ME, Cole JB, Sonstegard TS, Connor EE, Winters M, Van Kaam JBCHM, Valentini A, Van Doormaal BJ, Faust MA, Doak GA. Genomic imputation and evaluation using high-density Holstein genotypes. *J Dairy Sci*, 2013, 96(1): 668–678. [DOI]
- [54] Liu Z, Goddard ME, Reinhardt F, Reents R. A single-step genomic model with direct estimation of marker effects. *J Dairy Sci*, 2014, 97(9): 5833–5850. [DOI]
- [55] Legarra A, Christensen OF, Aguilar I, Misztal I. Single Step, a general approach for genomic selection. *Livest Sci*, 2014, 166: 54–65. [DOI]
- [56] Gao HD, Christensen OF, Madsen P, Nielsen US, Zhang Y, Lund MS, Su GS. Comparison on genomic predictions using three GBLUP methods and two single-step blending methods in the Nordic Holstein population. *Genet Sel Evol*, 2012, 44: 8. [DOI]
- [57] Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, Lawlor TJ. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J Dairy Sci*, 2010, 93(2): 743–752. [DOI]
- [58] Christensen OF, Madsen P, Nielsen B, Ostensen T, Su G. Single-step methods for genomic evaluation in pigs. *Animal*, 2012, 6(10): 1565–1571. [DOI]
- [59] Ostensen T, Christensen OF, Madsen P, Henryon M. Sparse single-step method for genomic evaluation in pigs. *Genet Sel Evol*, 2016, 48: 48. [DOI]
- [60] Legarra A, Aguilar I, Misztal I. A relationship matrix including full pedigree and genomic information. *J Dairy Sci*, 2009, 92(9): 4656–4663. [DOI]
- [61] Misztal I, Legarra A, Aguilar I. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J Dairy Sci*, 2009, 92(9): 4648–4655. [DOI]
- [62] Tsuruta S, Misztal I, Aguilar I, Lawlor TJ. Multiple-trait genomic evaluation of linear type traits using genomic and phenotypic data in US Holsteins. *J Dairy Sci*, 2011, 94(8): 4198–4204. [DOI]
- [63] Chesnais JP, Cooper TA, Wiggans GR, Sargolzaei M, Pryce JE, Miglior F. Using genomics to enhance selection of novel traits in North American dairy cattle. *J Dairy Sci*, 2016, 99(3): 2413–2427. [DOI]
- [64] Buch LH, Sørensen MK, Berg P, Pedersen LD, Sørensen AC. Genomic selection strategies in dairy cattle: Strong positive interaction between use of genotypic information and intensive use of young bulls on genetic gain. *J Anim Breed Genet*, 2012, 129(2): 138–151. [DOI]
- [65] Andersen-Ranberg, I.A., Grindfleck, E. Implementation of genomic selection in Norsvin genetic program: genetic gain in production and maternal traits in Norsvin Landrace. In: Proceedings of the 10th World Congress Genet. Appl. Livest. Prod., Vancouver, Canada, 2014, 508. [DOI]
- [66] Miar Y, Plastow G, Bruce H, Moore S, Manafiazar G, Kemp R, Charagu P, Huisman A, Van Haandel B, Zhang CY, McKay R, Wang ZQ. Genetic and phenotypic correlations between performance traits with meat quality and carcass characteristics in commercial crossbred pigs. *PLoS One*, 2014, 9(10): e110105. [DOI]
- [67] Simianer H, Pimentel ECG. Robust QTL fine mapping by applying a quantitative transmission disequilibrium test to the Mendelian sampling term. *J Anim Breed Genet*, 2009, 126(6): 432–442. [DOI]
- [68] Cleveland MA, Forni S, Deeb N, Maltecca C. Genomic breeding value prediction using three Bayesian methods and application to reduced density marker panels. *BMC Proc*, 2010, 4 (Suppl. 1): S6. [DOI]
- [69] Long N, Gianola D, Rosa GJM, Weigel KA, Avendaño S. Machine learning classification procedure for selecting SNPs in genomic selection: application to early mortality in broilers. *J Anim Breed Genet*, 2007, 124(6): 377–389. [DOI]
- [70] González-Recio O, Gianola D, Rosa GJ, Weigel KA, Kranis A. Genome-assisted prediction of a quantitative trait measured in parents and progeny: application to food conversion rate in chickens. *Genet Sel Evol*, 2009, 41: 3. [DOI]
- [71] Simeone R, Misztal I, Aguilar I, Vitezica ZG. Evaluation of a multi-line broiler chicken population using a single-step genomic evaluation procedure. *J Anim Breed Genet*, 2012, 129(1): 3–10. [DOI]
- [72] Wolc A, Zhao HH, Arango J, Settar P, Fulton JE, O'Sullivan NP, Preisinger R, Stricker C, Habier D, Fernando RL, Garrick DJ, Lamont SJ, Dekkers JCM. Response and inbreeding from a genomic selection experiment in layer chickens. *Genet Sel Evol*, 2015, 47(1): 59. [DOI]
- [73] Liu TF, Qu H, Luo CL, Li XW, Shu DM, Lund MS, Su GS. Genomic selection for the improvement of antibody response to Newcastle disease and avian influenza virus in chickens. *PLoS One*, 2014, 9(11): e112685. [DOI]
- [74] Jiang J, Shen B, O'Connell JR, VanRaden PM, Cole JB, Ma L. Dissection of additive, dominance, and imprinting effects for production and reproduction traits in Holstein cattle. *BMC Genomics*, 2017, 30: 18(1): 425. [DOI]